Hydrogeology Journal (2024) 32:1979–1991
https://doi.org/10.1007/s10040-024-02843-6

PAPER

Spatial thinking skills used by hydrogeology practitioners and students while completing a hydrogeology task

Peggy M. McNeal[1] · Heather L. Petcovic[2] · Joel Moore[1] · Donald M. Reeves[3] · Oluwarotimi Popoola[4] · Matthew Gordon[1]

Received: 13 December 2023 / Accepted: 28 September 2024 / Published online: 25 October 2024

Abstract

A typical classroom exercise in hydrogeology is to develop a conceptual model of a contaminated site, identify groundwater flow direction(s), and predict the location and mass of a contaminant plume. This requires knowledge of key hydrogeological concepts and is highly visuospatial in nature. Among multiple discrete spatial thinking skills identified by cognitive science, the combination of visual penetrative ability and working in multiple frames of reference were identified to significantly predict performance on a hydrogeology task and showed that together with hydrogeology knowledge, these spatial thinking skills account for 49% of the variability on task performance. Seventy-two hydrogeology practitioners and students with varying levels of expertise were administered multiple spatial thinking tests and an assessment of hydrogeology knowledge before completing a hydrogeology task that was developed for the study. Using spatial thinking and knowledge test scores as predictor variables, a hierarchical regression analysis was conducted with performance on the hydrogeology task as the outcome variable. The resulting model predicts that at low levels of hydrogeology knowledge, the identified spatial thinking skills account for more than a 25% difference on the hydrogeology task. This study provides empirical evidence that visual penetrative ability and working in multiple frames of reference are important skills in hydrogeology; thus, instructors are encouraged to recognize that underdeveloped spatial thinking skills could present hurdles for students and that targeted spatial thinking training may yield positive results for both weak and strong spatial thinkers.

Keywords General hydrogeology · Site characterization · Spatial thinking · Cognition · Education

Introduction

Geoscience education research (GER) is a robust area of inquiry that focuses on testing theory and producing generalizable findings that center on teaching, learning, and ways of thinking in the geosciences (NRC 2012). Hydrogeology, as a geoscience subdiscipline, can benefit from GER studies by using theory and findings to inform practice. Examples from two different geoscience subdisciplines are illustrative. In the subdiscipline of structural geology, over a decade of research suggests that the ability to relate, manipulate, and transform spatial information across multiple scales is essential for understanding the three-dimensional (3-D) nature of geologic structures both on maps and in the field. Similarly, in the subdiscipline of mineralogy, spatial skills related to visualizing and mentally rotating objects are essential to how students learn the internal crystal structures of minerals. Furthermore, students enter geoscience courses with a wide range of skills in spatial thinking, but using targeted spatial thinking curricula boosts students' ability to solve geological problems in both these domains (Ormand et al. 2017).

Spatial thinking includes thinking about shapes, locations, and paths, along with relationships among and between these things, and the frames of reference in which they exist. Spatial thinking also involves mentally transforming information

✉ Peggy M. McNeal
pmcneal@towson.edu

[1] Department of Physics, Astronomy, and Geosciences, Towson University, 8000 York Road, Towson, MD 21252, USA

[2] Department of Geological and Environmental Sciences and The Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI, USA

[3] Department of Geological and Environmental Sciences, Western Michigan University, Kalamazoo, MI, USA

[4] The Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI, USA

while manipulating, constructing, and navigating the physical world, and relating the physical world to internalized mental models (Newcombe and Shipley 2014). In everyday examples, individuals use spatial thinking to efficiently pack a bag, put together IKEA furniture, build a LEGO model, or use a map to navigate.

