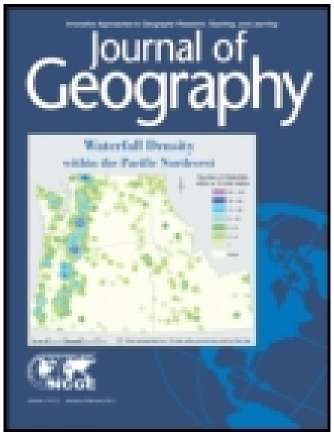

Journal of Geography
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/rjog20

Components of Spatial Thinking: Evidence from a Spatial Thinking Ability Test

Jongwon Lee [a] & Robert Bednarz [b]
[a] Ewha Womans University , Seoul , South Korea
[b] Texas A & M University , College Station , Texas , USA
Published online: 27 Dec 2011.

Components of Spatial Thinking: Evidence from a Spatial Thinking Ability Test

Jongwon Lee and Robert Bednarz

Abstract

This article introduces the development and validation of the spatial thinking ability test (STAT). The STAT consists of sixteen multiple-choice questions of eight types. The STAT was validated by administering it to a sample of 532 junior high, high school, and university students. Factor analysis using principal components extraction was applied to identify underlying spatial thinking components and to evaluate the construct validity of the STAT. Spatial components identified through factor analysis only partly coincided with spatial concepts used to develop the questions that compose the STAT and with the components of spatial thinking hypothesized by other researchers.

Key Words: *spatial thinking, spatial thinking ability test (STAT), factor analysis*

Dr. Jongwon Lee is an assistant professor of social studies education (geography education major) at Ewha Womans University, Seoul, South Korea. His research interests are in the areas of geography education: spatial cognition and skills, especially the role of geospatial technology in spatial thinking.

Robert Bednarz is professor of geography, Texas A & M University, College Station, Texas, USA, and North American Commissioning Editor of the Journal of Geography in Higher Education. His recent research has focused on spatial thinking, the impact of using geospatial technologies on spatial thinking skills, assessment of spatial thinking skills, and the implementation of geospatial technologies in science and geoscience curricula.

Introduction

Spatial thinking has been actively investigated during the last decade, especially with respect to its relationship to geospatial technologies and its relevance to problem solving in everyday life, the workplace, and science (Albert and Golledge 1999; Battersby, Golledge, and Marsh 2006; Bednarz 2004; Golledge 2002; Marsh, Golledge, and Battersby 2007). However, long before researchers began to focus on spatial thinking, psychologists and others sought to identify and measure spatial ability. Spatial ability—typically defined as spatial perception, visualization, and orientation—is seen as a narrower concept than spatial thinking (Committee on Support for Thinking Spatially 2006). It is beyond the scope of this article to provide a comprehensive review of the literature concerning the differences and distinctions between spatial ability, spatial reasoning, spatial cognition, spatial concepts, spatial intelligence, and environmental cognition. *Learning to Think Spatially,* published by the National Research Council, while recognizing that no clear consensus as yet exists concerning spatial thinking, provided a significant step toward understanding its nature and its importance in the school curriculum. The Committee (26) saw spatial ability as "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. The concept derives in part from the psychometric tradition of intelligence measurement and testing . . . " The Committee viewed spatial thinking, on the other hand, as a constructive amalgam of three mutually reinforcing components: the concept of space, tools of representation, and processes of reasoning. In order for individuals to conceptualize space, understand representations, and reason spatially, they must possess the appropriate spatial skills (Committee on Support for Thinking Spatially 2006).

The Committee (2006) also recognized the educational value of spatial thinking, arguing that it can be taught and learned; thus spatial thinking should be an important part of the educational curriculum at all levels. The Committee further suggested that GIS and other geospatial technologies can play a powerful role in promoting spatial thinking. In fact, many studies have pointed to the advantage of integrating GIS into the classroom (e.g., Allen 2007; DeMers and Vincent 2007; Doering and Veletsianos 2007; Milson and Earle 2007; Patterson, Reeve, and Page 2003) and have shown explicit links between GIS learning and students' spatial thinking skills (Kerski 2008; Lee and Bednarz 2009; Schultz, Kerski, and Patterson 2008).

However, researchers have also argued that "to be most effective, GIS teaching and curriculum development strategies should begin with an assessment of student understanding of spatial relationships. . . " (Wigglesworth 2003, 282), emphasizing the importance of establishing viable spatial thinking assessment based on a scientifically rigorous definition (Eliot and Czarnolewski 2007). Unfortunately, such a standardized measure of essential knowledge and skills does not exist. In fact, the Committee stated explicitly that "[t]here are neither content standards nor valid and reliable assessments for spatial thinking" (Committee on the Support for Thinking Spatially 2006, 232).

This article begins with a brief discussion of concepts of spatial thinking skills and the instruments available to measure them. Next, the article presents the development and validation procedures of the spatial thinking ability test

(STAT) that is modeled after the spatial skills test (Lee and Bednarz 2009). Data are presented that support the validity and reliability of STAT based upon a field test of 532 junior high, high school, and university students. The differences in the performance of these three levels of students are explored and tested for significance using ANOVA. In addition, factor analysis is applied to identify underlying spatial thinking components, to determine if the identified components support the structure of spatial thinking proposed by other researchers, and to evaluate the construct validity of the STAT.

Concepts of Spatial Thinking

Because the term "spatial thinking" has been used in both nonacademic and academic areas extensively, a variety of definitions exist (Committee on the Support for Thinking Spatially 2006; Eliot and Czarnolewski 2007; Gersmehl 2005; Gersmehl and Gersmehl 2006; 2007; Golledge and Stimson 1997; Harris 1981; Marsh, Golledge, and Battersby 2007; Montello *et al.* 1999). In addition, substantial disagreement continues to occur about the scale (from tabletop scale to geographic scale) and dimensions (thinking in, about, and with space) of spatial thinking, about the nature of cognitive processes involved, about the number of major components, and, as noted in the introduction, about the relationship, if any, between spatial ability and spatial thinking.

A few studies provide valuable input for the development of spatial thinking assessments. These studies suggest a series of spatial thinking concepts and describe differences between expert and novice performance in spatial thinking. Some of the most useful studies include *Learning to Think Spatially* (Committee on the Support for Thinking Spatially 2006), Gersmehl's (2005) spatial thinking taxonomy, and Golledge and others' categorization of geospatial concepts (Battersby, Golledge, and Marsh 2006; Golledge 2002; Golledge and Stimson 1997; Golledge, Marsh, and Battersby 2008; Marsh, Golledge, and Battersby 2007).

