International Journal of Assessment Tools in Education
2023, Vol. 10, No. 1, 76–97

https://doi.org/10.21449/ijate.1102435

Published at https://ijate.net/ https://dergipark.org.tr/en/pub/ijate ***Research Article***

Spatial ability test for university students: Development, validity and reliability studies

Kubra Acikgul[1,*], Suleyman Nihat Sad[2], Bilal Altay[2]

[1]İnönü University, Faculty of Education, Department of Mathematics and Science Education, Türkiye
[2]İnönü University, Faculty of Education, Department of Educational Sciences, Türkiye
[3] İnönü University, Faculty of Education, Department of Mathematics and Science Education, Türkiye

ARTICLE HISTORY

Received: Apr. 12, 2022

Accepted: Feb. 26, 2023

Keywords:

Student evaluation,
Test development,
Spatial ability.

Abstract: This study aimed to develop a useful test to measure university students' spatial abilities validly and reliably. Following a sequential explanatory mixed methods research design, first, qualitative methods were used to develop the trial items for the test; next, the psychometric properties of the test were analyzed through quantitative methods using data obtained from 456 university students. As a result, a multiple-choice spatial ability test with 27 items and five options was created, divided into three subtests: spatial relations, spatial visualization, and spatial orientation. The results suggested that scores obtained from the spatial ability test and its subtests are valid and reliable.

1. INTRODUCTION

Spatial ability is regarded as a critical component of human abilities (Lohman, 1993) and a prerequisite for scientific thinking (Clements & Battista, 1992). Spatial ability has an important role in the assimilation and use of preexisting knowledge as well as in the development of new knowledge and creativity (Kell et al., 2013). For example, the mental rotation skill is apparently an inevitable spatial ability for some popular professions, including dentistry, medicine, architecture, interior design, engineering, navigation, etc. (Kerkman et al., 2000). The decisive role of spatial abilities in the development of knowledge and skills in the fields of science, technology, engineering, and mathematics (STEM) is emphasized in many studies (e.g., Contreras et al., 2018; Gilligan et al., 2017). More specifically, spatial abilities are reported to have a critical role in enhancing the performance of learning mathematics and geometry (Battista et al., 1982; Gilligan et al., 2017; Sarama & Clements, 2009) and in developing mathematical thinking skills (Young et al., 2018).

Despite the importance and key role attributed to spatial abilities in many fields, the lack of a clear consensus on the definition and components of spatial ability confuses measuring spatial ability (D'Oliveira, 2004; Eliot & Hauptman, 1981; National Research Council (NRC), 2006). D'Oliveira (2004) reviewed the four main reasons for this confusion as follows: 1) Different definitions ascribed to spatial ability, 2) Different numbers of components of spatial ability, 3)

---

*CONTACT: Kübra AÇIKGÜL ✉ kubra.acikgul@inonu.edu.tr İnönü University, Faculty of Education, Department of Mathematics and Science Education, Türkiye

e-ISSN: 2148-7456 /© **IJATE** 2023

Different names given to the components of spatial ability, and 4) Quite a variety of spatial ability tests. In a similar vein, Eliot and Hauptman (1981) asserted that inconsistency among the methods and tools used to measure spatial ability further complicated the problem of a lack of consensus in the spatial ability literature. Thus, this study aimed to present a detailed review of the literature on measuring spatial ability first and then to develop a spatial ability test to include spatial relations, spatial visualization, and spatial orientation factors, which can measure the spatial abilities of university students in a valid, reliable, and useful way.

1.1. Literature Review

1.1.1. *Spatial ability and its components*

Spatial ability research started in the late 1800s with studies aimed at demonstrating that spatial ability is a separate factor from general intelligence and continued with studies aimed at identifying and defining the composition of spatial ability (Mohler, 2008). Since the concept of spatial ability was first introduced, many terms, including spatial ability, spatial reasoning, spatial concepts, spatial intelligence, spatial cognition, mental maps, environmental cognition, and cognitive mapping, have been used interchangeably in the spatial ability literature (NRC, 2006), and the term "spatial ability" has been defined in different ways (D'Oliveira, 2004; Martín-Dorta et al., 2008). In the present study, the concept of spatial ability was preferred since it is used more frequently in the field.

Gardner (1983), one of the leading theorists who popularized spatial ability with a different name, namely spatial intelligence, defined it as "the capacities to perceive the visual world accurately, to perform transformations and modifications upon one's initial perceptions, and to be able to re-create aspects of one's visual experience, even in the absence of relevant physical stimuli." (p. 173). Linn and Petersen (1985) defined spatial ability as a “skill in representing, transforming, generating, and recalling symbolic, nonlinguistic information.” (p.1482). Lohman (1993) defined it as “the ability to generate, retain, retrieve, and transform well-structured visual images.” (p.3). According to the National Research Council (2006), "spatial ability" is " a trait that a person has and as a way of characterizing a person’s ability to perform mentally such operations as rotation, perspective change, and so forth." (p.26). Tartre (1990) defined spatial skills as “mental skills concerned with understanding, manipulating, reorganizing, or interpreting relationships visually.” (p.216). According to Carroll (1993), individuals' spatial and other visual abilities refer to “searching the visual field, apprehending the forms, shapes, and positions of objects as visually perceived, forming mental representations of those forms, shapes, and positions, and manipulating such representations mentally." (p.304). Based on these definitions, the present study defines spatial ability as the ability to generate, retain, retrieve, manipulate, interpret, and reorganize the mental representations of visual objects by perceiving their forms and positions.

Psychometric studies on spatial ability have indicated that spatial ability does not have a monolithic structure, but is made up of a composition of factors consisting of sub-skills (Lohman, 1979, 1993; Guilford et al., 1952; Mohler, 2008). D'Oliveira (2004) also reported that one should refer to a domain of spatial abilities instead of a single spatial ability. Lohman (1993) noted that there are several spatial abilities, each focusing on different processes such as generating, retaining, retrieving, and transforming images. On the other hand, there are remarkable discrepancies and confusion in terms of the number and naming of factors in the literature and in terms of the tests used to measure each factor (D'Oliveira, 2004; Martín-Dorta et al., 2008). Relevant literature typically classifies spatial ability under two (Clements, 1998; Guilford et al., 1952; McGee, 1979; Pellegrino et al., 1984), three (Barnea, 2000; Contero et al., 2005; D'Oliveira, 2004; Linn & Petersen, 1985; Lohman, 1979); or five (Carroll, 1993; Maier, 1996) factors. These factors are summarized in Table 1.

Table 1. *The names of the factors classified by different researchers.*

Factors Author(s)	(Spatial) Visualization	Spatial Relations	(Spatial) Orientation	Spatial Perception	Mental Rotation	Perceptual Speed	Closure Speed	Flexibility of Closure
Clements (1998)	✓		✓					
McGee (1979)	✓		✓					
Pellegrino et al. (1984)	✓	✓						
Guilford et al. (1952)	✓	✓						
Linn and Petersen (1985)	✓			✓	✓			
Lohman (1979)	✓	✓	✓					
Barnea (2000)	✓	✓	✓					
Contero et al. (2005)	✓	✓	✓					
D'Oliveira (2004)	✓	✓	✓					
Maier (1996)	✓	✓	✓	✓	✓			
Carroll (1993)	✓	✓				✓	✓	✓

When these classifications are examined, it can be said that the most commonly-mentioned factors of spatial ability in the literature are Spatial Relations (Barnea, 2000; Carroll, 1993; Contero et al., 2005; D'Oliveira, 2004; Guilford et al., 1952; Lohman, 1979; Maier, 1996; Pellegrino et al., 1984), Spatial Visualization (Barnea, 2000; Carroll, 1993; Clements, 1998; Contero et al., 2005; D'Oliveira, 2004; Guilford et al., 1952; Linn & Petersen, 1985; Lohman, 1979; Maier, 1996; McGee, 1979; Pellegrino et al., 1984) and Spatial Orientation (Barnea, 2000; Clements, 1998; Contero et al., 2005; D'Oliveira, 2004; Lohman, 1979; Maier, 1996; McGee, 1979). Therefore, it was decided to include spatial visualization, spatial relations, and spatial orientation factors as components of the spatial ability test developed in this study.

1.1.2. *Measuring spatial ability*

Different definitions of spatial ability and its components have led to the use of many different tests for measuring these abilities (Martín-Dorta et al., 2008). There is a large variety of spatial ability tests, which confuses their names and content (D'Oliveira, 2004).

