Journal of Theoretical Educational Science, 18(2), 280-301, April 2025
Kuramsal Eğitimbilim Dergisi, 18(2), 280-301, Nisan 2025
[Online]: http://dergipark.org.tr/akukeg
DOI number: http://doi.org/10.30831/akukeg.1581042

Developing a Spatial Thinking Skills Test in Geography Teaching*

Coğrafya Öğretiminde Mekânsal Düşünme Beceri Testi Geliştirme

Atakan YALÇIN** **Cennet ŞANLI***** **Adnan PINAR******

Received: 7 November 2024 **Research Article** **Accepted:** 13 March 2025

ABSTRACT: This study aimed to develop a test to measure university students' spatial thinking skills. The research was conducted using a survey design, with a sample of 260 undergraduate students from geography teaching and geography departments. GIS software was used to incorporate maps and satellite images, enhancing the spatial representation in the test. The test was administered online. Item analysis was conducted for each question item, and items with a difficulty index above 0.70 and a discrimination index below 0.19 were removed from the test. Additionally, the scores of the lower and upper groups were compared using an independent samples t-test. Exploratory factor analysis (EFA) based on the tetrachoric correlation coefficient was performed to assess the construct validity of the test. This analysis yielded a single-factor structure that accounted for 30% of the total variance. Items with unsatisfactory factor loadings were also removed. The internal consistency coefficient calculated using the KR20 formula yielded a value of 0.83, indicating good internal consistency. The stability of the test was assessed using the test-retest method, confirming its stable structure. With these refinements, the test was finalized with 26 items. The study concluded that the developed test is a valid and reliable tool for measuring spatial thinking skills among undergraduate students studying geography or geography teaching.

Keywords: Geography education, spatial thinking, spatial thinking skills, test development.

ÖZ: Bu çalışmada, üniversite öğrencilerinin mekânsal düşünme becerilerini ölçmek amacıyla bir test geliştirmek hedeflenmiştir. Araştırma, coğrafya öğretmenliği ve coğrafya bölümlerinden 260 lisans öğrencisiyle, tarama deseni kullanılarak yürütülmüştür. CBS yazılımı kullanılarak teste harita ve uydu görüntüleri eklenmiş ve mekânsal temsil güçlendirilmiştir. Test çevrim içi olarak uygulanmıştır. Soru maddelerinin her birine madde analizi yapılmıştır. Madde güçlüğü 0.70'in üzerinde olan soru maddeleri ve madde ayırt ediciliği 0.19'un altında olan soru maddeleri testten çıkarılmıştır. Ayrıca, alt ve üst grupların puanları bağımsız örneklem t-testi ile karşılaştırılmıştır. Testin yapı geçerliğini değerlendirmek için tetrakorik korelasyon katsayısına dayalı açımlayıcı faktör analizi (AFA) yapılmıştır. Bu analiz, toplam varyansın %30'unu açıklayan tek faktörlü bir yapı ortaya koymuştur. Düşük faktör yüküne sahip maddeler testten çıkarılmıştır. Testin iç tutarlılığı KR20 formülü ile hesaplanmış ve 0.83 bulunmuştur. Testin kararlılığı test-tekrar test yöntemiyle değerlendirilmiş ve kararlı bir yapıda olduğu doğrulanmıştır. 31 madde olarak tasarlanan testten yapılan istatistiksel analizler sonunda toplam 5 madde çıkarılmıştır. Testin son hali 26 maddeden oluşmaktadır. Araştırma sonucunda, geliştirilen testin coğrafya ve coğrafya öğretmenliği lisans öğrencilerinin mekânsal düşünme becerilerini ölçmede geçerli ve güvenilir bir araç olduğunu sonucuna varılmıştır.

Anahtar kelimeler: Coğrafya eğitimi, mekânsal düşünme, mekânsal düşünme becerisi, test geliştirme.

* This study is produced as master thesis of the first author under the supervision of the second author and third author.

** *Corresponding Author*: Res. Asst., Necmettin Erbakan University, Konya, Türkiye, atakanylcin@gmail.com, https://orcid.org/0000-0002-2242-3536
*** Assoc. Prof. Dr., Pamukkale University, Denizli, Türkiye, csanli@pau.edu.tr, https://orcid.org/0000-0003-3285-0950
**** Prof. Dr., Necmettin Erbakan University, Konya, Türkiye, apinar@erbakan.edu.tr, https://orcid.org/0000-0002-5256-7901

Citation Information
Yalçın, A., Şanlı, C., Pınar, A. (2025). Developing a spatial thinking skills test in geography teaching. *Kuramsal Eğitimbilim Dergisi [Journal of Theoretical Educational Science], 18*(2), 280-301.

ISSN: 1308-1659

In today's complex world, individuals must be capable of generating solutions to the challenges that they face by leveraging their creativity, skills, inquisitiveness, and sense of responsibility. Skill-based education plays a critical role in developing these competencies. Therefore, geography education aims to equip individuals with key skills. Among these skills, spatial thinking has garnered significant attention in recent years. Responding to this growing emphasis on spatial skills, the revised 2024 Geography Curriculum in Türkiye introduces a spatial information technologies unit across all secondary education grades. In this unit, students are encouraged to achieve learning outcomes such as analyzing and reading maps, linking events and, phenomena, finding locations, understanding the concept of location, using Geographic Information Systems (GIS), creating spatial data, and preparing maps using GIS. This emphasis on spatial thinking in the updated curriculum reflects the increasing interest in spatial thinking within geography education but also signals its growing significance in future research.

The concept of spatial thinking carries various interpretations across academic contexts, scales, and disciplines (DiBiase, 2013). Thus, several definitions of spatial thinking exist. The "Learning to Think Spatially" report provides a widely accepted definition, describing spatial thinking as a combination of three interconnected components: spatial concepts, methods for representing spatial information, and spatial reasoning processes (National Research Council, 2006). Similarly, Sinton, Bednarz, Gersmehl et al. (2013) describe spatial thinking as a cognitive skill involving the visualization and evaluation of location, position, distance, direction, environmental relationships, movement, and change over time across various situations and scales. Collectively, these definitions highlight spatial thinking as a process of understanding and reasoning with spatial concepts, while employing spatial representation tools such as maps and graphics to express spatial concepts and relationships.

Due to its broad applications, spatial thinking is fundamental to numerous everyday activities. People engage in spatial thinking when they interact with their environment and thereby interpret the meaning of objects within spatial contexts (Montello et al., 2014). Spatial thinking is evident in activities such as packing a suitcase, assembling furniture by following instructions, setting a dining table, selecting a new home, navigating directions, and judging distances (Sinton et al., 2013; Verma, 2014). While obtaining spatial information and making sense of space, individuals actively use spatial thinking to visualize dimensions, arrangements of objects, distances to other objects, and directions (Sinton, 2011). Therefore, spatial thinking plays a significant role in the execution of all kinds of essential activities (Bodenhamer et al., 2010). This skill is crucial not only in daily life but also across diverse professional fields, including GIS, engineering projects, and surgical interventions. In STEM disciplines, it is specifically crucial for solving complex problems, performing analytical tasks, conceptualizing abstract ideas, and developing models.