Learning to Think Spatially (Committee on the Support for Thinking Spatially 2006) introduces three spatial contexts: life space (cognition in space), physical space (cognition about space), and intellectual space (cognition with space). The first of these involves thinking about the world in which we live. It often includes way-finding and navigation in the real, geographic world. It also includes other everyday activities including assembling it furniture by following instructions, packing the trunk of a car to maximize carrying capacity, etc. The second context, cognition about space, focuses on "a scientific understanding of nature, structure, and function of phenomena that range from the microscopic to the astronomical scales" (30). It is useful in explaining the structure of the atom or DNA, the movement and arrangement of the elements of the solar system, etc. Other examples include "shapes and structures of urban areas, the diffusion of cultures and agriculture, or the organization of the world economy" (Bednarz n.d.). The concept or object investigated through the third context is not necessarily spatial but can be spatialized by time-space coordination. For example, written linguistic symbols are spatially defined and spatially arranged, and readers must establish word order so that sentence and passage meaning can be determined. Patterns in complex numerical data can often be revealed and best understood by portraying the information graphically.

Although *Learning to Think Spatially* (Committee on the Support for Thinking Spatially 2006) provides multicontextual and interdisciplinary definitions of spatial thinking, it has been criticized for its lack of a conceptual framework, an essential prerequisite for development of assessment tools (e.g., Gersmehl and Gersmehl 2006). Previous research is not devoid of conceptual frameworks, however, as Gersmehl and Gersmehl (2006, 2007) and Golledge and others (Battersby, Golledge, and Marsh 2006; Golledge 1995, 2002; Golledge and Stimson 1997; Golledge, Marsh, and Battersby 2008; Marsh, Golledge, and Battersby 2007) have proposed hierarchies of spatial thinking skills and concepts. In a study to specify a taxonomy of spatial thinking, Gersmehl and Gersmehl (2006, 2007) defined spatial thinking as skills that geographers use to analyze the spatial relationships in the world. They identified thirteen modes of spatial thinking: defining a location, describing conditions (the geographic concept of site), tracing spatial connections (situation), making a spatial comparison, inferring a spatial aura (influence), delimiting a region, fitting a place into a spatial hierarchy, graphing a spatial transition, identifying a spatial analog, discerning spatial patterns, assessing a spatial association, designing and using a spatial model, and mapping spatial exceptions. They argued that brain research suggests that these modes of spatial thinking have distinct or independent neurological foundations. They offered no empirical evidence or other rigorous assessment to support their hypothesis that the modes they identified are independent, however.

A hierarchical set of spatial thinking concepts was proposed by Golledge and his colleagues (2008). This sequence progresses from four basic spatial concepts (or primitives) to more complex and abstract concepts through five different levels as follows (Golledge, Marsh, and Battersby 2008, 91–92): (1) primitive level (identity, location, magnitude, space-time); (2) simple level (arrangements, distribution, line, shape, boundary, distance, reference frame, sequence); (3) difficult level (adjacency, angle, classification, coordinate, grid pattern, polygon); (4) complicated level (buffer, connectivity, gradient, profile, representation, scale); and (5) complex level (area association, interpolations, map projection, subjective space, virtual reality). The spatial thinking skills suggested by the Gersmehls (2006, 2007) and geospatial concepts proposed by Golledge, Marsh, and Battersby (2008) share the common context of the geographic scale. However, whereas Gersmehls' concepts are related to geographic analysis, the geospatial concepts

Table 1. Core concepts of spatial thinking suggested by Gersmehl and Gersmehl (2007), Golledge, Marsh, and Battersby (2008), and Janelle and Goodchild (2009).

Gersmehl and Gersmehl (2007)	Golledge *et al.* (2008)	Janelle and Goodchild (2009)
Condition	Identity	Objects and Fields
Location	Location	Location
Connection	Connectivity	Network
	Distance	Distance
	Scale	Scale
Comparison	Pattern Matching	
Aura	Buffer	
Region	Adjacency, Classification	Neighborhood and Region
Hierarchy		
Transition	Gradient, Profile	
Analogy		
	Coordinate	
Pattern	Pattern, Arrangement, Distribution, Order, Sequence	
Spatial Association	Spatial Association, Overlay/Dissolve, Interpolation	Spatial Dependence, Spatial Heterogeneity
	Projection, Transformation	

identified by Golledge and his colleagues are intended primarily to address the functions of GIS.

The present investigation, using experiments that began in 2006, could not benefit from the most recent work concerning hierarchical geospatial ontology. However it did incorporate key spatial thinking concepts from several studies conducted by Golledge (1992, 1995, 2002). Along with Gersmehls' spatial thinking taxonomy, Golledge's list of geographic thinking elements presented in 2002 guided the development of the spatial thinking ability test on which this study is based. The following list from Golledge's 2002 study specifies the spatial thinking elements he thought were important and illustrates the ideas and concepts his work shares with Gersmehls' (2005):

> Comprehending spatial association (positive and negative); comprehending spatial classification (regionalization); comprehending spatial change and spatial spread (spatial diffusion); comprehending non-spatial and spatial hierarchy; comprehending spatial shapes and patterns; comprehending locations and places; comprehending integration of geographic features represented as points, networks, and regions; comprehending spatial closure (interpolation); and recognizing spatial form. (Golledge 2002, 4–6)

The core concepts of spatial thinking from three recent important sources including Gersmehl and Gersmehl (2007), Golledge, Marsh, and Battersby (2008), and Janelle and Goodchild (2009) are summarized in Table 1. Although the terms and number of core concepts that they used are different, it is not difficult to find similarity among them.

Measurement of Spatial Thinking Abilities

A variety of psychometric tests (Clements *et al.* 1997; Dean and Morris 2003; Hall-Wallace and McAuliffe 2002) have been widely used to measure individuals' spatial abilities, especially in psychological research. However, psychometric measures are limited to the assessment of psychologically and narrowly defined spatial abilities rather than spatial thinking as defined by the Committee (2006) (Hegarty *et al.* 2002; Lee and Bednarz 2009). Consistent with this view, Eliot and Czarnolewski (2007, 362) argued that "researchers need to go beyond the limits of existing spatial tests and consider the possibility that spatial intelligence is a more encompassing construct of human activities . . . ."

Self-assessment questionnaires are believed to assess broader aspects of spatial thinking (Hegarty *et al.* 2002), and there are a few examples of these available on the Web (e.g., Golledge 2000, 2001). A typical question from these instruments might ask people to rate on a five- or seven-point scale a statement such as, "When traveling, I take shortcuts as frequently as possible." Although researchers have found that self-report measures are capable of assessing spatial skills on both the small (or pictorial) and large (or environmental) scales (Hegarty *et al.* 2002) and are useful in assessing individuals' spatial behaviors in everyday life, they are more appropriate for classifying types of spatial behavior than determining levels of spatial ability. Another shortcoming of subjective self-report measures is that often the results from different instruments are incomparable.