Hydrogeologists likely use some of the spatial thinking skills previously identified in other geoscience subdisciplines, but hydrogeology is different in a few ways, suggesting that it may require distinct spatial thinking skills. Much as a structural geologist or stratigrapher might correlate rock and sediment between sparse outcrops (or boreholes), hydrogeologists work with sparse and spatially discontinuous subsurface data that require interpolation. Working largely with materials in the subsurface severely limits the features that a hydrogeologist can directly observe. Additionally, hydrogeologists need to contend with water (and/or other fluids), media, and possibly contaminants flowing dynamically through largely static regolith and rock at varying spatial and temporal scales. Although in practice, mathematical relationships and computer models describe groundwater flow and contaminant transport, presumably, and especially for students, a conceptual understanding of how a contaminant plume travels through the subsurface requires visualization and spatial thinking skills to create and manipulate a mental model. Because of these distinctions from other geoscience disciplines, the goal of this study is to characterize the spatial thinking skills used specifically in hydrogeology.

Although these examples could group together as generalized spatial thinking or visualization ability, human intelligence researchers recognize many distinct and separate visuospatial factors. Current human intelligence theory describes the structure of human cognitive abilities, including visuospatial skills, and psychology researchers have developed psychometric tests for measuring discrete factors. Psychometric studies of human spatial abilities (Carroll 1993; McGee 1979; McGrew 2009) provide a theoretical basis for identification of discrete spatial thinking skills (or factors) based on factor-analytic results (French et al. 1963). Cattell-Horn-Carroll (CHC) theory describes 16 broad factor groups that contribute to general human intelligence (Schneider and McGrew 2012), one of which is visual processing, although “spatial ability” is its more commonly used name in the psychometric literature (Buckley et al. 2018). Within the visual processing/spatial ability factor group, CHC theory identifies and defines 11 discrete factors broadly grouped into three categories. The first group, “visualization”, lists factors that describe how spatial information is manipulated through actions such as mental rotation or mental cutting (Uttal et al. 2013). The second group, “perceptual”, lists factors such as perceptual alternations, which describe how individuals perceive and encode spatial information. The third group, “memory”, lists factors such as visual memory, which describe how individuals store or hold spatial information in working or short-term memory.

Working specifically to apply CHC theory to science, technology, engineering, and math (STEM) education, Buckley et al. (2018) considered the number of visual processing/spatial ability factors within CHC theory and proposed that the current theory underrepresented some spatial factors, particularly those related to dynamic spatial ability. Within the three categories described previously, Buckley et al. (2018) noted that three purely dynamic factors—directional judgment, speed judgment, and movement detection—inform the mental models that learners construct when they integrate ideas about dynamic movement. They extended the original 11 spatial factors to a total of 25 spatial factors considered relevant to STEM learning, especially by categorically dissociating static and dynamic spatial factors. From an intelligence research perspective, this extension of spatial factors may advance work to develop a cognitive map of the human mind.

Buckley et al. (2018) recognized that the broader impacts of this work led to development of interventions to improve spatial ability, which have already had significant positive effects on STEM students who lack strong spatial thinking skills. Prior research demonstrates that spatial thinking is foundational to learning in STEM (Kell et al. 2013; Shea et al. 2001; Wai et al. 2009). A longitudinal study conducted by Shea et al. (2001) found that students who ultimately ended up in STEM careers had higher levels of spatial ability at age 13. While all individuals have some amount of inherent spatial ability, studies also show that spatial thinking ability can develop and improve with training and practice. Uttal et al. (2013) conducted a meta-analysis of 217 studies that investigated the trainability of spatial skills and concluded that spatial training improves spatial thinking, especially when compared to control groups. These findings suggest that characterizing the spatial thinking used uniquely with hydrogeology can support learning in hydrogeology courses through targeted training. The potential to build curricula that will more effectively teach students hydrogeology motivated this study, but identifying which spatial thinking skills are essential in hydrogeology was a necessary first step. The overall goal of this study was to identify spatial thinking skills used by hydrogeologists with two research questions: (1) What spatial thinking skills are essential to successfully completing a hydrogeology task? and (2) What effect do these spatial thinking skills have on performance on a hydrogeology task?