Spatial thinking also plays a central role in geography as the discipline examines interactions between humans and the natural world and focuses on the analysis of space. Some researchers refer to the application of spatial thinking within geographical contexts as "geospatial thinking" (Golledge et al., 2008). This idea can be better illustrated in Kerski’s (2008) definition of geography as a spatial science that explores relationships among social and environmental phenomena. Therefore, geography can be seen as a science fundamentally based on the analysis of spatial relations. This perspective aligns with what is known as the "geographical perspective", a concept used by geographers to understand human-environment interactions (Cutter et al., 2002; Golledge, 2002; Hanson, 2004). The geographical perspective refers not only to the locations of certain elements but also to more complex processes that involve spatial relationships, distances, networks, and hierarchies among those elements and geographical features. This perspective is closely related to spatial thinking skills (Montello, 1993). The concepts, representation tools, and reasoning processes used by a spatial thinker are, in fact, the fundamental elements used by a geographer to analyze the characteristics and spatial relationships of physical or human elements (Sinton et al., 2013). Thus, individuals who have developed acquired spatial thinking skills use them to engage in a more effective geographical inquiry across all subjects of interest in geography and offer more rational solutions to geographical problems that they encounter.

Given the centrality of spatial thinking in human life, a critical question emerges: Can spatial thinking skills be developed through education? Research suggests that spatial thinking can indeed be learned and improved through carefully designed educational approaches (Huynh & Sharpe, 2013; National Research Council, 2006; Uttal et al., 2013). Geography education that incorporates the foundational elements of spatial thinking such as spatial reasoning and visualization appears particularly effective in fostering these skills (Battersby et al., 2006). To foster spatial thinking in geography classes, it is essential to assess the spatial thinking skills of pre-service geography teachers who will soon be entering the teaching profession. Understanding the reasons behind any deficiencies and identifying their current levels are necessary steps for closing gaps and strengthening spatial thinking skills among future teachers. To develop students' spatial thinking, geography classes should strongly emphasize spatial thinking, and teachers should be well-equipped with methods and techniques to teach these skills.

Numerous studies have attempted to measure and improve spatial thinking. Studies aimed at measuring spatial thinking have been conducted mostly within the field of psychology, often using the umbrella term "spatial skills." These studies have typically attempted to measure specific sub-factors, which are considered to be the components of spatial skills, such as spatial orientation and spatial visualization. Examples include the "Card Flipping Test" developed by French, Ekstrom and Price (1963), the "Embedded Figures Test" developed by Witkin (1950) and mental rotation and surface development tests used by Yurt and Tünkler (2016). These tests were generally designed using various geometric shapes and figures. Although later versions of similar tests were developed for different age groups using a range of visuals, the fundamental idea remained the same (Charcharos et al., 2015). Studies using the term "spatial skills" have most often examined sub-factors such as spatial visualization, mental rotation, and spatial orientation (Arıkan, 2023). While geographers generally agree on these sub-factors, they also point out an additional sub-factor related to spatial patterns and relationships, referred to as "spatial relationships" (Albert & Golledge, 1999; Gilmartin & Patton, 1984; Golledge & Stimson, 1997; Self et al., 1992; Self & Golledge, 1994). The spatial relationships sub-factor, which is particularly important for geographers, entails functions such as understanding and using distance, spatial connections, and relationships; developing spatial hierarchies; understanding and linking spatial distributions and patterns; classifying information into spatial units such as regions; visualizing maps using verbal expressions; drawing maps; and examining maps though overlay analysis (Gilmartin & Patton, 1984; Golledge et al., 1995; Golledge & Stimson, 1997; Self & Golledge, 1994). Considering these functions of spatial relationships, it becomes evident that they play a crucial role in the discipline of geography.

Because geographers place a greater emphasis on spatial relationships within spatial thinking, spatial skills tests developed in psychology were considered insufficient to measure spatial thinking from the perspective of geography. This is because psychology-based assessments often lack the dimension of spatial relationships, a factor that geographers consider central to spatial skills (Bednarz & Lee, 2011; Goldstein et al., 1990; Golledge, 1993; Lee & Bednarz, 2009; Newcombe & Dubas, 1992; Self et al., 1992). Additionally, the validity and reliability of such measures are still debated (Lee & Bednarz, 2012). Thus, in geography education, several tools have been developed to measure spatial thinking (Huynh & Sharpe, 2013; Lee & Bednarz, 2012; Şanlı, 2021), However, these tools remain relatively limited. Given that pre-service geography teachers are critical to advancing spatial thinking skills, assessing their spatial thinking skills is vital for identifying gaps and implementing measures. This study aims to contribute to the literature by developing a spatial thinking skills test (STST) that can be adapted across cultural and linguistic contexts.

The STST, created by refining items from existing tests, integrates a variety of spatial reasoning processes and geographic themes. In the test, each spatial representation tool in the items consists of geographical elements and is related to different geographical topics. The STST is designed for use in experimental and correlational research, as well as educational applications. The test is expected to provide a reliable and valid measurement of spatial thinking among pre-service geography teachers, thereby supporting future research and educational efforts in this area.

Method

Research Design

This study is a skills test development study using a survey design. Survey research aims to capture the views or characteristics of larger populations by collecting data from representative samples (Büyüköztürk et al., 2020). The overarching purpose of survey research is to uncover general tendencies, attitudes, or opinions within a population by examining a selected sample (Creswell, 2017, p. 155).

Sampling

The population consisted of university students studying geography teaching or geography at state universities in Türkiye. The sample consisted of 260 university students who enrolled in these programs during the 2023-2024 academic year. Convenience sampling was used in sampling. In this method, researchers select participants who are most easily accessible until the required sample size is reached (Cohen & Manion, 1998). Detailed information about the participants is presented in Table 1.

Table 1

Demographics of the Participants

Variable	Category	*f*	%
Gender	Female	155	60
	Male	105	40
Year of Study	First	66	%25
	Second	65	%25
	Third	90	%35
	Fourth	39	%15
Faculty	Faculty of Education	115	%44
	Faculty of Humanities and Social Sciences	80	%31
	Faculty of Arts and Sciences	65	%25
Department	Geography Teaching	115	%44
	Geography	145	%56
GIS Course	Attended	166	%64
	Did not attend	94	%36

Measures

The STST developed by researchers was used as the data collection tool in this study. The test development stages outlined by Baykul & Turgut (2019) were followed during the development of the test. These stages are detailed below.

Determining the Purpose of the Test

This study aimed to develop a test specifically designed to measure spatial thinking skills from a geographical perspective. The developed test seeks to include items that assess spatial thinking related to the concept of space as perceived in a geographical context, as well as the spatial analysis of problems that may be encountered in daily life.

Determining the Characteristics to be Measured by the Test

To establish the specific characteristics to be measured by the test, an extensive literature review was conducted. This review examined various definitions of spatial thinking, identifying the key components within each definition. Focusing on widely accepted frameworks, the focus was concentrated on three principal components recognized in the international literature: spatial concepts, spatial representation tools, and spatial reasoning processes (National Research Council, 2006). In determining the scope of the STST developed in this study, established classifications of spatial concepts in the literature were considered (Golledge, 2002; Gersmehl & Gersmehl, 2007, 2006; Golledge et al., 2008; Jo & Bednarz, 2009). These classifications, especially those developed specifically for spatial concepts, were instrumental in selecting the features the test should measure.