Table 2 below provides various definitions of spatial visualization, spatial relations, and spatial orientation factors, which allow the readers to examine the definitions more clearly in a comparative manner and the most popular tests measuring the different components of spatial ability.

As can be seen from the definitions above, the confusion in defining spatial ability is also true for its components. To illustrate, both spatial visualization and spatial relations abilities are defined in terms of mental rotation ability. This situation also confuses the measurement of these abilities, as many researchers have stated (e.g., Carroll, 1993; D'Oliveira, 2004; Eliot & Hauptman, 1981). As a result, distinguishing the differences between spatial visualization and spatial relations abilities, as well as the types of items to be used for their measurement, is critical.

Table 2. *Some tests measuring the different components of spatial ability.*

	Components of spatial ability		
	Spatial visualization	Spatial relations	Spatial orientation
Definitions	"the ability to manipulate visual objects mentally.", (Guilford et al., 1952, p.62)	the ability to resolve mental rotation problems quickly (Lohman, 1979).	"the ability to imagine how a stimulus array will appear from another perspective" (Lohman, 1979, p.127).
	"the ability to mentally rotate, manipulate, and twist two- and three-dimensional stimulus objects." (McGee, 1979, p.896)	"the ability to visualize objects in space, when rotated." (Carroll, 1993, p.209)	"understanding and operating on the relationships between the positions of objects in space with respect to one's own position." (Clements & Battista, 1992, p.444)
	"comprehension and performance of imagined movements of objects in two- and three-dimensional space." (Clements & Battista, 1992, p.444)	"the ability to mentally rotate objects in two dimensions" (Contero et al., 2005, p.25)	"understanding and operating on relationships between different positions in space, at first with respect to one's own position and your movement through it, and eventually from a more abstract perspective that includes maps and coordinates at various scales." (Sarama & Clements, 2009, p.161)
	"the ability to understand accurately three-dimensional objects from their two-dimensional representation." (Barnea, 2000, p.308)	"the ability to visualise the effects of operations such as rotation, reflection and inversion, or to mentally manipulate objects." (Barnea, 2000, p.308)	"an ability to perceive spatial patterns or maintain orientation with respect to objects in space." (McGee, 1979, p. 892)
	"the mental manipulation and integration of stimuli consisting of more than one part or movable parts." (Olkun, 2003, p.2)	"the ability to comprehend the spatial configuration of objects or parts of an object and their relation to each other." (Maier, 1996, p.70)	"the ability to orient oneself physically or mentally in space" and it requires "a person's own orientation in any particular spatial situation." (Maier, 1996, p.71)
Tests	Paper Folding Test, Form Board Test, Surface Development Test (Ekstrom et al., 1976)	Flags Test (Thurstone & Thurstone, 1941)	Spatial Orientation Test (Guilford & Zimmerman, 1948)
	Purdue Spatial Visualization Test: Developments (Guay, 1977)	Purdue Spatial Visualization Test: Rotations (Guay, 1977)	Purdue Spatial Visualization Test: Views Test (Guay, 1977)
	Revised Minnesota Paper Form Board Test (Likert & Quasha, 1941)	Card Rotation Test and Cube Comparisons Tests (Ekstrom et al., 1976)	Middle Grades Mathematics Project (MGMP) Spatial Visualization Test (Winter et al., 1989)
	The Embedded Figures Test (Witkin, 1950)	Mental Rotation Tasks (Shepard & Metzler, 1971)	Spatial Orientation: Object Perspective/Map Perspective Tests (Kozhevnikov & Hegarty, 2001)

As an example of this discrepancy, while Olkun (2003, p. 2) defines spatial relations as "imagining the rotations of 2D and 3D objects as a whole body," Burnett and Lane (1980) and Olkun (2003) explain spatial visualization as a holistic and piece-by-piece imagination of the rotations of objects and their parts in 3D space. As can be understood from the definitions, while in the spatial relations ability, 2- and 3-dimensional objects are moved as a whole, in the spatial visualization subtest, the rotation of 3-dimensional objects happens with the whole and its parts. On the other hand, it has been frequently reported that while speed is more important in spatial relations tests, power is more important in spatial visualization test items (Olkun 2003; Pellegrino et al., 1984), problems in spatial relationships tests contain less complex stimuli than spatial visualization problems (Olkun, 2003), and more mental processing and coordination are required to solve spatial visualization problems (Pellegrino et al., 1984). In problems about spatial relations, the students have to find the rotated or twisted version of the original figure from among a group of objects given on a piece of paper (Olkun, 2003; Pellegrino et al., 1984). Pellegrino and Kail (1982) stated that spatial relations tests include problems measuring 2D and 3D mental rotation and cube comparison abilities. The tests measuring spatial visualization include form board problems (Linn & Petersen, 1985; Olkun, 2003; Pellegrino & Kail, 1982), paper folding (Contero et al., 2005; Linn & Petersen, 1985; McGee, 1979; NRC, 2006; Olkun, 2003; Pellegrino & Kail, 1982), and surface development (Contero et al., 2005; Linn & Petersen, 1985; Olkun, 2003; Pellegrino & Kail, 1982).

Lohman (1979) suggests that in a valid spatial orientation test, subjects must imagine being redirected in space and then interpret the situation. Spatial orientation tasks do not require moving an object mentally; only the perceptual perspective of the person viewing the object is changed or moved (Tartre, 1990). Measuring spatial orientation ability is difficult because it requires mental rotation of the stimulus rather than the rotation of the picture itself (Lohman, 1979). Tartre (1990) pointed out that there is no consensus among researchers on the classification of spatial orientation tasks, and stated that spatial ability tasks may involve organizing a visual representation, reorganizing, interpreting, seeing it, or seeing it from a different angle, but by moving the object mentally. Problems used to measure spatial orientation ability include finding directions on a map (Campos & Campos-Juanatey, 2020; Kozhevnikov & Hegarty, 2001), imagining the view of an object from different angles, determining the number of cubes in an object made up of cubes (Winter et al., 1989), finding the view of an object in a cube-shaped glass bell from different angles (Guay, 1977), etc.

1.2. Rationale

There are numerous spatial ability tests referred to in the relevant literature (see Table 2). As a result of the rapid proliferation of spatial ability tests, different researchers have given different names to similar factors or, conversely, the same names were used to describe different factors, which has measured the components of spatial ability even more complicated (Eliot & Hauptman, 1981). These confusions also affected the results of the factor analysis studies conducted to determine test structures. In a study, Carroll (1993) re-analyzed factor analytic studies in the literature and found that items are not always consistently loaded on relevant factors due to considerable confusion in the identification of factors. D'Oliveira (2004) also argued that variations in the format of the tests and specific administration procedures can be responsible for inconsistent results. Thus, D'Oliveira (2004) especially emphasized the importance of clarifying the spatial factor or capability covered by the test items, regardless of the names assigned to test items (or tasks) by previous researchers (Carroll, 1993). Due to this confusion in the literature, in this study, firstly, the definitions for the factors of spatial ability and question types were examined in detail and differentiated. Eventually, it was aimed to develop a spatial ability test in which question types and factors are clearly defined and statistically tested. The current test is planned to cover the three factors of spatial ability (spatial relations, spatial visualization, and spatial orientation) most commonly mentioned in the relevant literature. It is intended to expand the scope of the test by involving different types of problems in each factor.

Eliot and Hauptman (1981) pointed out that items can yield different factor loadings in different samples. This situation reveals the importance of developing the test in accordance with the characteristics of a particular group and testing its psychometric properties. What makes this test distinct from its antecedents is that while usually the same tests are used for different groups (Bakker, 2008; Battista et al., 1982; Ekstrom et al., 1976; Guay, 1977; Hegarty & Waller, 2004; Kozhevnikov & Hegarty, 2001; Lord, 1985; Sorby & Baartmans, 2000), the present test was developed specifically for university students studying at different programs/departments or candidate university students who plan to study programs that require spatial capability and to assess their professional competencies. In addition, several studies (e.g., Kim & Irizarry, 2021; Olkun et al., 2009; Patkin & Dayan, 2013; Sisman et al., 2021) have put forward the idea that spatial ability can be improved through well-designed training programs. In this context, it will be possible to accurately measure the spatial ability development among students and reveal the effect of the training only by using a valid and reliable spatial ability test.