Materials and methods

This quantitative study used a cross-sectional design to identify spatial thinking skills that correlate with hydrogeology performance and to model the effect of predictor variables on hydrogeology performance. A deliberately selected sample of practicing and student hydrogeologists allowed characterization of individual differences while identifying essential spatial thinking skills. The Institutional Review Boards at the researchers' universities approved the study and all individual participants provided informed consent.

Development and selection of measures

Prior to data collection, the researchers selected spatial thinking tests and developed measures for additional predictor (hydrogeology knowledge) and outcome (hydrogeology performance) variables (Fig. 1).

Selection of spatial thinking tests

Starting with the 25 spatial factors identified by Buckley et al. (2018), the researchers sought to ultimately select three to four measurable spatial thinking skills that are relevant to practicing hydrogeology that participants could complete without mental fatigue. Experienced hydrogeologists on the research team and advisory board selected nine potentially relevant spatial thinking skills. The researchers invited members of the National Ground Water Association (NGWA) to complete a survey hosted on Qualtrics (a survey management software) about perceived use of each of the nine spatial thinking skills in their hydrogeology work and used responses from nine individuals to identify three spatial thinking skills to test. Survey respondents ranked spatial scanning and visual penetrative ability highly, and the researchers added working in multiple frames of reference based on this quote from one of the responses, "It is very important to not only discuss spatial relationships but also spatial relationships as they pertain to 2-dimension and 3- dimension planes…it is imperative that 2-D and 3-D be thoroughly understood". The researchers included a fourth skill (mental rotation) on the recommendation of a cognitive scientist because its wide use in spatial thinking research would enable the researchers to make cross-study comparisons. Descriptions of each spatial thinking skill and the tests (Fig. 2) used to measure them follow in the next sections.

Spatial scanning French et al. (1963) define spatial scanning as speed in visually exploring a wide or complicated spatial field. Tests that measure this skill require scanning a field quickly for openings, following paths with the eye, and quickly rejecting false leads. The researchers used the Education Testing Service's (ETS) Map Planning Test (MPT) to measure this spatial thinking skill (Fig. 2a). In the test, participants see city maps with streets blocked at various points by barriers represented by circles. The participant must plan routes between given points that do not cross roadblocks (circles). The shortest available route will only pass one building (numbered squares). The participant looks for the shortest route as quickly as possible and responds with the number of the building passed. The test includes two