There have been some important attempts to measure specific aspects of spatial thinking skills (e.g., Albert and Golledge 1999; Battersby, Golledge, and Marsh 2006; Gilmartin and Patton 1984; Golledge 1992; Kerski 2000; Lloyd and Bunch 2003). For example, Golledge (1992) investigated how completely people understand spatial concepts such as "nearest neighbor" using a map-based laboratory experiment. Battersby, Golledge, and Marsh (2006) devised a task assessing individuals' understanding and ability to apply one of the most essential GIS functions—map overlay. In that study, participants were provided two maps of the same area and asked to derive conclusions about the spatial relationships. In order to answer the questions properly, participants had to combine two thematic layers of information and perform logical functions (i.e., Boolean logic). In a similar study, Albert and Golledge (1999) used a simplified set of thematic layers to evaluate how well GIS users could select appropriate map layers and operations and visually verify map overlay processes to achieve a specific result.

Another type of spatial task used to assess individuals' spatial ability includes map-reading skills such as following directions, judging distances, comprehending geographic characteristics, and recognizing patterns (Carswell 1971; Gilmartin and Patton 1984). Map-reading skills devised by Gilmartin and Patton (1984) provided students with representations of a country's population distribution, topography, and climate that they then used to answer multiple-choice questions such as "Which of the country's three major cities has the largest population?" Also included in the same study was a road map-reading task to assess abilities such as distance estimation, route comparison (e.g., visually compare two straight-line distances and judge which is shorter), and pattern recognition (e.g., choose which of four generalized diagrams best represents the overall road pattern in the study area). Finally, Kerski (2000) created a task that assessed both the spatial concept, "best location," and map reading simultaneously. He asked students to analyze geographic information and select the best location for a fast food restaurant in a hypothetical area based on a given set of variables including traffic volume, existing fast food locations, locations of high schools, and annual median income.

Development of the Spatial Thinking Ability Test (STAT)

One goal of the present study was to develop a standardized test of spatial thinking abilities (the spatial thinking ability test (STAT)) that integrates geography content knowledge and spatial skills. Currently no standardized instrument for assessing the set of spatial thinking skills discussed previously exists. In addition, the reviewed studies using questionnaires or other measures to assess spatial skills often ignored issues of reliability and validity.

The current study extends the authors' research (Lee and Bednarz 2009) that developed and deployed spatial skills tests (SST) to measure changes in students' spatial skills after they completed GIS coursework. That research found a significant relationship between the completion of one or more geospatial technology courses and students' scores on the spatial skills test. The components of spatial relations as defined by Golledge and Stimson (1997) provided guidelines for developing test items. The spatial skills tests consist of a set of multiple-choice questions and performance tasks that were designed to evaluate students' skills including overlaying and dissolving a map, reading a topographic map, evaluating several factors to find the best location, recognizing spatially correlated phenomena, constructing isolines based on point data, and differentiating among spatial data types.

The initial motivation to revise and augment the original spatial skills test was to measure students' mastery of the content and skills contained in the Association of American Geographers' *Teachers' Guide to Modern Geography* (TGMG) project materials. The primary aim of the TGMG, funded by the U.S. Department of Education, is to improve the preparation and ability of geography teachers to incorporate spatial thinking skills into their classes. The TGMG project produced a variety of print and digital materials for preservice and in-service teacher preparation programs, for example, a multimedia CD with animated instructional units that deal with the analytical skills specified in the *National Geography Standards*, such as measuring direction, distance, slope, and density; analyzing map patterns and making rigorous map comparisons; formulating and testing hypotheses; identifying exceptions to patterns predicted by hypotheses; and buffering, overlaying, windowing, and other methods of spatial analysis. The spatial thinking ability test (STAT) was designed to assess individuals' growth in spatial thinking skills and to help determine the effectiveness of the TGMG materials in promoting the spatial thinking skills of teachers. The revised and expanded spatial skills test also provided a data set that can be used to provide a preliminary assessment of the reliability and validity of the previously noted spatial thinking conceptualizations proposed by other researchers.

The initial step in the construction of the STAT was the delineation of the assessment objective and the description of the test contents to be measured. Two sets of spatial thinking concepts were analyzed and combined and served to inform the development of STAT. The first set of concepts was identified by Gersmehl (2005) whose ideas served as the theoretical foundation of the TGMG project. The second set was comprised of Golledge's (2002) list of spatial thinking skills, which played a key role in the development of the original spatial thinking ability test. Golledge's concepts were especially useful because they were detailed enough to develop test items, potentially leading to improvement of test content validity. Additionally, both lists share some common concepts and features as noted previously.

Each test item was designed to measure one or two components of spatial thinking identified by one or both of these two studies. The aspects of spatial thinking abilities covered by STAT include: (1) comprehending orientation and direction; (2) comparing map information to graphic information; (3) choosing the best location based on several spatial factors; (4) imagining a slope profile based on a topographic map; (5) correlating spatially distributed phenomena; (6) mentally visualizing 3-D images based on 2-D information; (7) overlaying and dissolving maps; and (8) comprehending geographic features represented as point, line, or polygon (see Table 2).

During the development of STAT, we focused on a central problem related to test construction: how to ensure practicability while at the same time providing

Table 2. Description of question types and spatial thinking components to measure.

Type (Item Number)	Item Description	Spatial Thinking Components to Measure
I (#1, #2)	In order to solve item #1 and #2, participants should visually navigate road maps using verbal information including participant's current location, directions to destination, street information, etc. (See Fig. 1)	Item #1 and #2 evaluate the trait of "comprehending orientation and direction (e.g., forward-backward; left-right; up-down; back-front; horizontal-vertical; north/south/east/west)" (Golledge 2002).
II (#3)	In order to solve item #3, participants should recognize map patterns and represent them in graphic form.	Item #3 assesses the trait of "discerning spatial patterns" (Gersmehl 2005) and "graphing a spatial transition" (Gersmehl 2005).
III (#4)	In order to solve item #4, participants should select an ideal location for a fictitious facility based on multiple pieces of spatial information such as land use, elevation, population density, etc.	The basic rationale behind item #4 is to assess the trait "comprehending overlay and dissolve" (Golledge 2002) and "inferring a spatial aura (influence)" (Gersmehl 2005).
IV (#5)	In order to solve item #5, participants should create a profile of topography along a proposed line on a contour map. In addition, the participants need to properly orient themselves in situ.	In solving item #5, participants deal with several cognitive traits including "recognizing spatial form (such as cross-sections to three-dimensional block diagrams or image)" (Golledge 2002), "being able to transform perceptions, representations and images from one dimension to another and the reverse" (Golledge 2002) and "graphing a spatial transition" (Gersmehl 2005).
V (#6, #7)	In order to solve item #6, participants should identify spatial correlations between sets of maps. Additionally, item #7 asks participants to display the identified spatial relationship in a graphic form. (See Fig. 1)	Item #6 and #7 evaluate the trait "comprehending spatial association (positive and negative)" (Golledge 2002), "making a spatial comparison" (Gersmehl 2005), and "assessing a spatial association" (Gersmehl 2005). Item #7 additionally assesses the trait of "graphing a spatial transition" (Gersmehl 2005).
VI (#8)	In order to solve item #8, participants need to mentally visualize a 3-D image based on 2-D information. (See Fig. 1)	Item #8 assesses the trait of "being able to transform perceptions, representations and images from one dimension to another and the reverse" (Golledge 2002).
VII (#9, #10, #11, #12)	In order to solve item #9, #10, #11, and #12, participants should visually verify a map overlay process and then select the appropriate map layers involved in the overlay. (See Fig. 1)	Item #9, #10, #11, and #12 correspond to the trait "overlaying and dissolving maps" (Golledge 2002).