To evaluate the ability to use spatial representation tools, which is a key component of spatial thinking, the test was designed to include tasks involving map creation using GIS, as well as spatial analysis methods that require diverse reasoning processes. Additionally, to align the STST with existing research and address any limitations identified in prior studies, existing spatial thinking skills tests were reviewed. The scope of the test was then defined in line with research recommendations and known gaps.

For reference, the Spatial Thinking Ability Test (STAT) by Lee and Bednarz (2012) measures eight key characteristics: (1) understanding direction and orientation, (2) comparing the information on the map with the information on the graph, (3) selecting optimal locations based on criteria, (4) visualizing a profile on a topographic map, (5) associating phenomena based on their spatial distribution, (6) visualizing 3D images in the mind based on 2D information, (7) overlaying and analyzing maps, (8) understanding geographical features represented as points, lines and areas. Another test developed by Şanlı (2021) measures nine characteristics, adding spatial hierarchy to the existing list: (1) finding location and direction, (2) showing patterns on maps graphically, (3) selecting ideal locations within spatial contexts, (4) profiling on topographic maps, (5) understanding spatial pattern correlations, (6) converting 2D topographic maps to 3D maps, (7) performing overlay operations, (8) representing geographical data as points, lines, and areas, and (9) spatial hierarchy.

In determining the features of the STST developed here, an effort was made to assess unique aspects of spatial thinking beyond those included in previous tests (Lee & Bednarz, 2012; Şanlı, 2021). The features measured in the STST include (1) mentally visualizing the profile of a topographic map, (2) finding locations and directions, (3) graphically representing spatial distributions on maps, (4) performing overlay operations and problem-solving tasks in spatial elements, (5) analyzing spatial concepts such as size, distance, and area of influence on maps, (6) conducting spatial network analysis, (7) conducting spatial risk analysis, (8) selecting optimal locations in the organization of space based on specific criteria, (9) understanding spatial correlations, (10) understanding spatial hierarchies, (11) associating geographical data (points, lines, areas) with spatial elements, and (12) understanding spatial distributions. By incorporating these features, the STST offers a broader, more comprehensive scope of measurement for spatial thinking skills, addressing a wide range of cognitive and analytical skills essential in geography education.

Creating Test Items

In developing the test items, various spatial representation tools such as maps, graphics, and satellite images were used to address the spatial concepts relevant to the characteristics intended to be measured in each question. GIS software (ArcGIS 10.7) was used to design maps and adapt satellite images according to the specific requirements of each question. This approach allowed the researcher to create customized maps tailored to measure each targeted feature. Each item included a prepared map or satellite image, accompanying question text, and five choices, with only one correct answer among them. Correct answers receive 1 point, and incorrect answers receive 0 points. Through this process, a question pool was compiled.

Revising Test Items

The initial question pool was reviewed by three faculty members, each an expert in the field of education, to ensure the quality and effectiveness of the test items. The experts were asked to provide feedback on various aspects, including the relevance of each question to the intended characteristics outlined in the specification table, the clarity of the wording, the suitability and comprehensibility of the representation tools used (such as maps, graphics, and satellite images), and the overall appropriateness of the questions for the target audience.

Preparing the Draft Form and Piloting

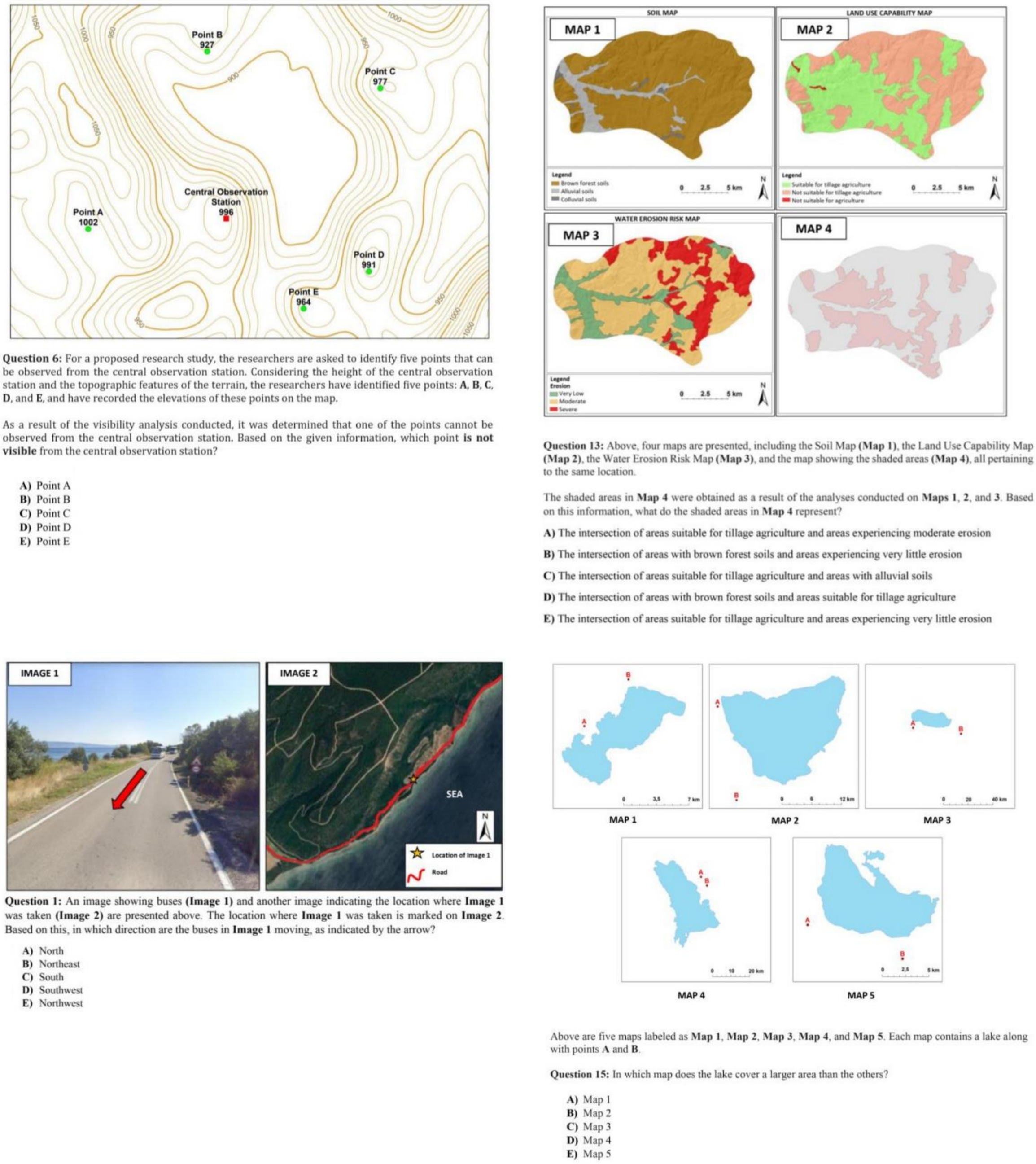
Following the expert review, revisions were made to certain items based on their feedback, and a draft form was created, containing 31 items selected from the question pool. This draft was administered to a small group of 10 students in the geography teaching department to determine the time needed to complete the test, evaluate the clarity of the questions from a student perspective, and gather preliminary data for statistical analysis.