On the other hand, the training applied may reveal different effects on the level of spatial ability in different cultures. For example, Turgut and Nagy-Kondor (2013) found a significant difference between the spatial visualization scores of Hungarian and Turkish pre-service mathematics teachers, favoring the former. Olkun et al. (2009) compared the initial spatial skills of primary school teacher candidates in four countries, i.e., Taiwan, Finland, the United States, and Türkiye, and evaluated the development of these skills through interactive computer programs. As a result, it was seen that the spatial visualization scores were the highest among Finnish students, followed by Taiwanese students, and the scores of the American and Turkish students were very close to each other. However, the researchers pointed out that while students from two eastern countries, Türkiye and Taiwan, made progress after the implementation, students from the USA and Finland did not make sufficient progress. Researchers stated that this situation may be due to cultural differences and suggested that the reason why spatial education is more successful in Taiwan and Türkiye than in the USA and Finland is that the former countries have relatively more formal class cultures. It is noteworthy that the spatial ability levels of Turkish students were reported as rather low in both of the abovementioned comparative studies. This reveals the importance of researching the spatial abilities of Turkish students. According to the literature review studies examining the tendency of spatial ability studies in Türkiye (Dokumacı Sütçü, 2021; İpekoğlu et al., 2020; Ozcakir Sumen, 2019), most of the studies were conducted with secondary school students and the effect of a particular teaching method (mostly computer-assisted teaching) on spatial ability was investigated. Since the transition to Piaget's formal operations stage coincides with the secondary school level, it is very important to focus on the development of students' spatial abilities during this period. However, since the spatial ability is very important in many professions, it is thought that the development of a spatial ability test to be used to measure the spatial ability levels of the students who are to get professional education at the university will be useful for researchers, educators, and curriculum developers.

The most commonly used tests were developed in the 1970s, and there are concerns about their psychometric properties since they have been administered to a wide range of different groups. In Türkiye, Purdue Spatial Visualization Test developed by Guay (1977) and the MGMP Spatial Ability Test developed by Michigan State University mathematics department faculty members (1983) were generally used to measure students' spatial visualization skills (İpekoğlu et al., 2020). However, the adaptation of the tests used in Turkish culture mainly concentrated on textual translation, and the equivalence of the tests in Turkish culture and their psychometric properties at the applied level were not adequately examined (Sevimli, 2009). Therefore, the present study is also promising because a more comprehensive test development procedure has been followed following a sequential exploratory mixed methods research design (qualitative followed by quantitative phases) and psychometric properties were tested through comprehensive analysis (item analysis, confirmatory factor analysis, reliability analysis, and the difference between 27% of the lower and upper groups).

2. METHOD

2.1. Design

The sequential exploratory mixed method was used to design this study, which aimed to develop a spatial ability test. The sequential exploratory mixed method is a common way of developing quantitative instruments, wherein in the first stage, the researcher starts to explore the subject using qualitative methods and then continues to validate the instrument using quantitative methods based on the themes from the first stage (Creswell & Plano Clark, 2011). In this study, qualitative methods (literature review, expert opinions, and student opinions) were used to develop the initial test form; and quantitative methods were used to test the psychometric properties regarding content validity, construct validity, and reliability.

2.2. Study Group

In the qualitative stage, 10 students (Female= 7, Male= 3) studying an elementary school mathematics education program were consulted to check the clarity, comprehensibility, and suitability of the draft test items. In addition, an expert panel consisting of 8 scholars (4 mathematics experts, 3 mathematics education experts, and one measurement and evaluation specialist) was consulted for their opinions about the content and face validity of the draft test.

In the quantitative stage, the validity and reliability studies of the test were conducted with a total of 456 university students (58% female), studying at different departments/programs of Malatya İnönü University, a state university located in eastern Türkiye. Participants were chosen from departments/programs where either recruitment or studying is considered to be facilitated by possessing good spatial abilities. Accordingly, participants involved 29 students (38% female) from the Graphic Design program, 61 students (7% female) from the Civil Engineering Department, 47 students (21% female) from the Mechanical Engineering Department, 266 students (77% female) from the Elementary School Mathematics Education Program, 39 students (67% female) from the Landscape Architecture Department, and 14 students (50% female) from the Art Teaching Program.

2.3. Procedure

In the development process of the spatial ability test, the stages of test development were followed, which included: 1) determining the purpose of the test; 2) determining the scope of the test, 3) determining test properties and writing items, 4) validity and reliability studies, and 5) preparing a guide for the test. Accordingly, first, an overall plan regarding the test development process was prepared by the researchers, which was then evaluated by a measurement and evaluation specialist. The plan was revised in accordance with the experts' opinions and put into practice as described below:

2.3.1. Determining the purpose of the test

The purpose of the spatial ability test is to measure the spatial ability levels of university students in a valid and reliable way.

2.3.2. Determining the scope of the test

Downing (2006) emphasized that determining the content of the test is one of the most important tasks at the earliest stages of the test development process. Due to the critical importance of the scope of the spatial ability test, we set out with a detailed literature review first. As a result of the comprehensive literature review, it was seen that spatial ability has a multifactorial structure, and different researchers explain spatial ability under different factors (D'Oliveira, 2004; Lohman, 1993; Mohler, 2008). Since it would not be possible to include all factors of spatial ability mentioned in the studies in terms of usefulness, reliability, and content validity, it was decided to include three domains of spatial ability most commonly referred to in the literature: *spatial relations, spatial visualization, and spatial orientation.* After deciding on the factors to be included in the test, a second literature review was conducted to examine the conceptual and operational (how they are measured) definitions of these factors. The definitions of the factors of the spatial ability test developed in this study are presented in Table 3.

Table 3. *The definitions of the factors of spatial ability test.*

Factor	Definition
Spatial Ability	the ability to generate, retain, retrieve, manipulate, interpret, reorganize the mental representations of visual objects by perceiving their forms and positions (Carroll, 1993; Linn & Petersen, 1985; Lohman, 1993; NRC, 2006; Tartre, 1990).
Spatial Relations	the ability to mentally manipulate 2D and 3D objects as a whole with processes such as rotation, reflection, and inversion (Barnea, 2000; Carroll, 1993; Contero et al., 2005; Olkun, 2003).
Spatial Visualization	the ability to mentally rotate, manipulate, and twist a 3-dimensional object composed of more than one part or movable parts in a holistic and piece-by-piece (Burnett & Lane, 1980; McGee, 1979; Olkun, 2003).
Spatial Orientation	the ability to understand the relations between the positions of objects in space relative to one's own position (Clements & Battista, 1992; Sarama & Clements, 2009) and to imagine how an object will look in space from a different perspective by mentally orienting oneself (Barnea, 2000; Contero et al., 2005; Lohman, 1979; Maier, 1996; McGee, 1979).

2.3.3. Determining test properties and writing items

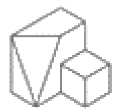
In this study, spatial ability test items were planned to be developed in a multiple-choice format with 5 options. A total of 38 original test items in different problem types were developed by the researchers. In addition, 2 items about rotating 3D objects (Item 12, Item 13) in the draft spatial relations subtest and 3 items unfolding 3D objects (Item 27, Item 28, Item 29) in the draft spatial visualization subtest were driven from Guay's (1977) Purdue Spatial Visualization Test: Rotations and Purdue Spatial Visualization Test: Developments tests, respectively. As a result, an item pool of 43 items was developed, including 17 items in the spatial relations subtest, 12 items in the spatial visualization subtest, and 14 items in the spatial orientation subtest as can be seen in Table 4.

Table 4. *Components, categories, and numbers of items.*

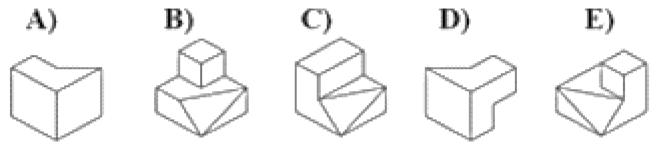
Component	Category	Item No	Number of items
Spatial Relations	Card rotation	1, 2, 3	3
	Rotating the 2D figures and symmetry	4, 5, 6, 7, 8	5
	Rotating the 3D figures	9, 10, 11, 12, 13	5
	Comparing cubes	14, 15, 16, 17	4
Spatial Visualization	Unfolding cubes	18, 19, 20, 21	4
	Cutting paper	22, 23, 24	3
	Unfolding 3D objects	25, 26, 27, 28, 29	5
Spatial Orientation	The view of an object made up of cubes from different angles	30, 31, 32, 33, 37, 38, 39, 40	8
	Number of cubes	34, 35, 36	3
	The view of an object in a cube-shaped glass bell from different angles	41, 42, 43	3

Examples of items are presented below.