(Continued on next page)

maximum comprehensibility of spatial thinking concepts. A number of other factors were also considered in the design of the STAT. These factors included (1) cognitive process (i.e., maximizing spatial processes and minimizing verbal processes); (2) psychometric rationale; (3) mode of representation (text, picture, graph, map, color versus black and white, etc.); and (4) practical constraints (e.g., amount of time required to complete the test).

The current version of the test is fourteen pages long and has two equivalent forms (one that can be used for a pretest and one for a post-test) allowing for the evaluation of changes in spatial thinking skills over a period of time. The pre- and post-tests were composed of slightly different questions covering the same spatial thinking skills. Each form, containing sixteen multiple-choice questions, consists of eight different types of questions (Table 2). Figure 1a and 1b contain a sample of items from the STAT. We also constructed a three-item questionnaire to collect information about the subject's gender, academic major (geography major or not), and amount of geospatial coursework completed (e.g., GIS and cartography).

Formal and informal review of STAT took place in a variety of venues, mostly conducted as part of the evaluation plan of the TGMG project. After a draft of STAT was completed, all items were carefully reviewed by a team of experts consisting primarily of the TGMG project team and steering committee members. The team included two individuals who teach geographic education courses for undergraduate and graduate students; these members conducted an informal review and then pilot-tested the instrument with their students. Twenty-seven undergraduate students participated in a pilot test that helped estimate the difficulty level of STAT and reduce the incidence of errors in test administration.

Table 2. Description of question types and spatial thinking components to measure. (*Continued*)

Type (Item Number)	Item Description	Spatial Thinking Components to Measure
VIII (#13, #14, #15, #16)	In order to solve item #13, #14, #15, and #16, participants should visually extract types of spatial data from verbally expressed spatial information. (See Fig. 1)	Item #13, #14, #15, and #16 measure the trait “comprehending integration of geographic features represented as points, networks, and regions” (Golledge 2002) and “comprehending spatial shapes and patterns” (Golledge 2002).

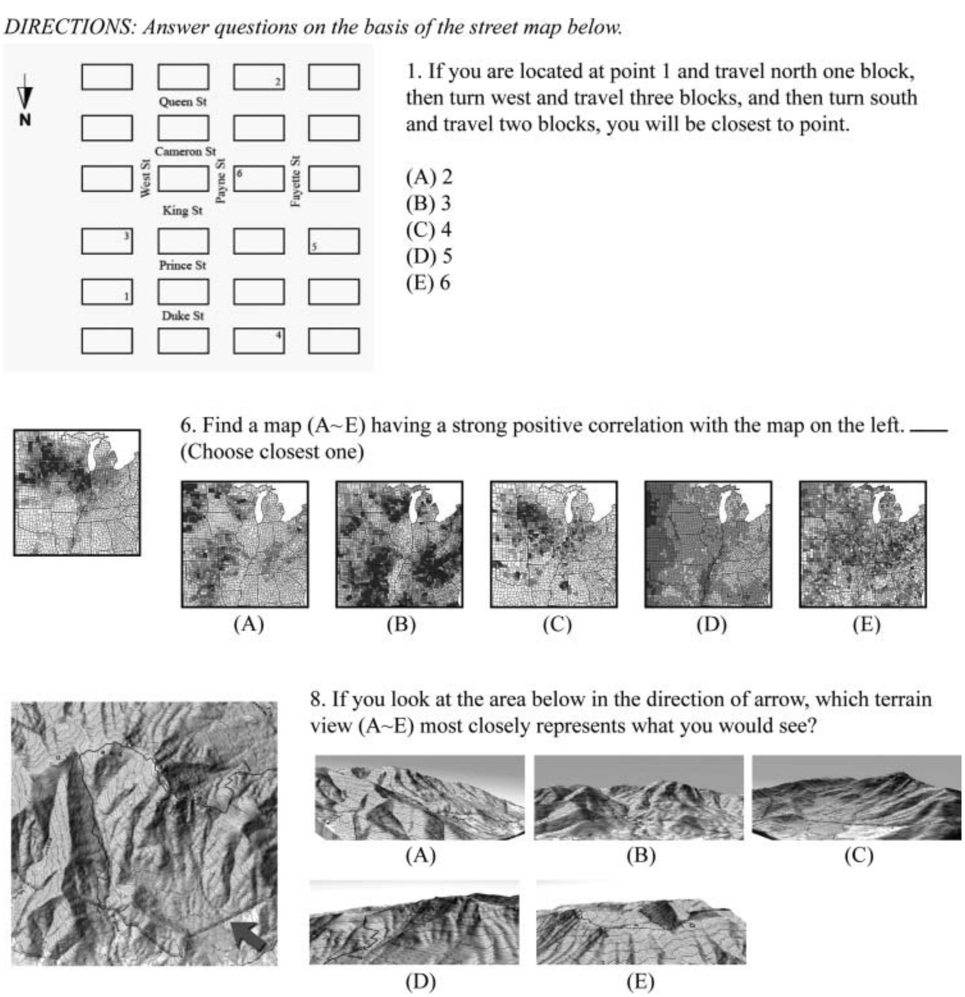

Figure 1a. Selected items from the STAT. Each item corresponds to Type I, V, and VI, respectively.

Revisions were made as a consequence of the pilot tests: questions that were perceived to have more than one correct answer or were difficult to score objectively were eliminated or revised; a pair of questions (for pre- and post-test) that proved to have different levels of difficulty were adjusted; and items whose graphics or directions were unclear were improved.

When workshops to review the TGMG project materials were scheduled during a variety of geography meetings and conferences, STAT review sessions were also conducted. For example, a STAT review session occurred in conjunction with the annual meeting of the National Council for Geographic Education. During that session twenty-two TGMG workshop participants including a preservice teacher, education students, secondary school geography teachers, and professional geographers, took and commented on the STAT.