Finalizing and Administering the Test

Using the data from the pilot study, further revisions were made to enhance certain questions in the finalized version of the test. The final test items were then uploaded to Google Forms, allowing easy access via mobile devices. Before the test was administered, the participants were given a brief introduction to the STST. The test was administered in a classroom with either the researcher or an instructor present to address any questions, ensuring a reliable and supportive environment for data collection.

Figure 1

Sample Question Items on the STST



Data Analysis

To evaluate the validity and reliability of the test, an initial analysis was conducted on the distribution of participant scores by calculating skewness and kurtosis coefficients. When these coefficients fall between +1 and -1, it indicates that the data follow a normal distribution (Tabachnick & Fidell, 2007). To assess the ability of the STST to differentiate between participants with varying levels of spatial thinking skills, the top 27% of students (upper group) were identified as the most successful, while the bottom 27% (lower group) represented the least successful group. The mean scores of the two groups were then compared using an independent samples t-test. Item difficulty and discrimination indices were calculated for each test item. The item discrimination index was derived by subtracting the number of correct answers in the lower group from those in the upper group and dividing by the number of respondents in the upper group.

The item difficulty index was calculated by summing the number of correct answers in upper and lower groups and dividing by the total number of respondents.

Construct and content validity were examined to establish the validity of the test. In the content validity analysis, each item was evaluated to assess whether it was sufficient to measure the intended characteristic. Preparing a specification table and seeking expert opinion are common methods used to ensure content validity (Büyüköztürk et al., 2020). The content validity of the STST was assessed by three faculty members who are experts in geography education. In the construct validity analysis, the Kaiser-Meyer-Olkin (KMO) value and Bartlett's test results were examined to assess the suitability of the data for factor analysis. A KMO value of 0.6 or higher and significant Bartlett's test results indicate that the data are appropriate for factor analysis (Tabachnick & Fidell, 2013). The data were found to be suitable, and exploratory factor analysis (EFA) was subsequently conducted on the test scores to identify the factor structure. Given that STST items were scored dichotomously (1 for correct and 0 for incorrect answers), the data were classified as binary. For dichotomous, normally distributed data, tetrachoric factor analysis is commonly used to reveal the factor structure of a measure (Baykul & Güzeller, 2022; Dokumacı Sütçü & Oral, 2019; Hambleton & Swaminathan, 1985; Olsson, 1979; Uebersax, 2015). To assess the construct validity of the STST, a factor analysis based on the tetrachoric correlation coefficient was conducted using the “Factor 12.04.05” software developed by Lorenzo-Seva & Ferrando (2021).

To assess the reliability of the test, the internal consistency coefficient was calculated, and the stability of the test was evaluated. Internal consistency was calculated using the Kuder-Richardson (KR20) formula. For a test to be considered reliable, the coefficient calculated using internal consistency formulas must be at least 0.7 (Heale & Twycross, 2015). However, for tests used in educational settings, a KR20 coefficient of at least 0.80 is preferred (Özçelik, 2010). The stability of the test was verified using the test-retest method. The STST was re-administered to a subset of 50 participants after a 25-day interval. The consistency between the two sets of scores was measured using the Pearson product-moment correlation coefficient.

Ethical Procedures

Before the study was conducted, permission was obtained from the Ethics Committee for Scientific Research in Social and Human Sciences at Necmettin Erbakan University, with the decision number 2023/113 dated 10.03.2023. During the administration of the test to students, it was stated that test results would be kept confidential and used solely for this study. Additionally, students were reminded that participation was voluntary, and they could withdraw from the study at any time.

Results

Descriptive Analysis Results

The STST consists of 31 items, with each correct response scored 1 point and each incorrect response 0 points. Among the 260 participants, the scores ranged from a minimum of 3 to a maximum of 29 points, with an average score of 14.65 and a standard deviation of 5.96 (Table 2). To assess the distribution of the test scores, the skewness and kurtosis values were calculated, yielding a skewness of 0.132 and a kurtosis of -0.694 (Table 2). These values indicate that the scores obtained from the test follow a normal distribution.

Table 2

Descriptive Statistics of Test Scores

Lowest Score	Highest Score	*Mean*	Standard Deviation	Variance	Skewness	Kurtosis
3	29	*14.65*	5.96	35.558	0.132	-0.694

Item Analysis Results

Item analyses were conducted to evaluate the difficulty and discrimination properties of the test items. The participants were ranked from highest to lowest based on their test scores. According to the scores of the participants (n = 260), the top 27% (70 students) were classified as the upper group, while the bottom 27% (70 students) were identified as the lower group. To assess item discrimination, an independent samples t-test was conducted to determine whether there was a significant difference between the scores of the upper and lower groups. The results showed that the average score of the upper group was 22.17, while the average score of the lower group was 7.30. A significant difference was observed between the total test scores of the lower and upper groups ($t = 35.689$, $p < .05$) (Table 3). For each test item, the discrimination coefficient was calculated to determine the extent to which each item distinguished between the upper and lower groups. This was achieved by subtracting the number of correct responses in the lower group from those in the upper group and then dividing by the total number of students in the upper group. Item discrimination coefficients (Rjx) were then calculated based on these statistics.

Table 3

T-Test Results for the Test Scores of the Upper and Lower Groups

	n	X (min-max)	SD	*t*	*p*
Upper Group	70	22.17 (19-29)	2.724	35.689	.001
Lower Group	70	7.30 (3-11)	2.205		

As shown in Table 4, items with a discrimination index of 0.19 or lower (#17) were classified as having poor discrimination, while items with a discrimination index between 0.20 and 0.29 (#5, #14, #23) were classified as having average discrimination. Items with a discrimination index between 0.30 and 0.39 (#1, #2, #10, #13, #15) demonstrated good discrimination, and items with a discrimination index of 0.40 or higher (#3, #4, #6, #7, #8, #9, #11, #12, #16, #18, #19, #20, #21, #22, #24, #25, #26, #27, #28, #29, #30, #31) were classified as having very good discrimination. The overall average item discrimination of the test was calculated by summing the discrimination indices of all items and dividing by the total number of items. The average item discrimination of the test was calculated to be 0.48, indicating a generally strong level of discrimination across the test items.