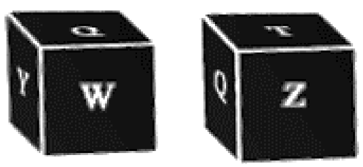
Spatial Relation:



10. Which of the following is the rotated form of the object on the left?



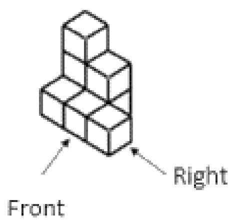
Spatial Visualization:



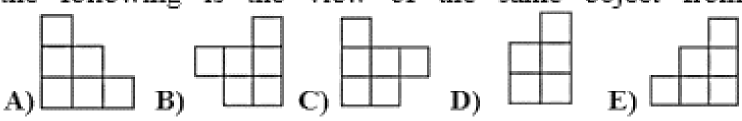
20. Images of the same cube in different positions are presented on the left labeled with letters X, Y, T, Z, Q, and W. Which letter is located opposite the letter X?

A) Y B) Q C)W D) Z E) T

Spatial Orientation:



30. On the left is the FRONT-RIGHT corner view of an image. Which of the following is the view of the same object from the REAR?



2.3.4. Validity and reliability studies

An expert panel consisting of 4 mathematics experts, 3 mathematics education experts, and 1 measurement and evaluation specialist was asked to evaluate the content and face validity of the test. The evaluation criteria included scientific accuracy, comprehensibility and responsiveness of the question roots and options, and the suitability of the figures. The experts evaluated each item using the 4-point scale offered by Davis (1992): 4-Highly relevant, 3-Quite relevant, 2-Somewhat relevant, and 1-Not relevant. The criteria for the Content Validity Index (CVI), which is computed as the number of experts rating an item either 3 or 4, divided by the total number of experts, is set to a minimum of 0.80 (Davis, 1992). Based on the expert ratings, CVIs for all items were found to satisfy the minimum criteria of 0.80. In addition, the revision suggestions from the experts were done and the draft test form was developed with 43 items. Further, to assess the clarity, understandability, and appropriateness of the test form to the target audience, within the scope of the think-aloud protocol, another 10 prospective primary school mathematics teachers were asked to take the test and verbally express their mental processes while solving each test item (Irwing et al., 2018). This way the test items were checked to ensure whether they can measure the constructs which they were actually meant to test. To ensure the reliability and validity of the test results, the figures and the question roots were checked for readability during the preparation and printing of the booklets (Downing, 2006).

Next, the test was applied to 456 university students to examine the item and test statistics. To test the construct validity of the instrument, item difficulty index, item discrimination index, and item-total correlation coefficients were calculated, and the significance of *t* values regarding the differences between 27% lower and upper groups were examined. In addition, a second-order confirmatory factor analysis was performed to test the 3- factor (spatial relations, spatial visualization, and spatial orientation) construct of the spatial ability test. The reliability of the scores obtained from the test was calculated using KR-20 and Split-Half (odd-even) with Spearman-Brown reliability coefficients.

2.3.5. Preparation of a guide for test users

It is planned to provide users with information about the application of the test through a guide, which specifies the purpose of the test, its theoretical background, scoring procedures, and descriptive statistics at the end of the study.

2.4. Data Analysis

The analysis of the data obtained from 456 participants was made via the Test Analysis Program (TAP) (Brooks & Johanson, 2003), SPSS 22, and Lisrel software programs. Correct and incorrect or blank answers were scored 1-0, respectively. The skewness and kurtosis coefficients for the data set were estimated at 0.363 and 0.322, respectively. Since the skewness and kurtosis coefficients were within the acceptable range, it was understood that the data set comes from a normal distribution. While item difficulty refers to the percentage or probability (P) of test takers who answer the item correctly (Ebel & Frisbie, 1991; Hingorjo & Jaleel, 2012; Wendler & Walker, 2006), item discrimination is the tendency of an item to be answered correctly by test takers who are strong in terms of the skill or knowledge intended to be measured and to be answered incorrectly by test takers who are not strong in this respect (Livingston, 2006). The item difficulty indices were kept in the 0.30 to 0.70 range, with fewer items in the easier or more difficult ranges, because in large-scale standardized tests, test taker levels are typically assumed to be normally distributed, and items in the middle range of difficulty have the most variance and the greatest potential to discriminate test takers (Bandalos, 2018, p. 122). Hambleton and Jirka (2006) categorized values around 0.25 as "difficult," values around 0.50 as "moderate," and values around 0.75 as “easy” in terms of item difficulty. Items with a discrimination index of 0.40 or higher were considered very good; items with a discrimination index of 0.30 to 0.39 were considered reasonably good but could be developed; items with a discrimination index of 0.20 to 0.29 were considered poorly discriminative but could be corrected or improved; and items with a discrimination index of 0.19 or lower were considered very poor and could not be corrected or improved. The ideal item-total correlation coefficient was set to a minimum of 0.30 (Wendler & Walker, 2006). In addition, to estimate how discriminative the individual test items are, the differences between the scores of the 27% upper and lower groups were compared using independent samples t-test since the scores were close to the normal distribution (skewness and kurtosis values ± 2 (Cameron, 2004)). The significance level was set to $p < 0.05/27 = 0.002$ (n= 27 t-tests for differences between the scores of the 27% lower and upper groups) with a Bonferroni correction (Abdi, 2010).

3. FINDINGS

3.1. Findings about the Item Analysis of the Spatial Ability Test

The construct validity of the Spatial Ability Test was tested through item analysis. Accordingly, item difficulty indices, item discrimination indices, and item-total correlation coefficients calculated for the preliminary spatial ability test consisting of 43 items are presented in Table 5.

Table 5. *Results of item analysis.*

Item no	Item difficulty(P)	Item discrimination (d)	Item-total correlation (r)
1	0.68	0.39	0.36
2	0.73	0.47	0.43
3	0.38	0.34	0.32
4	0.68	0.54	0.45
5	0.68	0.41	0.36
6	0.50	0.28	0.27
7	0.44	0.37	0.37
8	0.57	0.46	0.37
9	0.51	0.35	0.32
10	0.72	0.46	0.43
11	0.48	0.36	0.33
12	0.35	0.35	0.32
13	0.38	0.30	0.26
14	0.40	0.31	0.27
15	0.38	0.21	0.19
16	0.61	0.42	0.37
17	0.54	0.56	0.44
18	0.45	0.32	0.31
19	0.23	0.06	0.07
20	0.51	0.52	0.43
21	0.20	0.12	0.14
22	0.41	0.33	0.32
23	0.30	0.38	0.41
24	0.20	0.21	0.24
25	0.24	0.21	0.24
26	0.29	0.13	0.07
27	0.31	0.15	0.14

Item no	Item difficulty(P)	Item discrimination (d)	Item-total correlation (r)
28	0.16	0.20	0.29
29	0.20	0.24	0.30
30	0.48	0.43	0.35
31	0.19	0.16	0.18
32	0.18	0.05	0.17
33	0.72	0.49	0.43
34	0.37	0.42	0.37
35	0.34	0.34	0.29
36	0.31	0.15	0.12
37	0.26	0.19	0.17
38	0.31	0.27	0.25
39	0.29	0.38	0.34
40	0.29	0.33	0.31
41	0.47	0.47	0.36
42	0.44	0.34	0.33
43	0.38	0.45	0.38

In Table 5, it was decided to exclude 19, 21, 26, 27, 31, 32, 36, and 37 items with a discrimination index below 0.20. To decide whether items with an item discrimination index between 0.20-0.40 to be corrected or excluded from the test, their item-total correlations were examined. Accordingly, it was decided to exclude items 6, 13, 14, 15, 24, 25, 35, and 38 with item-total correlations below 0.30. Table 4 presents components, categories, and numbers of items. During the item analysis, attention was paid to retain at least 2 items in each category in the final test so that the content validity was not impaired. Thus, despite their relatively low item discrimination indices, items 28 (d= 0.20) and 29 (d= 0.24) were decided to be kept in the test to ensure content validity. Starting from the item with the lowest discrimination index and item-total correlation value, the problematic items were removed successively and the item analysis was repeated. Item analysis results for the final test are presented in Table 6.