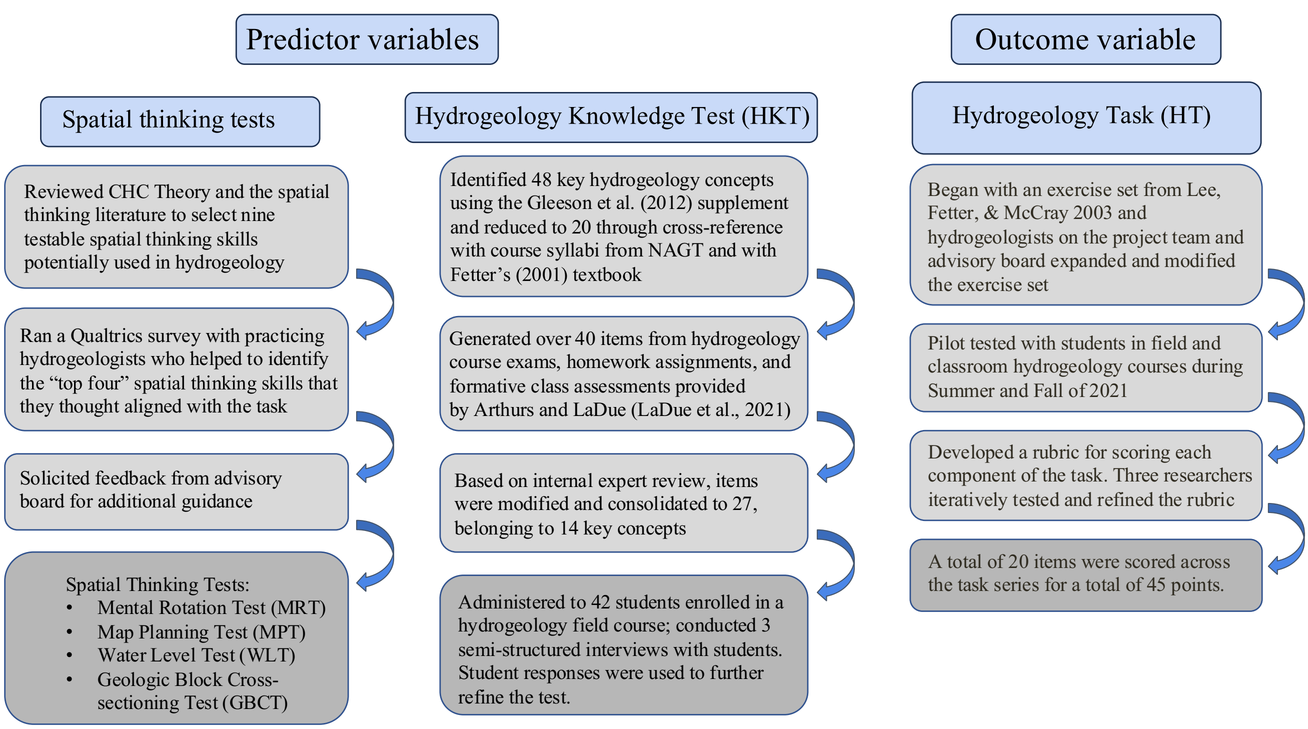

Fig. 1 Process for developing and selecting measures

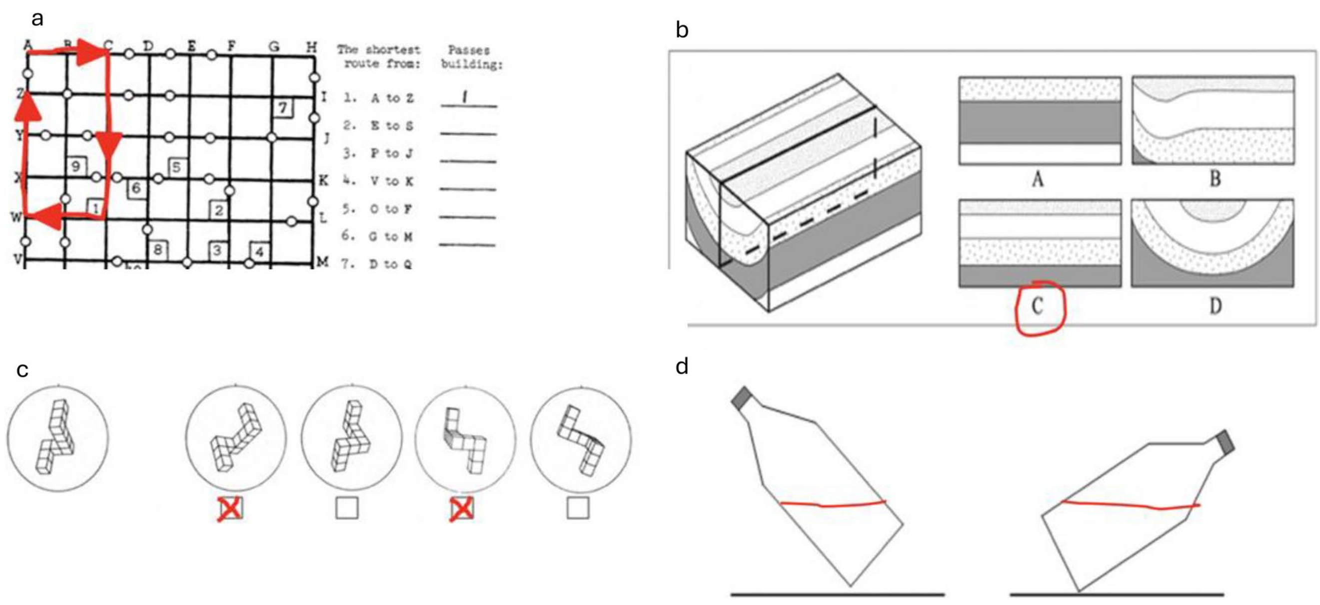

Fig. 2 Example items from spatial thinking tests used in the study: **a** Map Planning Test (ETS 1964) to test spatial scanning; **b** Geological Block Cross-sectioning Test (Ormand et al. 2014) to test visual penetrative ability; **c** Vandenberg and Kuse Test of Mental Rotation (Vandenberg and Kuse 1978) to test mental rotation; and **d** Piaget's Water Level Test (Piaget and Inhelder 1956) to test frames of reference