Table 4

Discriminatory Power of Test Items (Rjx)

Item Number	Item Discrimination (Rjx)	Item Number	Item Discrimination (Rjx)
#1	0.34	#17	0.15
#2	0.38	#18	0.52
#3	0.42	#19	0.64
#4	0.51	#20	0.42
#5	0.27	#21	0.58
#6	0.5	#22	0.67
#7	0.44	#23	0.24
#8	0.54	#24	0.71
#9	0.54	#25	0.6
#10	0.31	#26	0.6
#11	0.51	#27	0.48
#12	0.64	#28	0.78
#13	0.34	#29	0.65
#14	0.22	#30	0.62
#15	0.32	#31	0.54
#16	0.5		

To calculate the item difficulty coefficients (Pj) of the test, the number of students in the upper and lower groups who answered each item correctly was summed and then divided by the total number of students in both groups. Accordingly, items with a difficulty index below 0.29 (#13, #14, #15, #17) were classified as very difficult, items with a difficulty index between 0.30 and 0.49 (#1, #2, #3, #6, #7, #9, #10, #12, #18, #M20, #21, #23, #25, #26) were considered moderately difficult, items with a difficulty index between 0.50 and 0.69 (#4, #8, #11, #16, #19, #22, #24, #27, #28, #29, #30, #31) were classified as easy, and items with a difficulty index between 0.70 and 1.00 (#5) were classified as very easy (Table 5) (Hasançebi et al., 2020).

To determine the average difficulty of the test, the difficulty indices of all items were summed up and divided by the total number of items. The average difficulty of the test was found to be 0.47, suggesting that the test, on average, was moderately difficult.

Table 5

Difficulty of Test Items (Pj)

Item Number	Item Difficulty (Pj)	Item Number	Item Difficulty (Pj)
#1	0.34	#17	0.25
#2	0.30	#18	0.47
#3	0.45	#19	0.55
#4	0.6	#20	0.34
#5	0.86	#21	0.49
#6	0.49	#22	0.57
#7	0.32	#23	0.39
#8	0.55	#24	0.51
#9	0.44	#25	0.45
#10	0.47	#26	0.45
#11	0.62	#27	0.5
#12	0.49	#28	0.56
#13	0.27	#29	0.51
#14	0.27	#30	0.6
#15	0.29	#31	0.64
#16	0.60		

Validity Analysis Results

To establish the content validity of the test, a specification table was created to clearly outline the specific features each item was designed to measure. This specification table (Table 6), along with the test items, was submitted for review to three faculty members with expertise in geography education. Following their evaluation, it was concluded that the test items effectively measured spatial thinking from a geographical science perspective, and the test provides sufficient content validity for assessing spatial thinking in this context.

Table 6

Characteristics To Be Measured by the Test

Characteristics to be measured by the test	Item Number
Visualizing the profile of a topographic map	6, 9, 31
Finding locations and directions	1, 2, 10
Graphically representing spatial distributions on maps	7, 8
Performing overlay operations and problem-solving tasks in spatial elements	12, 13, 14
Analyzing spatial concepts such as size, distance, and area of influence on maps	15, 16, 25
Conducting spatial network analysis	11, 19
Conducting spatial risk analysis	20, 22, 23, 24
Selecting optimal locations in the organization of space based on specific criteria	17, 18
Understanding spatial correlations	4, 5
Understanding spatial hierarchies	21

Associating geographical data (points, lines, areas) with spatial elements	3
Understanding spatial distributions	26, 27, 28, 29, 30

EFA based on tetrachoric correlation was conducted to determine the factor structure of the STST. First, the KMO value and Bartlett's test results were examined to assess the adequacy of the sample for factor analysis. A KMO value of 0.6 or higher and significant Bartlett's test results indicate that the data are appropriate for factor analysis (Tabachnick & Fidell, 2013). The analysis revealed a KMO value of 0.837, and Bartlett's test of sphericity was significant ($p = .000010$), confirming that the data were appropriate for factor analysis. Items with insufficient discrimination (M17) and items that did not meet the desired levels of difficulty and discrimination (M5) were excluded from the factor analysis. When determining the number of factors in measurement instruments, eigenvalues, scree plots, and explained variance ratios are generally considered (Büyüköztürk, 2002; Tabachnick & Fidell, 2013). In factor analysis, factors with eigenvalues of 1 or higher are typically considered significant (Büyüköztürk, 2007).

As shown in Table 7, the factor analysis revealed a 10-factor structure with eigenvalues of 1 or higher. However, after the first factor, the eigenvalues of subsequent factors decreased significantly. Looking at the variance explained by the factors, it is evident that the first factor accounts for 27% of the variance, while the second factor explains only 6%. The factors after the first were considered insufficient to explain the total variance (Pallant, 2007). Additionally, the output file of the factor analysis indicated that a parallel analysis based on principal component analysis recommended a single-factor structure (advised number of dimensions: 1). Based on these findings, it was concluded that the test exhibits a single-factor structure. Subsequently, the factor loadings of each item within this structure were examined (Table 8). Items with factor loadings of 0.32 or lower are typically removed from the measurement tool (Tabachnick & Fidell, 2013). Consequently, items with factor loadings below 0.32 (#10, #14, #23) were removed from the test. Factor analysis was repeated after these items were excluded. The final analysis revealed that the factor loadings of the remaining items ranged from 0.392 (#13) to 0.802 (#28), with no item falling below the threshold of 0.32 (Table 9). After conducting the final factor analysis on the refined 26-item test, the variance explained by the single-factor structure was examined. Finally, as a result of the final factor analysis conducted on the 26-item version of the test, the variance explained by the single-factor structure was examined. For single-factor measurement tools, an explained variance of 30% or higher is considered sufficient (Büyüköztürk, 2007; Çokluk et al., 2016). The repeated tetrachoric factor analysis showed that the single-factor structure explained 30% of the total variance. The reliability of the single-factor structure was found to be 0.908 (Table 10).

Table 7

Eigenvalues and Variances Explained in the Factor Analysis

Factor	Eigenvalue	Variance Explained
1	8.365	0.27
2	1.882	0.06
3	1.767	0.05
4	1.601	0.05
5	1.527	0.05
6	1.292	0.04
7	1.183	0.03
8	1.119	0.03
9	1.092	0.03
10	1.050	0.03

Table 8

Factor Loadings of the Items in the First Analysis

Item Number	Factor Loadings	Communality	Item Number	Factor Loadings	Communality
#1	0.432	0.186	#18	0.500	0.250
#2	0.427	0.183	#19	0.614	0.377
#3	0.415	0.172	#20	0.508	0.258
#4	0.456	0.208	#21	0.592	0.350
#6	0.455	0.207	#22	0.623	0.388
#7	0.469	0.220	#23	0.218	0.047
#8	0.532	0.283	#24	0.681	0.463
#9	0.485	0.235	#25	0.617	0.380
#10	0.244	0.060	#26	0.602	0.362
#11	0.579	0.336	#27	0.447	0.200
#12	0.621	0.386	#28	0.793	0.629
#13	0.414	0.171	#29	0.645	0.416
#14	0.273	0.074	#30	0.564	0.318
#15	0.444	0.197	#31	0.587	0.345
#16	0.549	0.302			

Table 9

Factor Loadings of the Items in the Repeated Analysis

Item Number	Factor Loadings	Communality	Item Number	Factor Loadings	Communality
#1	0.428	0.183	#18	0.513	0.263
#2	0.458	0.209	#19	0.609	0.371
#3	0.412	0.170	#20	0.522	0.273
#4	0.446	0.199	#21	0.607	0.369
#6	0.462	0.213	#22	0.616	0.380
#7	0.474	0.225	#24	0.690	0.476
#8	0.525	0.276	#25	0.615	0.379
#9	0.501	0.251	#26	0.599	0.359
#11	0.576	0.331	#27	0.443	0.196
#12	0.600	0.360	#28	0.802	0.643
#13	0.392	0.154	#29	0.628	0.395
#15	0.461	0.213	#30	0.557	0.310
#16	0.563	0.317	#31	0.594	0.353