According to Table 6, the difficulty indices of the items in the final test ranged between 0.16 and 0.73 (mean = 0.475). Accordingly, 4 items in the final test were difficult (items 28, 29, 39, and 40), 3 items were easy (items 2, 10, and 33), and the remaining 20 items were moderate in terms of difficulty. Discrimination indices of the items ranged between 0.23 and 0.55 (mean = 0.421), and the item-total correlation coefficient ranged between 0.28 and 0.47 (mean = 0.381). In the final test, items 1, 2, 3, 4, 5, 7, 8, 9, 10, 11, 12, 16, and 17 measure spatial relations ability, 18, 20, 22, 23, 28, and 29 items measure spatial visualization ability, and items 30, 33, 34, 39, 40, 41, 42, 43 measure spatial orientation ability. The average difficulty index of the 27 items in the Spatial Ability test was 0.475; while the average discrimination index was 0.421 and the average item-total correlation was 0.381.

Table 6. *Results of item analysis and descriptive analysis for the items in the final test.*

Item no	Item difficulty (P)	Item discrimination (d)	Item-total correlation (r)
1	0.68	0.43	0.38
2	0.73	0.49	0.47
3	0.38	0.42	0.37
4	0.68	0.51	0.44
5	0.68	0.42	0.37
7	0.44	0.42	0.37
8	0.57	0.48	0.41
9	0.51	0.38	0.34
10	0.72	0.45	0.45
11	0.48	0.43	0.36
12	0.35	0.37	0.34
16	0.61	0.44	0.38
17	0.54	0.53	0.45
18	0.45	0.38	0.35
20	0.51	0.55	0.44
22	0.41	0.33	0.33
23	0.30	0.45	0.41
28	0.16	0.27	0.30
29	0.20	0.23	0.28
30	0.48	0.43	0.38
33	0.72	0.47	0.43
34	0.37	0.43	0.40
39	0.29	0.33	0.34
40	0.29	0.36	0.36
41	0.47	0.43	0.36
42	0.44	0.44	0.37
43	0.38	0.50	0.41

3.2. Differences between 27% lower and upper group scores

Another method used to test the construct validity of the test through the discrimination potential of the items is to compare, for each item, the average scores from 27% lower and upper groups using the independent t-test. The results of the independent samples t-test regarding the comparison of the averages of the 27% lower group ($n = 144$) and 27% upper group ($n = 131$) are presented in Table 7.

When Table 7 is examined, statistically significant differences were found between the lower and upper groups of 27% for all items ($p <0.002$). Therefore, in addition to item discrimination and item-total correlation coefficient analyses, it was proven once again that each item is able to significantly distinguish between the upper group with the highest spatial ability and the lower group with the lowest spatial ability.

Table 7. *t-test results regarding the significance of the differences between the scores of the lower and upper groups (27%).*

Item no	Group	Mean	Std. Deviation	*t*	*df*	*p*
Item 1	Lower 27%	0.46	0.50	-8.516	245.781	.000*
	Upper 27%	0.89	0.32			
Item 2	Lower 27%	0.46	0.50	-10.591	203.058	.000*
	Upper 27%	0.95	0.23			
Item 3	Lower 27%	0.16	0.37	-7.928	238.462	.000*
	Upper 27%	0.58	0.50			
Item 4	Lower 27%	0.42	0.50	-10.707	223.761	.000*
	Upper 27%	0.92	0.27			
Item 5	Lower 27%	0.47	0.50	-8.457	241.682	.000*
	Upper 27%	0.89	0.31			
Item 7	Lower 27%	0.19	0.39	-7.708	248.363	.000*
	Upper 27%	0.60	0.49			
Item 8	Lower 27%	0.30	0.46	-9.087	272.999	.000*
	Upper 27%	0.78	0.42			
Item 9	Lower 27%	0.32	0.47	-6.845	271.439	.000*
	Upper 27%	0.70	0.46			
Item 10	Lower 27%	0.47	0.50	-9.444	222.183	.000*
	Upper 27%	0.92	0.27			
Item 11	Lower 27%	0.26	0.44	-7.894	273	.000*
	Upper 27%	0.69	0.46			
Item 12	Lower 27%	0.16	0.37	-7.014	236.955	.000*
	Upper 27%	0.53	0.50			
Item 16	Lower 27%	0.40	0.49	-8.530	263.698	.000*
	Upper 27%	0.84	0.37			
Item 17	Lower 27%	0.26	0.44	-10.295	273	.000*
	Upper 27%	0.79	0.41			
Item 18	Lower 27%	0.27	0.45	-6.753	265.644	.000*
	Upper 27%	0.65	0.48			
Item 20	Lower 27%	0.26	0.44	-10.806	272.897	.000*
	Upper 27%	0.81	0.39			
Item 22	Lower 27%	0.29	0.46	-5.721	266.061	.000*
	Upper 27%	0.62	0.49			
Item 23	Lower 27%	0.12	0.32	-8.894	220.074	.000*
	Upper 27%	0.57	0.50			
Item 28	Lower 27%	0.04	0.20	-6.171	173.113	.000*
	Upper 27%	0.31	0.47			
Item 29	Lower 27%	0.11	0.32	-4.719	222.117	.000*
	Upper 27%	0.34	0.48			
Item 30	Lower 27%	0.27	0.45	-7.902	273	.000*
	Upper 27%	0.70	0.46			
Item 33	Lower 27%	0.47	0.50	-10.142	209.860	.000*
	Upper 27%	0.94	0.24			
Item 34	Lower 27%	0.19	0.39	-8.027	249.299	.000*

Item no	Group	Mean	Std. Deviation	*t*	*df*	*p*
	Upper 27%	0.62	0.49			
Item 39	Lower 27%	0.14	0.35	-6.231	228.730	.000*
	Upper 27%	0.47	0.50			
Item 40	Lower 27%	0.14	0.35	-6.947	228.396	.000*
	Upper 27%	0.50	0.50			
Item 41	Lower 27%	0.28	0.45	-7.912	273	.000*
	Upper 27%	0.71	0.46			
Item 42	Lower 27%	0.24	0.43	-8.180	263.595	.000*
	Upper 27%	0.68	0.47			
Item 43	Lower 27%	0.15	0.36	-8.180	263.595	.000*
	Upper 27%	0.66	0.48			

*$p<0.002$ (with a Bonferroni correction of 0.05/27=0.002)

3.3. Second-order Confirmatory Factor Analysis

Second-order confirmatory factor analysis (CFA) was performed to test the three-factor construct (spatial visualization, spatial relations, and spatial orientation) of the 27-item spatial ability test. The model was estimated using the asymptotic covariance matrix and analyzed using the Diagonally Weighted Least Squares method in the Lisrel software program (Jöreskog & Sörbom, 1993; Kline, 2011). While evaluating CFA results, the goodness of fit indices was considered excellent when $\chi^2/df \leq 2$; GFI, AGFI, CFI, IFI, NNFI ≥ 0.95; RMSEA, SRMR ≤ 0.05, and they were indicated acceptable when $\chi^2/df \leq 5$; GFI, AGFI, CFI, IFI, NNFI ≥ 0.90; SRMR, RMSEA ≤ 0.08 (e.g. Brown, 2006; Hair et al., 2014; Hu & Bentler, 1999; Jöreskog & Sörbom, 1996; Tabachnick & Fidell, 2013). As a result of the first analysis, the goodness of fit indices was estimated $\chi^2/df = 1.39$ (445.99/321), RMSEA =0.029, SRMR = 0.075, GFI = 0.96, AGFI = 0.95, CFI = 0.98, IFI = 0.98, NNFI = 0.98. According to these values, χ^2/df, RMSEA, GFI, AGFI, CFI, IFI NNFI values were excellent, SRMR value indicated acceptable. Standardized factor loadings ranged between 0.60 and 0.86 in the spatial relations subtest, between 0.69 and 0.88 in the spatial visualization subtest, and between 0.59 and 0.82 in the spatial orientation subtest. These results suggested that the three-factor construct of the spatial ability test is confirmed and a total Spatial Ability score can be calculated for all 27 items. The path diagram is presented in Figure 1.