Test Results

Reliability and Construct Validity of STAT

Test results from 352 university students from four different U.S. states who took STAT were used to examine the reliability and validity of STAT. The number of students who completed the tests at the four universities varied from 11 to 146. The variation of the sample sizes resulted from the access to students by faculty volunteers who agreed to administer STAT at each of the schools. As a measure of internal consistency, Cronbach’s alpha was calculated. Cronbach’s alpha is a measure of the intercorrelation of items, measuring the extent to which item responses obtained at the same time correlate with each other. A value of 0.7 for Cronbach’s alpha is generally considered to indicate a reliable set of items (de Vaus 2002). The Cronbach’s alpha for the latest version was 0.721 and 0.701 for Forms A and B,

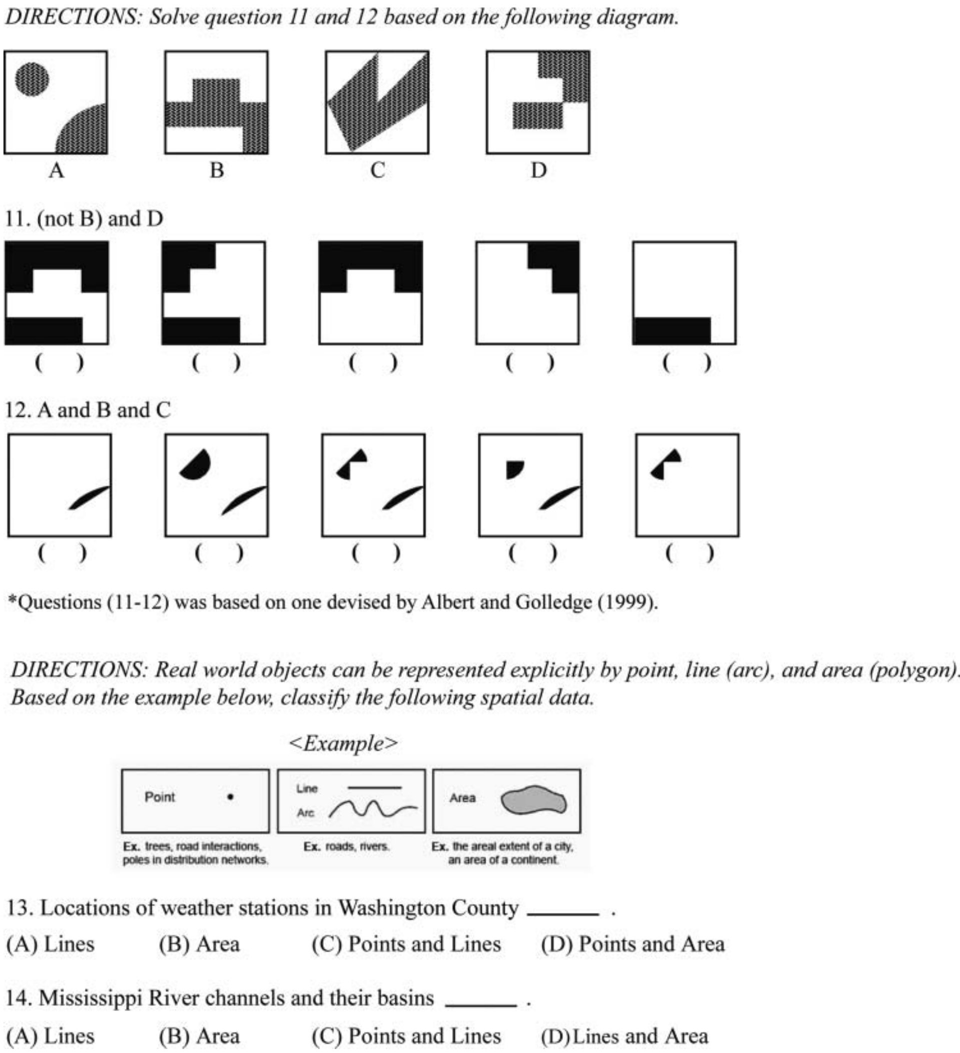

Figure 1b. Selected items from the STAT. Each item corresponds to Type VII and VIII, respectively. *Note*: all items of STAT may be viewed at http://home.ewha.ac.kr/~ziriboy/STAT.pdf.

respectively. When STAT was administered for the first time, we were somewhat disappointed by the relatively low validity and reliability statistics. As we reconsidered the results, however, we realized that recent conceptualizations of spatial thinking skills support the notion that spatial thinking skills are composed of several elements that may be at least somewhat independent of one another. Therefore, it is not surprising that some individuals might perform significantly differently on questions that assessed different skills thereby lessening the internal consistency or intercorrelation. Although originally Golledge and Stimson (1997) proposed "spatial relations" as an additional spatial ability to visualization or orientation, it seems likely that "spatial relations" included a variety of skills that are likely uncorrelated.

In order to explore to what extent spatial thinking skills are composed of distinct components, the construct validity of STAT was examined using factor analysis. Factor analysis is a statistical technique used to identify the minimal underlying factors needed to explain the intercorrelations among the test items. Principal components analysis revealed six factors with eigen-values of 1.0 or more, accounting for 54.66 percent of cumulative variance. In general, a factor analysis accounting for 60–70 percent or more of the total variance is considered a good fit to the data. Varimax rotation was then applied to the six factors. This procedure rotates the set of individual scores within the space defined by principal component axes, thereby creating a new set of factor loadings (increasing the difference between high and low loadings) for the factors that have already been found. The rotated factor matrix is presented in Table 3. The nature of each of the factors in Table 3 is determined by the characteristics of the variables that have high loadings on these factors. Six factors accounted for 11.1, 10.7, 10.5, 7.9, 7.6, and 6.9 percent of the variance, respectively.

If the skills tested by the eight question types displayed in Table 2 are independent components of spatial thinking, we would expect the factor analysis to yield factors that reflect those components. That is, questions that assess a specific component should be grouped. Although some factors directly or indirectly show high levels of correspondence, others do not. For example, four of the questions that load on factor 2, items #13, #14, #15, and #16, are based on the spatial skill "visually extract types of spatial data from verbally expressed spatial information" (Type VIII) although #10, which is not a Type VIII item, also loads as highly on factor 2 (0.484). Factor 3 generates heavy loadings for the items related to question type VII (#9, #10, #11, and #12) requiring "participants to visually verify a map overlay process and then select the appropriate map layers involved in the overlay." It is interesting that, in addition to the type VII questions, item #7 had the high loading (0.529) on this factor. This result might be explained

Table 3. Results of the factor analysis of the pretest of the STAT.