Table 10

Characteristics of the Single-Factor Structure

Factor	Eigenvalue	Variance Explained	Reliability Estimation
1	7.867	0.302	0.908

Reliability Analysis Results

The KR20 formula was used to assess the reliability of the STST. The KR20 formula is a method for assessing reliability based on the variance of the items in the test. It is typically used for tests scored dichotomously, where correct answers are scored as "1" and incorrect answers as "0" (Özçelik, 2010). The KR20 value of the test was found to be 0.83. The test-retest method was employed to assess the stability of the test. The test was administered to a subset of 50 participants after a 25-day interval. The Pearson product-moment correlation coefficient was calculated to determine the relationship between the scores of the two administrations. As shown in Table 11, the analysis yielded a significant ($p<0.05$) strong positive correlation ($r = 0.876$) between the first administration and the retest. Accordingly, there was a strong positive correlation between the scores from the first and second administrations, indicating a significant increase in the scores of the 50 students across both tests. Therefore, it can be concluded that the test demonstrates stability based on the test-retest method (Table 11).

Table 11

Test-Retest Results

		Retest
First Application	Pearson r	0.876
	P	0.001
	n	50

Discussion and Conclusion

This study aimed to develop a spatial thinking skills test (STST) to assess pre-service geography teachers' spatial thinking. Several stages of test development were followed during the creation of the STST. After these stages were completed, the validity and reliability analyses of the test were conducted using data collected from the finalized version of the test. A maximum score of 31 points can be obtained from the test, with the participants' average score being 14.65. This suggests that students in geography teaching and geography departments demonstrated a moderate level of success in the STST. Looking at the participants' average scores for specific items, it was observed that the students particularly scored lower on items related to "finding locations and directions" and "performing overlay operations and problem-solving tasks in spatial elements." These findings suggest areas where curriculum adjustments could be made to enhance skills such as directional awareness, spatial distribution, map analysis, and overlay operations. Additionally, these skills could be further sharpened through modifications to the content and implementation of the GIS course at universities, along with relevant practical activities.

Looking at the scores of the lower and upper groups, the average score of the lower group was found to be 7.30, while the average score of the upper group was 22.17. An independent samples t-test revealed a significant difference between the scores of the two groups. Additionally, the discrimination coefficients for each item were calculated. Items with a discrimination index (Rjx) of 0.19 or lower should be removed from the test, those with a discrimination index of 0.20 to 0.29 should be revised and improved, items with a discrimination index of 0.30 to 0.39 are considered quite good and can be retained, and items with a discrimination index of 0.40 or higher are classified as very good (Hasançebi et al., 2020). Accordingly, items with a discrimination index of 0.19 or lower (#17) had poor discrimination, items with a discrimination index of 0.20 to 0.29 (#5, #14, #23) were moderately discriminative, and items with a discrimination index of 0.30 to 0.39 (#1, #3, #10, #15) had very good discrimination. Items with a discrimination index of 0.40 or higher (#2, #4, #6, #7, #8, #9, #11, #12, #16, #18, #19, #20, #21, #22, #24, #25, #26, #27, #28, #29, #30, #31) were classified as having excellent discrimination.

An item is considered very difficult if the item difficulty coefficient is below 0.29, moderately difficult if it is between 0.30 and 0.49, easy if it is between 0.50 and 0.69, and very easy if it is between 0.70 and 1.00 (Hasançebi et al., 2020). Item difficulty values are generally expected to be around 0.50 (Büyüköztürk et al., 2020). Accordingly, items with a difficulty index below 0.29 (#2, #13, #14, #15, #17) were classified as very difficult, items with a difficulty index between 0.30 and 0.49 (#1, #7, #9, #10, #12, #20, #23, #26) were moderately difficult, items with a difficulty index between 0.50 and 0.69 (#3, #4, #6, #8, #11, #16, #18, #19, #21, #22, #24, #25, #27, #28, #29, #30, #31) were easy, and items with a difficulty index between 0.70 and 1.00 (#5) were classified as very easy. Based on the item discrimination and item difficulty coefficients, the following items were removed from the test: #17, which had poor discrimination, and #5 which was very easy. The analysis results revealed that the average discrimination of the entire test was 0.46, and the average difficulty value was 0.49. In test development, it is generally undesirable for a test to be either too difficult or too easy. The average difficulty of a test is typically expected to be around 0.50, and the average discrimination should be above 0.4 (Yılmazer, 2012). Given the findings on the discrimination and difficulty of the STST, it was concluded that the test had an average difficulty level (0.49) and demonstrated a very good ability to distinguish between successful and unsuccessful students in spatial thinking skills (0.46).

For validity assessment, content and construct validity were evaluated. To determine the validity of the test, both content validity and construct validity were assessed. Preparing a specification table is a commonly used method to ensure content validity. In this context, a specification table was developed to detail the specific spatial thinking skills measured by each item, and expert opinion was sought to confirm content validity. Similarly, previous studies on test development for measuring spatial thinking skills have used specification tables to ensure content validity (Huynh & Sharpe, 2013; Lee & Bednarz, 2012; Şanlı, 2021). In dichotomously scored tests, tetrachoric factor analysis is typically used to determine construct validity (Köroğlu et al., 2023; Özkılıç et al., 2023; Salar & Uğurel, 2020; Taşdemir, 2022). Thus, tetrachoric factor analysis was employed to assess the construct validity of the STST. The EFA results revealed that the test consisted of a single factor, which explained 30% of the total variance. It is considered sufficient for single-factor measurement tools to explain 30% or more of the total variance (Büyüköztürk, 2007; Çokluk et al., 2016).

Looking at the factor loadings of each item in the single-factor structure, some items were found to have factor loadings below 0.32. It is recommended that items with factor loadings below 0.32 be removed (Tabachnick & Fidell, 2013). Consequently, items with factor loadings below 0.32 (#10, #14, #23) were removed. The content validity of the STST was not affected because the remaining items in the test measured the same features as the removed items. It is believed that the identification of a single-factor structure in the STST is related to the emphasis on the spatial relations dimension, recognized as one of the sub-factors of spatial thinking particularly by geographers. This is because, in the test development process, the items were designed in a way that required the integration of various spatial concepts, spatial representation tools, and different reasoning processes while considering the spatial relations dimension. Therefore, it is thought that the test demonstrates a single-factor structure as a result of the analyses. The internal consistency of the test, measured by the KR20 formula, yielded a coefficient of 0.83. It is stated that the internal consistency coefficient in measurement tools should be at least 0.70 (Heale & Twycross, 2015). An internal consistency coefficient of 0.83 for the STST indicates that the test items are consistent with each other and the overall test. Therefore, it can be concluded that the internal consistency of the test is sufficient. The validity and reliability of the STST were assessed using various statistical methods. It was determined that the STST is a valid and reliable measurement tool for assessing spatial thinking skills among university students in geography teaching and geography departments.