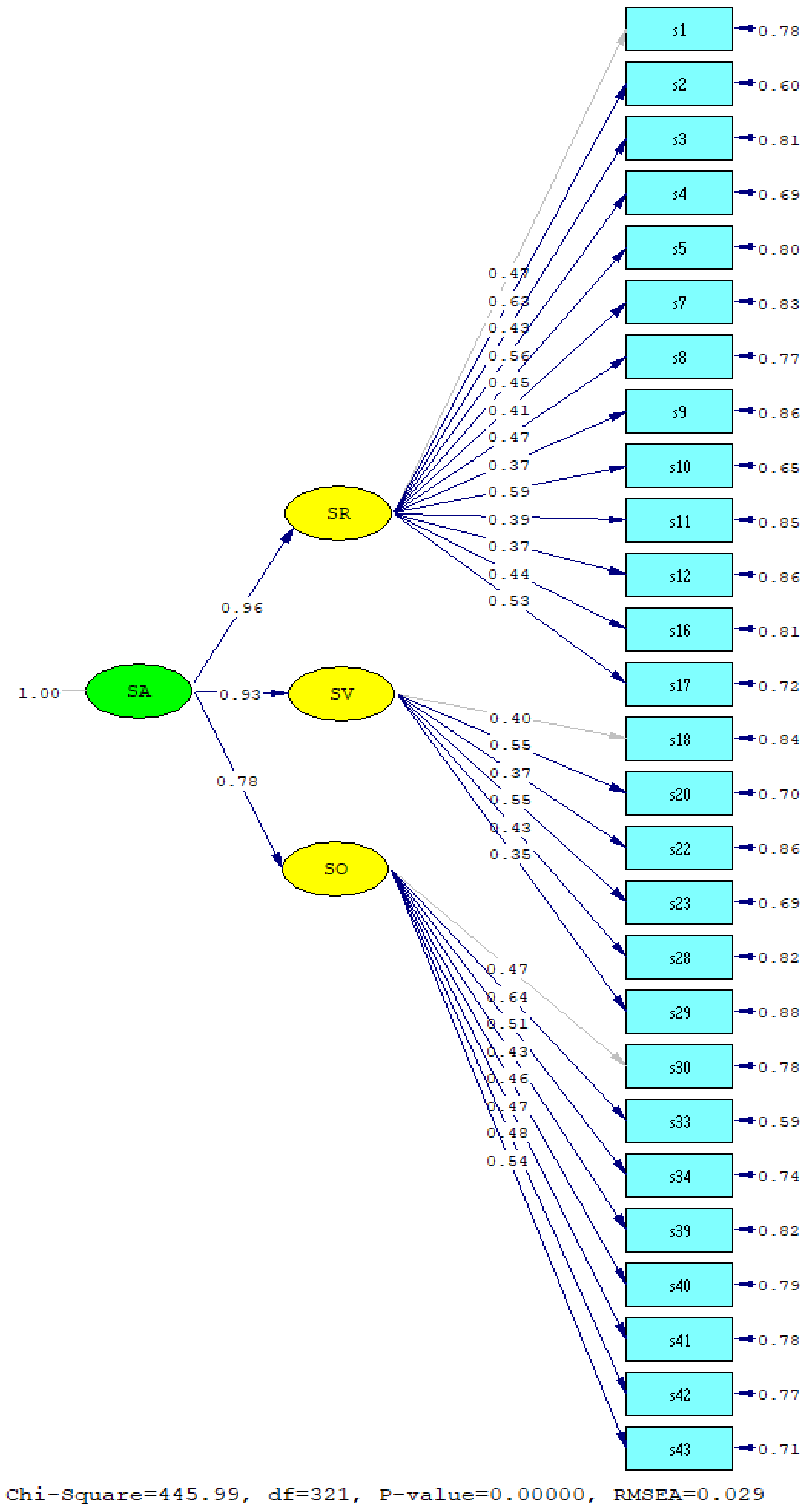
3.4. Reliability Analysis

The KR-20 reliability coefficient of the total test was estimated 0.775, and Split-Half (odd-even) with Spearman-Brown was estimated 0.798 suggesting acceptable internal consistency. Wells and Wollack (2003) put that the minimum value of the reliability coefficient is expected to be 0.70. Based on this reference value, it can be said that the reliability coefficients of the test are sufficient for the whole test.

3.5. The Guide for Test Users

The Spatial Ability Test is a multiple-choice test to measure university students' spatial abilities. The final test has three sub-tests with 27 items each offering 5 options: 13 items in the spatial relations subtest (items 1-13), 6 items in the spatial visualization subtest (items 14-19), and 8 items in the spatial orientation subtest (items 20-27). Correct and incorrect/blank answers are scored 1-0 respectively. For the test, a student has a minimum possible score of 0 and a maximum possible score of 27. The high scores are indicative of good level spatial ability, whilst low scores are indicative of low spatial ability. For average scores, 0-9.00 points can be interpreted as low, 9.01-18.00 points as medium, and 18.01-27.00 points as good spatial ability skills. The mean spatial ability score for the participants of this study was 12.85 (s= 4.89), indicating medium level of spatial ability.

Figure 1. *Path diagram of the model.*



Note. SA: Spatial Ability; SR: Spatial Relations; SV: Spatial Visualization; SO: Spatial Orientation

4. DISCUSSION and CONCLUSION

This study aimed to develop a useful test to measure university students' spatial abilities in a valid and reliable way. To develop a comprehensive and focused instrument, it was planned to develop items related to spatial relations, spatial visualization, and spatial orientation abilities as the subtests of spatial ability. The validity studies of the spatial ability test were carried out in detail and meticulously, and evidence for three criteria was collected to determine the validity of the test: content validity, face validity, and construct validity.

Prior to the quantitative pilot study, the opinions of experts evaluated according to the Davis (1992) technique proved the adequate level of content and face validity. In the process of testing the construct validity, item analysis was performed first. Crocker and Algina (2008) suggested that when developing a test, it was aimed to produce a final test including an optimum number of items, which meet the required reliability and validity criteria. Therefore, taking into account the usefulness of the test, it was aimed to develop a valid and reliable test with a minimum number of items while preserving the content validity. Accordingly, during item analysis, items with poor discrimination indexes and item-total correlation coefficients were successively excluded from the test, and a final test form with 27 items was obtained. The final test included 13 items in the spatial relations ability subtest, 6 items in the spatial visualization ability subtest, and 8 items in the spatial orientation ability subtest.

When the average discrimination indices for the spatial ability test and its subtests are examined, it can be said that the test as a whole and its subtests are highly discriminative (Ebel & Frisbie, 1991; Wells & Wollack, 2003). The average item-total correlation coefficients for the spatial ability test and its subtests indicated the adequacy of discrimination and internal consistency (Wendler & Walker, 2006). Moreover, we found statistically significant differences between the 27% lower and upper group scores for each item, which provided additional evidence for the existence of the items' discrimination, as the difference between the upper and lower groups of 27% reveals a more sensitive and stable item discrimination index about the test items (Crocker & Algina, 2008; Diederich, 1973).

The average difficulty index of the test is moderate (Hambleton & Jirka, 2006). Hingorjo and Jaleel (2012) point out that item difficulty and item discrimination indexes are generally interrelated. Similarly, it is well known that test developers should avoid including items that are answered correctly or incorrectly by the majority of students since such items would have standard deviations close to zero and cannot distinguish students with different ability levels (Crocker & Algina, 2008; Wells & Wollack, 2003; Wendler & Walker, 2006). Since the variance would be maximum when the item difficulty is 0.50, it has been suggested that most of the test items should be a moderate difficulty (around 0.50) to discriminate well between people with a wide range of abilities (Crocker & Algina, 2008; Gronlund, 1977; Wendler & Walker, 2006). The average difficulty level of the items in the final spatial ability test developed in this study was also around 0.50. Accordingly, it can be said that the average difficulty of the test increases the variance and contributes to the potential of the test to distinguish individuals with high and low spatial abilities. When the difficulty of the subtests is examined, it can be said that the average difficulty values of the spatial relations and spatial orientation tests are closer to moderate difficulty; however, the average difficulty of the spatial visualization test indicated a rather difficult test. Several studies (Linn & Petersen, 1985; Lohman, 1979; Olkun, 2003; Pellegrino et al., 1984) report that spatial visualization is more complex than other subskills. The result obtained in this study regarding the difficulty of the spatial visualization test compared to other subtests is in line with the literature. Thus, in this study, it can be said that the construct validity of the test was provided according to the results obtained from the item analysis of the test.