Question Item	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6
#4	.742	.126	.067	.236	−.110	.063
#5	.682	.134	.110	.118	.052	−.237
#12	.560	−.060	.189	−.190	.084	.235
#14	.407	.399	.074	.020	.130	.271
#15	.089	.773	.002	.042	−.016	.156
#13	.054	.615	.041	.351	−.189	−.153
#10	−.032	.484	.470	.087	.343	−.072
#16	.250	.451	.383	−.275	.195	.048
#9	.105	.113	.694	−.022	.157	.164
#11	.063	.065	.674	.262	−.290	.071
#7	.299	−.094	.529	.043	.024	−.280
#3	.119	.118	.049	.687	.042	−.023
#2	.040	.027	.166	.591	.243	.388
#1	.024	−.039	.020	.101	.857	.039
#6	.066	.103	.033	.104	.015	.695
#8	.292	.207	.044	.195	.288	−.332
Eigen values	1.781	1.705	1.676	1.265	1.218	1.101
% of variance	11.129	10.659	10.472	7.908	7.613	6.879
Cumulative %	11.129	21.788	32.260	40.168	47.781	54.660

by the existence of similar cognitive processes in solving type VII ("visually verify map overlay processes") and item #7 ("comprehending spatial association").

However, unlike factors 2 and 3, other factors were not clearly connected to specific question types. Factor 1, with heavy loadings for items #4, #5, #12, and #14, spans several question types. Furthermore, some items load equally on more than one factor such as item #8, which loads on four factors—1 (0.292), 2 (0.207), 4 (0.195) and 5 (0.288)—although at relatively low levels. This may occur partly because the question items are inadequately specified (i.e., they represent or require more than one spatial thinking skill) or because of the failure of STAT to capture the full range of spatial thinking.

Analysis of STAT Test Results

The STAT was administered to students at a wide range of academic levels—at four universities located in Texas, Ohio, Illinois, and Oregon, and at a junior high and high school in Ohio (Table 4).

In general, as students advanced from junior high to university their performance improved. For every question, the average score for high school students exceeded that for junior high school students. Similarly, the average scores for university students were greater than the scores for high school students for every question, although the average high school scores were greater for some questions than the average scores of university B and C students.

Analysis of the test data showed that a sizeable majority of university students could identify patterns on a map and choose a correct graphical display of a spatial pattern (item #3). In addition, a large percent of university students (63.6 to 93.1 percent) could find locations, understand orientations and directions, and navigate on road maps following directions (item #1 and #2). Whereas nearly 90 percent of the participants could comprehend spatial association between two maps (item #6), less than half were capable of transforming a spatial relationship into a graphic form (item #7). Although items #9, #10, #11, and #12 were designed to assess the same skill (Type VII), students performed most poorly on item #12 (see Fig. 1b), the question that required the most complex Boolean logic.

A few question items were found to have better discriminating power than other items. For instance, question items such as #1,

Table 4. A percentage of correct answers per item and mean score by groups.

Item	Univ. A (N = 29)	Univ. B (N = 11)	Univ. C (N = 59)	Univ. D (N = 146)	Junior High (N = 52)	High (N = 149)
#1	89.66	90.91	67.80	84.25	42.27	73.51
#2	93.10	63.64	77.97	88.36	43.75	67.58
#3	96.55	81.82	91.53	95.21	46.32	79.78
#4	65.52	45.45	54.24	70.55	24.21	38.46
#5	65.52	45.45	54.24	79.45	19.35	29.89
#6	96.55	100.00	83.05	94.52	55.32	71.27
#7	65.52	27.27	18.64	42.47	22.34	24.73
#8	51.72	27.27	28.81	44.52	18.48	24.73
#9	68.97	36.36	52.54	60.27	26.37	30.34
#10	79.31	54.55	67.80	82.19	31.91	60.44
#11	65.52	27.27	49.15	58.90	15.22	33.71
#12	34.48	18.18	22.03	32.19	15.22	24.72
#13	93.10	45.45	64.40	75.34	35.16	55.25
#14	68.97	36.36	55.93	67.12	30.53	37.36
#15	89.66	54.55	72.88	86.99	41.94	63.13
#16	68.97	9.09	45.76	69.18	22.83	36.11
Mean Score	11.93	7.64	9.07	11.32	4.60	7.58
(SD)	(2.64)	(3.67)	(2.79)	(2.82)	(2.47)	(2.76)

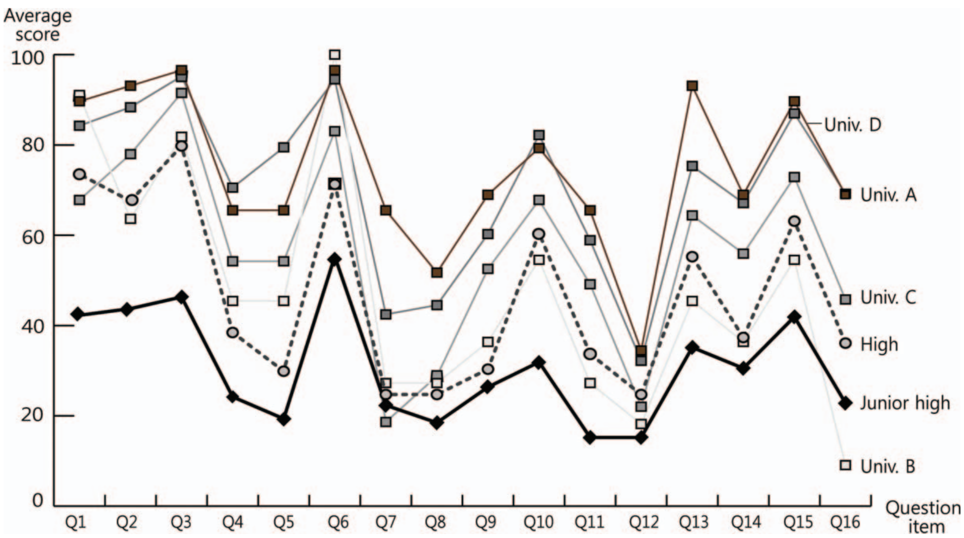

Figure 2. Score comparison per item by groups.

Analysis of variance (ANOVA) was used to determine whether significant differences in scores among four university groups existed. When analysis of variance (ANOVA) was performed on Form A scores of the four schools, significant differences were found ($p = .000$) (Table 5). *Post-hoc* comparisons (using Tukey method) revealed that students of university A and D scored significantly higher than those of university B and C (Table 6). No significant difference was found between scores of university A and D and between those of university B and C, respectively.

The number of geography majors in each group may account for the differences in scores (Table 7). School groups with the highest percentages of geography majors, university A (41.38% majors) and D (26.03%), scored higher on Form A than university B (9.09%) and C (16.95%). The format of STAT, including maps and spatial terms, may be more challenging or less familiar to nongeography majors.

#2, #3, #10, and #11 most clearly separate the junior high schools students from others (Fig. 2). Nevertheless, as the graph illustrates, students at all levels displayed similar performance patterns, in the sense that scores for all students were uniformly higher for some questions than others. The lines representing the scores for the six student samples are relatively parallel, indicating that students, from junior high to university, scored higher and lower on the same questions. This result would seem to indicate that some skills are more challenging than others and offer support for the argument that spatial thinking is composed of more than one skill or ability (in addition to the widely accepted spatial visualization and orientation abilities).