maps each with ten routes, and participants have three minutes to complete each map. The researchers scored the test by the number of completed and correctly identified paths.

Visual penetrative ability The ability to visualize spatial relations inside an object from outside or surface clues is penetrative thinking, also known as visual penetrative ability (Newcombe and Shipley 2014). Kali and Orion (1996) identified visual penetrative ability by using a geologic spatial ability test that included four problems requiring individuals to draw cross-sections of structures presented as block diagrams. Alles and Riggs (2011) describe visual penetrative ability as visualizing "the structure of a geologic object from surface clues and projecting elements of that structure into the interior of a block or underground". CHC theory does not specifically describe visual penetrative ability, although it is referenced outside of geoscience contexts (e.g., Cohen and Hegarty 2012). Kali and Orion (1996) note that items included in their geologic spatial ability test belong to the spatial visualization category described by McGee (1979).

The researchers used the Geologic Block Cross-sectioning Test (GBCT) developed by Ormand et al. (2014), which measures skill in visualizing the internal 3D geometries of geologic structures (Fig. 2b). In the test, participants see a geologic block diagram sectioned with a planar surface, and they choose from four multiple-choice responses of the interior surface that would match the given block diagram. The test included 16 items, and the researchers gave participants 8 min to complete the test. The researchers scored the test by the number of correctly identified cross-sections.

Mental rotation Skill with mentally rotating objects correlates strongly with success in STEM fields (Cheng and Mix 2014; Sorby 2007). In the Buckley et al. (2018) framework, both a speeded rotation and spatial orientation factor describe mental rotation. Newcombe and Shipley (2014) recognized the dynamic nature of mental rotation using a framework that characterized spatial skills as intrinsic or extrinsic and static or dynamic, due to rotation of the stimulus.

The researchers selected the Vandenberg and Kuse Mental Rotation Test (MRT; Vandenberg and Kuse 1978) to measure this skill because it assesses the ability to transform visual images, a key component of visual processing. The Vandenberg and Kuse MRT is a timed test that consists of symmetric and asymmetric figures of 3-D objects (Fig. 2c), drawn in a 2-D isometric format and rotated around a horizontal axis. Each item presents an initial figure followed by four others. From the latter four figures, subjects select two figures that match the same configuration as the initial figure, only rotated (Vandenberg and Kuse 1978). The test consists of 24 items administered in two sets of 12 with a 3-min time limit for each set. On average, males outperform females on timed tests of mental rotation; however, Newcombe and Stieff (2012) note weak support for biological causation and conclusive evidence that individuals can improve these skills with training. To eliminate the possibility of scoring guesses as correct, the researchers scored items as correct only if a participant selected both correct configurations (Vandenberg and Kuse 1978).

Frames of reference Buckley et al. (2018) include skill with working in multiple frames of reference within a group of perceptual illusion spatial factors that "describe the ability to not be fooled by illusions when encoding visual information". Buckley et al. (2018) restate Coren et al.'s (1976) description of frames of reference illusion as involving "the comparison of an element to its global frame of reference". The researchers measured skill with frames of reference using Piaget's Water Level Test (WLT; Piaget and Inhelder 1956). Originally designed for children, Piaget and Inhelder considered accuracy on this test as indicative of a mature Euclidean reference system (Piaget and Inhelder 1956). However, later studies demonstrated that difficulty with the test persisted into adulthood (Rebelsky 1964).