As a part of construct validity studies, the three-factor construct (spatial relations, spatial visualization, and spatial orientation) of the 27-item final test was examined with a second-order CFA. According to the widely accepted goodness of fit criteria (e.g., Brown, 2006; Hair et al., 2014; Tabachnick & Fidell, 2013), the goodness of fit indices for the three-factor construct were determined to be perfect, except for SRMR, which is also acceptable. Also, the standardized factor loadings were estimated between 0.59 and 0.88 and significant for all items. Brown (2006) stated that factor loadings greater than or equal to 0.30 or 0.40 are regarded as salient. The result of reliability analysis through KR-20 and Split-Half (odd-even) with Spearman-Brown coefficients proved to be favorable (Wells & Wollack, 2003). As a result, it can be said that the 27-item test is useful, valid, and reliable for measuring the spatial abilities of university students.

5. LIMITATIONS and FUTURE DIRECTIONS

The participants of this research are restricted to 456 students studying at the departments/programs of graphic design, civil engineering, mechanical engineering, elementary mathematics education, landscape architecture, and art education at a state university in eastern Türkiye. Therefore, the results of the research may not be practically generalized to students studying in all departments of the university and different levels (e.g, high school, graduate). In addition, the study may have shown context-based results and may limit the generalization of the results to other regions. Thus, the psychometric properties can be tested further with university students studying in different departments, with different levels (e.g, high school, graduate), or different regions of Türkiye. In addition, the participants in this study were recruited using the purposive and convenience sampling method. The psychometric properties of the test can be tested on a group determined by random assignment.

Considering the abovementioned limitations, the spatial ability test developed here can be used by high school guidance services to measure students' or candidates' spatial abilities to predict their potential for tertiary programs requiring such abilities as dentistry, medicine, architecture, engineering, navigation, mathematics, art, graphic design, etc. The same is also true for any admission committees that plan to measure candidates' spatial abilities for employment or program admission purposes. Moreover, the spatial ability test developed in this study can be used as a pre-posttest to test the effect of the potential intervention programs aiming at improving learners' spatial ability.

Acknowledgments

The findings of the research were presented as an oral paper at International EJER2020 Congress, held on 10-13 September at Eskişehir Anadolu University, Turkiye.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. Scientific and ethical principles were complied. It has been confirmed by İnönü University Social and Human Sciences Scientific Research Ethics Committee (24/03/2022- Ethics Committee Number: 2022/6-24). The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors.

Authorship Contribution Statement

Kubra Acikgul: Literature review, conceptualization, preparation of the item pool, data collecting, validity and reliability studies, writing- original draft preparation. **Suleyman Nihat Sad:** Literature review, conceptualization, validity and reliability studies, writing- original draft preparation, reviewing and editing. **Bilal Altay:** Preparation of the item pool, data collecting.

Orcid

Kubra Acikgul https://orcid.org/0000-0003-2656-8916
Suleyman Nihat Sad https://orcid.org/0000-0002-3169-2375
Bilal Altay https://orcid.org/0000-0002-2400-7122

REFERENCES

Abdi, H. (2010). Holm's sequential Bonferroni procedure. In N. Salkind (Eds.), *Encyclopedia of research design* (pp. 1-8). Sage Publication.

Bakker, M. (2008). *Spatial ability in primary school: Effects of the Tridio® learning material* [Master's thesis]. University of Twente.

Bandalos, D.L. (2018). *Measurement theory and applications for the social sciences* (1st ed.). Guilford Publications.

Battista, M.T., Wheatley, G.H., & Talsma, G. (1982). The importance of spatial visualization and cognitive development for geometry learning in preservice elementary teachers. *Journal for Research in Mathematics Education*, *13*(5), 332-340. https://doi.org/10.5951/jresematheduc.13.5.0332

Barnea, N. (2000). Teaching and Learning about Chemistry and Modelling with a Computer managed Modelling System. In Gilbert J.K., Boulter C.J. (Eds.), *Developing Models in Science Education* (pp. 307-323). Springer. https://doi.org/10.1007/978-94-010-0876-1_16

Brooks, G.P., & Johanson, G.A. (2003). TAP: Test Analysis Program. *Applied Psychological Measurement*, *27*(4), 303-304. https://doi.org/10.1177/0146621603027004007

Brown, T. A. (2006). *Confirmatory factor analysis for applied research* (1st ed.). The Guilford Press.

Burnett, S.A., & Lane, D.M. (1980). Effects of academic instruction on spatial visualization. *Intelligence*, *4*(3), 233-242. https://doi.org/10.1016/0160-2896(80)90021-5

Cameron, A. (2004). Kurtosis. In M. Lewis-Beck, A. Bryman and T. Liao (Eds.). *Encyclopedia of social science research methods* (pp. 544-545). SAGE Publications, Inc.

Campos, A., & Campos-Juanatey, D. (2020). Measure of spatial orientation ability. *Imagination, Cognition and Personality*, *39*(4), 348-357. https://doi.org/10.1177/0276236619896268

Carroll, J.B. (1993). *Human cognitive abilities: A survey of factor-analytic studies* (No. 1). Cambridge University Press.

Clements, D.H. (1998). *Geometric and spatial thinking in young children*. National Science Foundation.

Clements, D.H., & Battista, M.T. (1992). Geometry and spatial reasoning. In D.A. Grouws (Eds.), *Handbook of research on mathematics teaching and learning* (pp. 420-464). Macmillan.

Contero, M., Naya, F., Company, P., Saorín, J.L., & Conesa, J. (2005). Improving visualization skills in engineering education. *IEEE Computer Graphics and Applications*, *25*(5), 24-31. https://doi.org/10.1109/MCG.2005.107

Contreras, M.J., Escrig, R., Prieto, G., & Elosúa, M.R. (2018). Spatial visualization ability improves with and without studying Technical Drawing. *Cognitive processing*, *19*(3), 387-397. https://doi.org/10.1007/s10339-018-0859-4

Creswell, J.W., & Plano Clark, V.L. (2011) *Designing and conducting mixed methods research* (2nd ed.). Sage Publications.

Crocker, L., & Algina, J. (2008). *Introduction to classical and modern test theory*. Cengage Learning.

D'Oliveira, T.C. (2004). Dynamic spatial ability: An exploratory analysis and a confirmatory study. *The International Journal of Aviation Psychology*, *14*(1), 19-38. https://doi.org/10.1207/s15327108ijap1401_2

Davis, L.L. (1992). Instrument review: getting the most from a panel of experts. *Applied Nursing Research*, 5, 194-197. https://doi.org/10.1016/S0897-1897(05)80008-4

Diederich, P.B. (1973). Short-cut statistics for teacher-made tests. Educational testing service. https://files.eric.ed.gov/fulltext/ED081785.pdf