Because two versions of STAT were developed (Form A and Form B), it was important to verify that the two forms were equally difficult. This was achieved by comparing participants' mean scores on the two forms of the STAT using a t-test for independent samples. The student participants of university A, B, and C (students of university D took only Form A) were randomly divided into two groups each taking either Form A or Form B of the test. In order to verify the equivalency of two forms of the STAT, three separate t-tests were conducted for the three groups (Table 8). All of the t-tests indicated that the two forms of STAT are generally equivalent in difficulty except for university B where two students who completed Form A scored very low on the test (answering only two of sixteen questions correctly).

Table 5. ANOVA for Form A scores by groups (University students only).

	SS	*df*	MSE	*F*	*p*
Between groups	359.524	3	119.841	14.936	.000
Within groups	1933.643	241	8.023		
Total	2293.167	244			

Table 6. *Post-hoc* comparison of Form A scores by groups.

	Univ. A	Univ. B	Univ. C
Univ. B	4.295** (.000)		
Univ. C	2.863** (.000)	1.431 (.416)	
Univ. D	0.616 (.708)	−3.679** (.000)	−2.247** (.000)

**$p < .01$.

Discussion and Conclusions

In this study, we developed and evaluated standard measures of spatial thinking skills. Internal consistency reliability estimates for the STAT were in the moderate range. Although these results may raise concerns regarding the

Table 7. A comparison of Form A scores by major (geography major versus nongeography major).

	N	Mean	S.D.	t	p
Geography major	61	11.77	2.58	−3.264	0.067
Nongeography major	184	10.32	3.14		

Table 8. Independent sample t-test.

	Form	N	Mean	S.D.	t	p
University A	A	29	11.93	2.64	.908	.368
	B	26	11.23	3.08		
University B	A	11	7.64	3.67	−1.945	.065
	B	13	10.38	3.25		
University C	A	59	9.07	2.78	−0.771	.442
	B	47	9.49	2.81		

completeness of the measure, there are several issues that may account for the moderate reliability. With regard to internal consistency reliability, Cronbach's alpha increases as the number of items increases. *Ceteris paribus*, increasing the number of items can increase the level of alpha.

In terms of the construct validity, the factor analysis using principal components extraction with varimax methods provide mixed results with regard to the research hypothesis. We hypothesized that factor analysis would identify the independent components of spatial thinking by generating factors that reflected the eight components of previous researchers' spatial thinking conceptualizations that were represented by questions in STAT. Whereas some factors were directly or indirectly connected to the question types, some were not. This result might be attributed to the participants' styles of spatial problem solving. It is widely accepted that different people employ different strategies when solving spatial tasks (Kyllonen, Lohman, and Woltz 1984; Lohman and Kyllonen 1983). Furthermore, spatial tasks are often solved using nonspatial processing strategies. For instance, Just and Carpenter (1985) found that many spatial test items may also involve verbal analytic processing. They argued that verbal strategies are routinely employed for spatial tasks including 3D rotation, spatial orientation, and others. Thus, for at least some individuals, relative success on spatial items could be due to a verbal or another type of ability rather than this spatial ability. As mentioned previously, care was taken when developing STAT to maximize spatial processes and minimize verbal processes required to answer correctly. Because no information about how students solved questions was collected however, studies with additional items are needed to explore the spatial thinking processes employed by individuals engaged in spatial problem solving before this issue can be addressed reliably.

Perhaps another reason that the factor analysis did not identify eight independent components is that independence of the eight components is not as great or as complete as hypothesized. Spatial thinking skills may be comprised of fewer than eight components or some skills may be correlated to others, which may or may not be the same thing. If spatial thinking consists of fewer independent components, what do the results of the factor analysis suggest those components might be? Three of the four items loading highly on factor 1 require the skill to overlay or visualize spatial data. Four of the five items that load on factor 2 require the ability to distinguish among the map elements point, line, and area. Three of the items that load on factor 3 test respondents' skill in performing Boolean operations on geometric pattern; the fourth item requires identification of the nature of spatial correlation between two mapped distributions. The two items loading on factor 4 do not appear to have much in common: one concerns a way-finding task and the other requires the creation of a cross-section diagram from a mapped distribution. Factors 5 and 6 are, for the most part, comprised of one item, a way-finding question for factor 5 and identification of a positive spatial correlation for factor 6.

Thus, the analysis of STAT offers relatively little support for the existence of the independent spatial thinking components hypothesized in the literature. The analysis also suggests that Golledge and Stimson's spatial relations ability is almost certainly not a single ability but instead is comprised of a collection of different skills. Based on the clusters identified by the factor analysis, the following spatial thinking components emerge: map visualization and overlay, identification and classification of map symbols (point, line, area), generalized or abstract Boolean operations, map navigation or way-finding, and recognition of positive spatial correlation. We do not assert that these five components are *the* five spatial thinking skill sets. Nevertheless, intuitively these skills do seem different enough that individuals might be able to use one or more successfully while they are having difficulty with others. For example, it is not hard to believe that a person who is skilled at solving Boolean problems might not necessarily be a skilled navigator.

We do think that the analysis strongly supports the hypothesis that spatial thinking is a collection of different skills and that more work must be done to identify those component skill sets. The results also help explain why individuals perform well on some spatial thinking tasks while performing poorly at others. For geography educators these results suggest that because students perform well on some tasks does not mean that they will perform well on others. If different tasks require different skill sets, performance may be uneven. The results also suggest that giving students a variety of ways to demonstrate what they have learned might reveal a student's knowledge or ability that would go undetected if only one method of assessment is employed.

The two forms of STAT were equivalent in difficulty at baseline, and therefore, the two forms of STAT can be used for pre- and post-test designs to evaluate changes in spatial thinking abilities over a brief period. In addition, the field tests in several different environments showed STAT was useful for testing both university and high school students.

Our standardized measure needs to substantiate content validity more rigorously. This measure, however, provides rigorous bedrock for testing that can be expanded with new tests in the near future. The current version of STAT represents considerable developmental work based on a solid theoretical foundation. Additional research and refinement of the measures could strengthen their testing abilities and contribute to research on spatial literacy in the long run.

References

Albert, W. S., and R. G. Golledge. 1999. The use of spatial cognitive abilities in geographic information systems: The map overlay operation. *Transactions in GIS* 3 (1): 7–21.

Allen, T. R. 2007. Digital terrain visualization and virtual globes for teaching geomorphology. *Journal of Geography* 106 (6): 253–266.

Association of American Geographers. n.d. Teachers' Guide To Modern Geography. http://www.aag.org/cs/tgmg (last accessed November 2011).

Battersby, S., R. G. Golledge, and M. J. Marsh. 2006. Incidental learning of geospatial concepts across grade level: Map overlay. *Journal of Geography* 105 (4): 139–146.