The test (Fig. 2d) simply asks participants to draw a line inside a tipped bottle to show where the water would be if the bottle were half full. The test has no time limit and includes six bottles at different angles of tilt. Vasta and Liben (1996) refer to the difficulty that so many adults have with this test as an "enigma", and the literature since that time offers little resolution. Although it is unlikely that some adults have failed to establish a Euclidean reference system, it is possible that some adults fail to apply it (Vasta and Liben 1996). The researchers scored each item as two points if inside 5° of parallel to the surface, one point if inside 10°, and zero points if outside of 10°.

Hydrogeology knowledge test

For the purposes of the study, the researchers developed the Hydrogeology Knowledge Test (HKT) to measure knowledge of the domain. The HKT is currently undergoing full validation. Domain knowledge results from an accumulation of school and lifetime experiences. Researchers can measure domain knowledge using concept inventories and tests of the factual knowledge required for understanding a field. Not only is possessing sufficient domain knowledge fundamental to expertise (Wiley 1998), but the way that this knowledge is organized enhances its accessibility, functionality, and efficiency (Bédard and Chi 1992). Prior research demonstrates that domain knowledge and expertise are the strongest predictors of performance on problem-solving tasks in that domain (Hambrick et al. 2012; Hegarty et al. 2010; McNeal et al. 2019). Thus, it was important to measure hydrogeology knowledge in the investigation to determine its effect on performance and how it might interplay with the measured spatial thinking skills.

Many disciplines have developed generalized, validated concept inventories for assessing knowledge within specific subject domains; the most well-known is perhaps the Force Concept Inventory for physics (Hestenes et al. 1992). In the geosciences, researchers have developed validated concept inventories for entry-level geoscience courses (the Geoscience Concept Inventory; Libarkin and Anderson 2005), mineralogy (Scribner and Harris 2020), oceanography (Arthurs et al. 2015), and meteorology (Davenport and French 2020). In hydrogeology, however, a concept inventory does not yet exist.

The HKT went through four iterations, versions 1, 2, 3, and 4. First, the researchers used a review by Gleeson et al. (2012), to identify the top 15 topics considered crucial by their survey of hydrogeology instructors. By cross-referencing these topics with hydrogeology course syllabi compiled by the National Association of Geoscience Teachers (United States) and the chapters outlined in a commonly used hydrogeology textbook (Fetter 2013), the researchers identified 20 key concepts. Using this concept list, the researchers adopted (from LaDue et al. 2021), revised, or generated over 40 possible items. Review by four practicing hydrogeologists consolidated the knowledge test to include 27 items, 13 short-answer, and 14 multiple-choice, belonging to 14 key concepts. The researchers administered this preliminary version of the Hydrogeology Knowledge Test (version 1) to 49 students enrolled in a hydrogeology field camp and interviewed three students. In the semistructured interviews, the researchers sought content validity by ensuring that the students interpreted each question-and-answer option as intended.

To make the second version of the test faster to both complete and score, the researchers converted the short-answer items into multiple choice. The researchers categorized written student responses and used salient scientifically inaccurate ideas couched in student language to generate distractors (plausible incorrect responses) for multiple-choice items (after Arthurs et al. 2015) to produce the third version. Several additional rounds of revision consolidated similar items, simplified wording, and eliminated some items based on feedback from an advisory board. The final version (HKT; version 4) used for the study had a total of 21 questions (2 short-answer and 19 multiple-choice) that covered 11 key hydrogeology concepts (Table 1, see electronic supplementary material (ESM) for HKT version 4). The researchers gave participants unlimited time to complete the HKT and scored the HKT by the number of correctly answered items.