Determining the spatial thinking skill levels of future geography educators is essential for identifying potential areas of improvement and implementing necessary educational interventions. Geography teachers, in particular, play a key role in imparting spatial thinking skills to students at the secondary education level. This tool offers a means of assessing pre-service geography teachers' spatial thinking skills. In conclusion, the STST is expected to contribute to future research by providing a reliable assessment of spatial thinking and facilitating the development of strategies to enhance these skills.

Acknowledgements

This study was not supported by any institution.

Statement of Responsibility

Atakan Yalçın: Creating test items, methodology, data collection, data analysis and interpretation, writing- original draft, visualization.

Cennet Şanlı: Reviewing test items, validation, writing-review and editing, supervision.

Adnan Pınar: Validation, writing-review and editing, supervision.

Conflicts of Interest

The authors have no conflict of interest to disclose.

Author Bios:

Research assistant Atakan YALÇIN completed his undergraduate education in Geography Teaching at Gazi University in 2020. He earned his master's degree in Geography Education from Necmettin Erbakan University in 2024 and is currently pursuing a Ph.D. in Geography Education at the same university. His research interests include geography education, spatial thinking skills, and geographic information systems.

Assoc. Prof. Dr. Cennet ŞANLI completed both her master's and doctoral degrees in Geography Education at Necmettin Erbakan University and received her associate professorship in the same field. Since 2017, she has been working in the Department of Geography at Pamukkale University. Her research areas include geography education and spatial thinking skills.

Prof. Dr. Adnan PINAR completed his master's degree in Physical Geography and his Ph.D. in Remote Sensing in the United Kingdom. He received his associate professorship in Geography Education. Since 1988, he has been working in the field of geography education at Selçuk University and Necmettin Erbakan University. His research areas include geography education, remote sensing, geographic information systems, and the use of technology in education.

References

Albert, W. S., & Golledge, R. G. (1999). The use of spatial cognitive abilities in geographical information systems: The map overlay operation. *Transactions in GIS*, *3*(1), 7–21. https://doi.org/10.1111/1467-9671.00003

Arıkan, A. (2023). *Kuantum öğrenme modeli'nin ortaokul öğrencilerinin mekânsal beceri öz yeterliklerine ve mekânsal becerilerine etkisinin incelenmesi* [Doctoral dissertation]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Battersby, S. E., Golledge, R. G., & Marsh, M. (2006). Incidental learning of geospatial concepts across grade levels: Map overlay. *The Journal of Geography*, *105*(4), 139-146. https://doi.org/10.1080/00221340608978679

Baykul, Y., & Güzeller, C. (2022). *Sosyal bilimler için istatistik (9th ed.)*. Pegem Akademi.

Baykul, Y., & Turgut, M. F. (2019). *Eğitimde ölçme ve değerlendirme (8th ed.)*. Pegem Akademi. https://doi.org/10.14527/9786053640875

Bednarz, R., & Lee, J. (2011). The components of spatial thinking: Empirical evidence. *Procedia - Social and Behavioral Sciences*, *21*, 103–107. https://doi.org/10.1016/j.sbspro.2011.07.048

Bodenhamer, D. J., Corrigan, J., & Harris, T. M. (2010). *The spatial humanities: GIS and the future of humanities scholarship*. Indiana University Press.

Büyüköztürk, Ş. (2002). Faktör analizi: Temel kavramlar ve ölçek geliştirmede kullanımı. *Kuram ve Uygulamada Eğitim Yönetimi*, *32*(32), 470-483.

Büyüköztürk, Ş. (2007). *Sosyal bilimler için veri analizi el kitabı*. Pegem Akademi.

Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş., & Demirel, F. (2020). *Eğitimde bilimsel araştırma yöntemleri*. Pegem Akademi.

Charcharos, C., Tomai, E., & Kokla, M. (2015, October). Assessing spatial thinking ability in geospatial thinking. Paper presented at the *Educating the Future Spatial Citizen: GEOTHNK International Closing Conference*, Pallini, Greece.

Cohen, L. M., & Manion, L. (1998). *Research methods in education*. Routledge.

Creswell, J. W. (2017). *Nicel yöntemler: Nitel, nicel ve karma yöntem yaklaşımları* (S. B. Demir, Trans.; 3rd ed.). Eğiten Kitap Yayıncılık.

Cutter, S. L., Golledge, R., & Graf, W. L. (2002). The big questions in Geography. *The Professional Geographer*, *54*(3), 305–317. https://doi.org/10.1111/0033-0124.00332

Çokluk, Ö., Şekercioğlu, G., & Büyüköztürk, Ş. (2016). *Sosyal bilimler için çok değişkenli istatistik: SPSS ve LISREL uygulamaları (6th ed.)*. Pegem Akademi.

DiBiase, D. (2013). There's more to spatial thinking than you think. *Essays in geography and GIS* (Report No. 6, pp. 3–4). Esri. https://www.esri.com/content/dam/esrisites/sitecore-archive/Files/Pdfs/library/bestpractices/essays-on-geography-gis-vol6.pdf

Dokumacı Sütçü, N., & Oral, B. (2019). Uzamsal görselleştirme testinin geliştirilmesi: Geçerlik ve güvenirlik çalışmaları. *Kastamonu Eğitim Dergisi*, *27*(3), 1179–1195. https://doi.org/10.24106/kefdergi.2826

French, J. W., Ekstrom, R. B., & Price, L. A. (1963). *Manual for kit of reference tests for cognitive factors* (Rev. ed.). Educational Testing Service.

Gersmehl, P., & Gersmehl, C. A. (2006). Wanted: A concise list of neurologically defensible and assessable spatial-thinking skills. *Research in Geographic Education*, *8*(1), 5-38.

Gersmehl, P., & Gersmehl, C. (2007). Spatial thinking by young children: Neurologic evidence for early development and "educability." *Journal of Geography*, *106*(5), 181–191. https://doi.org/10.1080/00221340701809108

Gilmartin, P. P., & Patton, J. C. (1984). Comparing the sexes on spatial abilities: Map-use skills. *Annals of the Association of American Geographers*, *74*(4), 605–619. https://doi.org/10.1111/J.1467-8306.1984.TB01477.X

Goldstein, D., Haldane, D., & Mitchell, C. (1990). Sex differences in visual-spatial ability: The role of performance factors. *Memory & Cognition*, *18*(5), 546–550. https://doi.org/10.3758/BF03198487

Golledge, R. G. (1993). Geographical perspectives on spatial cognition. *Advances in Psychology*, *96*, 16–46. https://doi.org/10.1016/S0166-4115(08)60038-2

Golledge, R. G. (2002). The nature of geographic knowledge. *Annals of the Association of American Geographers*, *92*(1), 1–14. https://doi.org/10.1111/1467-8306.00276

Golledge, R. G., Dougherty, V., & Bell, S. (1995). Acquiring spatial knowledge: Survey versus route-based knowledge in unfamiliar environments. *Annals of the Association of American Geographers*, *85*(1), 134-158. https://doi.org/10.1111/j.1467-8306.1995.tb01798.x

Golledge, R. G., & Stimson, R. J. (1997). *Spatial behavior: A geographic perspective*. Guilford Press.