Bednarz, S. W. n.d. Maps and spatial thinking skills in the AP human geography. *AP Central*. http://apcentral.collegeboard.com/apc/public/courses/teachers_corner/151317.html (accessed August 2010).

Bednarz, S. W. 2004. Geographic information systems: A tool to support geography and environmental education? *GeoJournal* 60 (2): 191–199.

Carswell, R. J. B. 1971. The role of the user in the map communication process: Children's abilities in topographic map reading. *Cartographica: The International Journal for Geographic Information and Geovisualization* 8 (2): 40–45.

Clements, D. H., M. T. Battista, J. Sarama, and S. Swaminathan. 1997. Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal* 98 (2): 171–186.

Committee on Support for Thinking Spatially. 2006. *Learning to Think Spatially*. Washington, D.C.: The National Academies Press.

Dean, G. M., and P. E. Morris. 2003. The relationship between self-reports of imagery and spatial ability. *British Journal of Psychology* 94 (2): 245–273.

DeMers, M. N., and J. S. Vincent. 2007. ArcAtlas in the classroom: Pattern identification, description, and explanation. *Journal of Geography* 106 (6): 277–284.

de Vaus, D. 2002. *Surveys in Social Research*, 5th ed. London: Routledge.

Doering, A., and G. Veletsianos. 2007. An integration of the use of real-time, authentic geospatial data in the K–1 classroom. *Journal of Geography* 106 (6): 217–225.

Eliot, J., and M. Y. Czarnolewski. 2007. Development of an everyday spatial behavioral questionnaire. *The Journal of General Psychology* 134 (3):361–381.

Gersmehl, P. 2005. *Teaching Geography*. New York: The Guilford Press.

Gersmehl, P. J., and C. A. Gersmehl. 2007. Spatial thinking by young children: Neurologic evidence for early development and "educability". *Journal of Geography* 106 (5): 181–191.

———. 2006. Wanted: A concise list of neurologically defensible and assessable spatial-thinking skills. *Research in Geographic Education* 8 (1) : 5–38.

Gilmartin, P. P., and J. C. Patton. 1984. Comparing the sexes on spatial abilities: Map-use skills, *Annals of the Association of American Geographers* 74 (4): 605–619.

Golledge, R. G. 2002. The nature of geographic knowledge. *Annals of the Association of American Geographers* 92 (1): 1–14.

———. 2001. Geography and everyday life (again!). *Directions Magazine* (March). http://www. directionsmag.com/columns.php?column_id=40 (accessed August 2010).

———. 2000. Geography and everyday life. *Directions Magazine* (December). http://www. directionsmag.com/columns.php?column_id=37 (accessed August 2010).

———. 1995. Primitives of spatial knowledge. In *Cognitive Aspects of Human-Computer Interaction for Geographic Information Systems*, ed. T. L. Nyerges, D. M. Mark, R. Laurini, and M. J. Egenhofer, pp. 29–44. Boston, Massachusetts: Kluwer Academic Publishers.

———. 1992. Do people understand spatial concepts: The case of first-order primitives. In *Theories and Methods of Spatio-temporal Reasoning in Geographic Space*, ed. A. U. Frank, I. Campari, and U. Formentini, pp. 1–21. New York: Springer-Verlag.

Golledge, R. G., M. Marsh., and S. Battersby. 2008. Matching geospatial concepts with geographic educational needs. *Geographical Research* 46 (1): 85–98.

Golledge, R. G., and R. J. Stimson. 1997. *Spatial Behavior: A Geographic Perspective*. New York: The Guilford Press.

Hall-Wallace, M. K., and C. M. McAuliffe. 2002. Design, implementation, and evaluation of GIS-based learning materials in an introductory geoscience. *Journal of Geoscience Education* 50 (1): 5–14.

Harris, L. J. 1981. Sex variations in spatial skills. In *Spatial Representation and Behavior Across the Lifespan*, ed. L. S. Liben, A. H. Patterson, and N. Newcombe, pp. 31–54. New York: Academic Press.

Hegarty, M., A. E. Richardson, D. R. Montello, K. Lovelace, and I. Subbiah. 2002. Development of a self-report measure of environmental spatial ability. *Intelligence* 30 (5): 425–447.

Janelle, D. G., and M. F. Goodchild. 2009. Location across disciplines: Reflection on the CSISS experience. In *Geospatial Technology and the Role of Location in Science*, ed. H. J. Scholten, N. van Manen, and R. v.d. Velde, pp. 15–29. Dordrecht, Netherlands: Springer.

Just, M. A., and P. A. Carpenter. 1985. Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review* 92 (2): 137–172.

Kerski, J. J. 2008. The world at the student's fingertips. In *Digital Geography: Geo-spatial Technologies in the Social Studies Classroom*, ed. A. J. Milson and M. Alibrandi, pp. 119–134. Charlotte, North Carolina: Information Age.

———. 2000. The implementation and effectiveness of geographic information systems technology and methods in secondary education. Ph.D. diss., University of Colorado, Boulder.

Kyllonen, P. C., D. F. Lohman, and D. J. Woltz. 1984. Componential modeling of alternative strategies for performing spatial tasks. *Journal of Educational Psychology* 76 (6): 1325–1345.

Lee, J., and R. S. Bednarz. 2009. The effect of GIS learning on spatial thinking. *Journal of Geography in Higher Education* 33 (2): 183–198.

Lloyd, R., and R. L. Bunch. 2003. Technology and map-learning: Users, methods, and symbols. *Annals of the Association of American Geographers* 93 (4): 828–850.

Lohman, D. F., and P. C. Kyllonen. 1983. Individual differences in solution strategy on spatial tasks. In *Individual differences in cognition*, ed. D. F. Dillon and R. R. Schmeck, pp. 105–135. New York: Academic Press.

Marsh, M., R. Golledge, and S. E. Battersby. 2007. Geospatial concept understanding and recognition in G6–college students: A preliminary argument for minimal GIS. *Annals of the Association of American Geographers* 97 (4): 696–712.

Milson, A. J., and B. D. Earle. 2007. Internet-based GIS in an inductive learning environment: A case study of ninth-grade geography students. *Journal of Geography* 106 (6): 227–237.

Montello, D. R., L. L. Lovelace, R. G. Golledge, and C. M. Self. 1999. Sex-related differences and similarities in geographic and environmental spatial abilities. *Annals of the Association of American Geographers* 89 (3): 515–534.

Patterson, M. W., K. Reeve, and D. Page. 2003. Integrating geographic information systems into the secondary curricula. *Journal of Geography* 102 (6): 275–281.

Schultz, R. B., J. J. Kerski, and T. C. Patterson. 2008. The use of virtual globes as a spatial teaching tool with suggestions for metadata standards. *Journal of Geography* 107 (1): 27–34.

Wigglesworth, J. C. 2003. What is the best route? Route-finding strategies of middle school students using GIS. *Journal of Geography* 102 (6): 282–291.