Dokumacı Sütçü, N. (2021). Research tendencies towards spatial ability in Turkey. *International Journal* of Society *Researches*, *17*(36), 2605-2636. https://doi.org/10.26466/opus.839496
Downing, S.M. (2006). Twelve steps for effective test development. In S.M. Downing & T.M. Haladyna (Eds.), *Handbook of test development* (pp. 3–26). Lawrence Erlbaum Associates.
Ebel, R.L., & Frisbie, D.A. (1991). *Essentials of educational measurement* (5th ed.). Prentice-Hall.
Ekstrom, R.B., French, J.W., Harman, H.H., & Dermen, D. (1976). *Manual for kit of factor-referenced cognitive tests*. Educational Testing Service.
Eliot, J., & Hauptman, A. (1981). Different dimensions of spatial ability. *Studies in Science Education*, *8*(1), 45-66. https://doi.org/10.1080/03057268108559886
Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences*. Basic Books.
Gilligan, K.A., Flouri, E., & Farran, E.K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of experimental child psychology*, *163*, 107-125. https://doi.org/10.1016/j.jecp.2017.04.016
Gronlund, N.E. (1977). *Constructing achievement tests* (2nd ed.). Prentice-Hall.
Guay, R.B. (1977). *Purdue spatial visualization test.* Purdue Research Foundation.
Guilford, J.P., Fruchter, B., & Zimmerman, W.S. (1952). Factor analysis of the army air forces sheppard field battery of experimental aptitude tests. *Psychometrika*, *17*(1), 45–68. https://doi.org/10.1007/BF02288795
Guilford, J.P., & Zimmerman, W.S. (1948). The Guilford-Zimmerman Aptitude Survey. *Journal of applied Psychology*, *32*(1), 24-34. https://doi.org/10.1037/h0063610
Hair, J.F., Jr., Black, W.C., Babin, B.J., Anderson, R.E., & Tatham, R.L. (2014). *Multivariate data analysis* (7th ed.). Pearson New International Edition.
Hambleton, R.K., & Jirka, S.J. (2006). Anchor-based methods for judgmentally estimating item statistics. In Downing, S.M., & Haladyna, T.M. (Eds.). *Handbook of test development* (pp. 399-420). Lawrence Erlbaum Associates Publishers.
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
Hingorjo, M.R., & Jaleel, F. (2012). Analysis of one-best MCQs: the difficulty index, discrimination index and distractor efficiency. *JPMA-Journal of the Pakistan Medical Association*, *62*(2), 142-147. https://www.jpma.org.pk/PdfDownload/3255
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
Irwing, P., Booth, T., & Hughes, D.J. (2018). *The Wiley handbook of psychometric testing. A Multidisciplinary Reference on Survey, Scale and Test Development.* Wiley-Blackwell.
İpekoğlu, A., Kepceoğlu, İ. & Biber, A.Ç. (2020). Thematic and methodological trends of graduate theses related to spatial ability: the case of Turkey. *International Journal of Contemporary Educational Studies*, *6*(2), 681-699. https://dergipark.org.tr/en/pub/intjces/issue/59193/829670
Jöreskog, K.G. & Sörbom, D. (1993). *LISREL 8: Structural Equations Modeling with the SIMPLES Command Language*. Scientific Software, International.
Jöreskog, K.G. & Sörbom, D. (1996). *Lisrell 8 reference guide.* Scientific Software International.
Kell, H.J., Lubinski, D., Benbow, C.P., & Steiger, J.H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science*, *24*(9), 1831-1836. https://doi.org/10.1177/0956797613478615

Kerkman, D.D., Wise, J.C., & Harwood, E.A. (2000). Impossible "mental rotation" problems: A mismeasure of women's spatial abilities?. *Learning and Individual Differences*, *12*(3), 253-269. https://doi.org/10.1016/S1041-6080(01)00039-5

Kim, J., & Irizarry, J. (2021). Evaluating the use of augmented reality technology to improve construction management student's spatial skills. *International Journal of Construction Education and Research*, *17*(2), 99-116. https://doi.org/10.1080/15578771.2020.1717680

Kline, R.B. (2011). *Principles and practice of structural equation modeling*. 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

Likert, R., & Quasha, W.H. (1941). *Revised Minnesota Paper Form Board*. Psychological Corporation.

Linn, M., & Petersen, A.C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, *56*(6), 1479-1498. https://www.jstor.org/stable/1130467

Livingston, S.A. (2006). Item Analysis. In Downing, S. M., & Haladyna, T. M. (Eds.). *Handbook of test development* (pp. 421-441). Lawrence Erlbaum Associates Publishers.

Lohman, D.F. (1979). Spatial ability: A review and reanalysis of the correlational literature. *Technical Report. No. 8*, Stanford University. https://apps.dtic.mil/sti/pdfs/ADA075972.pdf

Lohman, D.F. (1993, July 8). *Spatial ability and g*. Paper presented at the first Spearman Seminar, University of Plymouth, England. https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.111.7385&rep=rep1&type=pdf

Lord, T.R. (1985). Enhancing the visuo-spatial aptitude of students. *Journal of research in science teaching*, *22*(5), 395-405. https://doi.org/10.1002/tea.3660220503

Maier, P.H. (1996, March). Spatial geometry and spatial ability–How to make solid geometry solid. In *Selected papers from the Annual Conference of Didactics of Mathematics* (pp. 69-81). http://webdoc.sub.gwdg.de/ebook/e/gdm/1996/maier.pdf

Martín-Dorta, N., Saorín, J.L., & Contero, M. (2008). Development of a fast remedial course to improve the spatial abilities of engineering students. *Journal of Engineering Education*, *97*(4), 505-513. https://doi.org/10.1002/j.2168-9830.2008.tb00996.x

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

Mohler, J.L. (2008). Examining the spatial ability phenomenon from the student's perspective. *The Engineering Design Graphics Journal*, *72*(3), 1-15. http://gpejournal.org/index.php/EDGJ/article/view/50

National Research Council (NRC) (2006). *Learning to think spatially*. National Academies Press.

Olkun, S. (2003). Making connections: Improving spatial abilities with engineering drawing activities. *International Journal of Mathematics Teaching and Learning*, *3*(1), 1-10. http://www.ex.ac.uk/cimt/ijmtl/ijabout.htm

Olkun, S., Smith, G.G., Gerretson, H., Yuan, Y., & Joutsenlahti, J. (2009). Comparing and enhancing spatial skills of pre-service elementary school teachers in Finland, Taiwan, USA, and Turkey. *Procedia-Social and Behavioral Sciences*, *1*(1), 1545-1548. https://doi.org/10.1016/j.sbspro.2009.01.271

Ozcakir Sumen, O. (2019). A meta-synthesis about the studies on spatial skills in Turkey. *International Online Journal of Educational Sciences*, *11*(4), 23-41. https://doi.org/10.15345/iojes.2019.04.003

Patkin, D., & Dayan, E. (2013). The intelligence of observation: improving high school students' spatial ability by means of intervention unit. *International Journal of Mathematical Education in Science and Technology*, *44*(2), 179-195. https://doi.org/10.1080/0020739X.2012.703335

Pellegrino, J.W., Alderton, D.L., & Shute, V.J. (1984). Understanding spatial ability. *Educational psychologist*, *19*(4), 239-253. https://doi.org/10.1080/00461528409529300

Pellegrino, J.W., & Kail, R.V. (1982). Process analyses of spatial aptitude. In R. J. Sternberg (Eds.), *Advances in the psychology of human intelligence* (Vol. I, pp. 311-365). Lawrence Erlbaum Associates.

Sarama, J., & Clements, D. (2009). *Early childhood mathematics education research: Learning trajectories for young children*. Routledge.

Sevimli, E. (2009). *Consideration of pre-services mathematics teachers' preferences of representation in terms of definite integral within the context of certain spatial abilities and academic achievement* [Master's thesis]. University of Marmara, Turkey.

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

Sisman, B., Kucuk, S., & Yaman, Y. (2021). The effects of robotics training on children's spatial ability and attitude toward STEM. *International Journal of Social Robotics, 13*(2), 379-389. https://doi.org/10.1007/s12369-020-00646-9

Sorby, S.A., & Baartmans, B.J. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first year engineering students. *Journal of Engineering Education, 89*(3), 301-307. https://doi.org/10.1002/j.2168-9830.2000.tb00529.x

Tabachnick, B.G., & Fidell, L.S. (2013). *Using multivariate statistics* (6th ed.). Pearson.

Tartre, L.A. (1990). Spatial orientation skill and mathematical problem solving. *Journal for research in Mathematics Education*, *21*(3), 216-229. https://doi.org/10.5951/jresematheduc.21.3.0216

Thurstone, L.L., & Thurstone, T.G. (1941). *The Chicago tests of primary mental abilities.* The American Council on Education.

Turgut, M., & Nagy-Kondor, R. (2013). Spatial visualization skills of Hungarian and Turkish prospective mathematics teachers. *International Journal for Studies in Mathematics Education*, *6*(1), 168-183. https://doi.org/10.17921/2176-5634.2013v6n1p%25p

Wells, C.S., & Wollack, J.A. (2003). *An instructor's guide to understanding test reliability*. Testing and Evaluation Services, University of Wisconsin.

Wendler, C.L., & Walker, M.E. (2006). Practical issues in designing and maintaining multiple test forms for large-scale programs. In Downing, S.M., & Haladyna, T.M. (Eds.), *Handbook of test development* (pp. 445-467). Lawrence Erlbaum Associates Publishers.

Winter, J.W., Lappan, G., Fitzgerald, W., & Shroyer, J. (1989). *Middle Grades Mathematics Project: Spatial Visualization*. Addison-Wesley.

Witkin, H.A. (1950). Individual differences in ease of perception of embedded figures. *Journal of Personality*, *19*, 1-15. https://doi.org/10.1111/j.1467-6494.1950.tb01084.x

Young, C.J., Levine, S.C., & Mix, K.S. (2018). The connection between spatial and mathematical ability across development. *Frontiers in psychology*, *9(1)*, 1-7. https://doi.org/10.3389/fpsyg. 2018.00755