Golledge, R., Marsh, M., & Battersby, S. (2008). Matching geospatial concepts with geographic educational needs. *Geographical Research*, *46*(1), 85-97. https://doi.org/10.1111/j.1745-5871.2007.00494.x

Hambleton, R., & Swaminathan, H. (1985). *Item response theory*. Kluwer Nijhoff Publishing.

Hanson, S. (2004). Who are "we"? An important question for geography's future. *Annals of the Association of American Geographers*, *94*(4), 715–722. https://doi.org/10.1111/j.1467-8306.2004.00425.x

Hasançebi, B., Terzi, Y., & Küçük, Z. (2020). Madde güçlük indeksi ve madde ayırt edicilik indeksine dayalı çeldirici analizi. *Gümüşhane Üniversitesi Fen Bilimleri Dergisi*, *10*(1), 224–240. https://doi.org/10.17714/gumusfenbil.615465

Heale, R., & Twycross, A. (2015). Validity and reliability in quantitative studies. *Evidence-Based Nursing*, *18*(3), 66–67. https://doi.org/10.1136/EB-2015-102129

Huynh, N. T., & Sharpe, B. (2013). An assessment instrument to measure geospatial thinking expertise. *Journal of Geography*, *112*(1), 3–17. https://doi.org/10.1080/00221341.2012.682227

Jo, I., & Bednarz, S. (2009). Evaluating geography textbook questions from a spatial perspective: Using concepts of space, tools of representation, and cognitive processes to evaluate spatiality. *Journal of Geography*, *108*(1), 4–13. https://doi.org/10.1080/00221340902758401

Kerski, J. J. (2008). The role of GIS in digital earth education. *International Journal of Digital Earth*, *1*(4), 326–346. https://doi.org/10.1080/17538940802420879

Köroğlu, M. N., Karaca, M., & Bektaş, O. (2023). Maddenin yapısı ve özellikleri konusunda başarı testi geliştirme: Geçerlik ve güvenirlik çalışması. *Anadolu Üniversitesi Eğitim Fakültesi Dergisi*, *7*(1), 1–32. https://doi.org/10.34056/aujef.997764

Lee, J., & Bednarz, R. (2009). Effect of GIS learning on spatial thinking. *Journal of Geography in Higher Education*, *33*(2), 183–198. https://doi.org/10.1080/03098260802276714

Lee, J., & Bednarz, R. (2012). Components of spatial thinking: Evidence from a spatial thinking ability test. *Journal of Geography*, *111*(1), 15–26. https://doi.org/10.1080/00221341.2011.583262

Lorenzo-Seva, U., & Ferrando, P. J. (2021). *Factor* (Version 12.04.05) [Computer software]. Universitat Rovira i Virgili. https://psico.fcep.urv.cat/utilitats/factor/

Montello, D. R. (1993). Scale and multiple psychologies of space. In A. U. Frank & I. Campari (Eds.), *Spatial information theory: A theoretical basis for GIS* (pp. 312–321). Springer. https://doi.org/10.1007/3-540-57207-4_21

Montello, D. R., Grossner, K., & Janelle, D. G. (2014). *Space in mind: Concepts for spatial learning and education*. The MIT Press.

National Research Council. (2006). *Learning to think spatially: GIS as a support system in the K–12 curriculum*. National Academies Press. https://doi.org/10.17226/11019

Newcombe, N., & Dubas, J. S. (1992). A longitudinal study of predictors of spatial ability in adolescent females. *Child Development*, *63*(1), 37–46. https://doi.org/10.1111/j.1467-8624.1992.tb03593.x

Olsson, U. (1979). Maximum likelihood estimation of the polychoric correlation coefficient. *Psychometrika*, *44*(4), 443–460. https://doi.org/10.1007/BF02296207/

Özçelik, D. (2010). *Test hazırlama kılavuzu*. Pegem Akademi.

Özkılıç, G., Bektaş, O., & Karaca, M. (2023). Sindirim sistemi ünitesine yönelik başarı testi geliştirme: Geçerlik ve güvenirlik çalışması. *Araştırma ve Deneyim Dergisi*, *8*(1), 115–154. https://doi.org/10.47214/adeder.1267535

Pallant, J. (2007). *SPSS survival manual: A step-by-step guide to data analysis using SPSS for Windows* (3rd ed.). Open University Press.

Salar, R., & Uğurel, E. (2020). Developing a measurement tool for the tenth class physics course electricity unit and determining the prior knowledge of the students. *Necatibey Eğitim Fakültesi Elektronik Fen ve Matematik Eğitimi Dergisi*, *14*(1), 217–239.

Şanlı, C. (2021). Mekansal düşünme beceri testinin geliştirilmesi. *International Journal Of Eurasia Social Sciences (IJOESS)*, *12*(43), 1–18. https://doi.org/10.35826/ijoess.2858

Self, C. M., & Golledge, R. G. (1994). Sex-related differences in spatial ability: What every geography educator should know. *Journal of Geography*, *93*(5), 234-243. https://doi.org/10.1080/00221349408979727

Self, C. M., Gopal, S., Golledge, R. G., & Fenstermaker, S. (1992). Gender-related differences in spatial abilities. *Progress in Human Geography*, *16*(3), 315–342. https://doi.org/10.1177/030913259201600301

Sinton, D. S. (2011). Spatial thinking. In J. P. Stoltman (Ed.), *21st century geography: A reference handbook* (pp. 733–745). Sage Publications.

Sinton, D. S., Bednarz, S. W., Gersmehl, P., Kolvoord, R. A., & Uttal, D. H. (2013). *The people's guide to spatial thinking*. National Council for Geographic Education.

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

Taşdemir, Ö. (2022). *Öğretmenler için üstün yetenekliliğe eşlik eden öğrenme güçlüğü aday gösterme ölçme aracı geliştirme çalışması* [Doctoral dissertation]. Yükseköğretim Kurulu Ulusal Tez Merkezi.

Uebersax, J. (2015). *Introduction to the tetrachoric and polychoric correlation coefficients*. https://www.john-uebersax.com/stat/tetra.htm

Uttal, D. H., Miller, D. I., & Newcombe, N. S. (2013). Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science*, *22*(5), 367-373. https://doi.org/10.1177/0963721413484756

Verma, K. (2014). *Geospatial thinking of undergraduate students in public universities in the United States* (Doctoral dissertation, Kent State University). ProQuest Dissertations & Theses Global. https://www.proquest.com/docview/1547037348

Yurt, E., & Tünkler, V. (2016). A study on the spatial abilities of prospective social studies teachers: A mixed method research. *Educational Sciences: Theory & Practice*, *16*, 965-986. https://doi.org/10.12738/estp.2016.3.0324

Witkin, H. (1950). Perception of the upright when the direction of the force acting on the body is changed. *Journal of Experimental Psychology*, *40*(1), 93–106. https://doi.org/10.1037/h0055162

Yılmazer, U. (2012). *Eğitimde ölçme ve değerlendirme*. Teorem Yayıncılık.