Hydrogeology task

To evaluate hydrogeology performance and serve as an outcome variable in a hierarchical regression analysis, the researchers developed a Hydrogeology Task (HT). The intent was to leverage knowledge and data analysis skills that most students possess by the time they complete 66–75% of an upper-level undergraduate hydrogeology course, while maintaining an ability to discriminate between the performance of practicing hydrogeologists

Table 1 Key concepts covered on the Hydrogeology Knowledge Test (with item numbers)

Key concepts	Item number
Aquifer and confining units	2, 4, 5, 12
Contaminant transport in groundwater	10, 11, 15
Darcy's Law and its applicability	8
Gradient and head	9, 18, 21
Hydraulic conductivity/intrinsic permeability	3, 17
Hydraulic testing	19
Primary and secondary porosity	16
Recharge and discharge areas	6, 14
Water budget	7
Water quality	1, 20
Water table and mapping	13

and students. Because the researchers were interested in identifying requisite spatial thinking skills, the task needed to use spatial thinking and visualization.

The task is a typical contaminated site investigation problem modeled on an exercise set from Lee et al. (2003), but expanded and modified to include spatial components. In its final form, the HT included a 3-D synthetic aquifer (high permeability)–aquitard (low permeability) system in which participants were asked to: (1) determine the geometry of the aquifer and underlying aquitard in a cross section using borehole data; (2) contour the potentiometric surface from well data and interpret the groundwater flow field; (3) complete a 3-point problem using potentiometric surface data from three wells and determine groundwater flow; and (4) integrate these results along with contaminant concentrations from monitoring wells to delineate the extent of subsurface contamination and predict plume migration. The researchers provide the entire HT in the ESM and include examples of completed work from students and practicing hydrogeologists.

To ensure clarity of instructions, and fidelity of the data, the researchers pilot-tested the HT with students in field and classroom hydrogeology courses during the summer and fall of 2021 and made minor revisions as a result. The researchers developed equally weighted scoring rubrics for each task component and awarded points for accuracy, consistency of interpretations with data, and geologic plausibility of responses over heuristics, neatness, and rote knowledge. Three research team members iteratively scored the completed HT, compared scores, discussed differences, and refined the rubric until reaching a high level of scoring agreement and consistency. The final rubric scores 20 items across all elements of the task and totals 45 points. The first and second author used this rubric to score all participant tasks independently, then met to compare and discuss scores and resolve differences until they reached consensus. The researchers provide the entire HT scoring rubric in the ESM.

Domain experience questionnaire

To gather demographic data and evaluate levels of expertise, the researchers modified valid and reliable surveys used in previous work (Baker et al. 2012; McNeal et al. 2019; Petcovic et al. 2016) to make them applicable to hydrogeologists. The researchers provide the Domain Experience Questionnaire (DEQ) in the ESM.

Data collection

Data collection began in September 2021 and continued through August 2022. The researchers initially recruited participants from attendees at the 2021 Geological Society of America (GSA) Annual Meeting by posting announcements in online professional society forums (e.g., GSA Hydrogeology Division), distributing recruiting flyers at the meeting, and making verbal announcements at topical sessions. The researchers employed snowball recruiting by encouraging individuals who participated to tell colleagues about the study. Several hydrogeologists participated in person at the meeting, with others participating after the meeting using a virtual (Zoom) format with mailed materials. This phase of recruitment netted mainly practicing hydrogeologists. To include students in the sample, the researchers contacted colleagues at three universities with robust hydrogeology programs. A contact person distributed email announcements to students enrolled in hydrogeology courses and to undergraduate majors and graduate students. Visits to these institutions took place for in-person data collection in the spring and summer of 2022. Additionally, two of the universities had state geological surveys on campus, from which the researchers recruited additional practicing hydrogeologists. To qualify for the study, participants needed to be age 18 or older and have completed or be currently enrolled in a hydrogeology course. Participants (Table 2) ranged from undergraduate students to hydrogeologists with more than 10 years of work experience.

The researchers conducted most data collection sessions with one researcher and one participant; however, in a few instances, one researcher worked with a small group of participants. Each session began with an opportunity for the participant to ask questions about the study and sign a consent form. Next, the researchers gave each participant a binder, containing paper copies of all the measures described previously, a ruler, pencil, pen, and eraser. The researcher gave instructions for each test, managed the timer, and answered questions. Each participant worked through the tests in the same order: Geologic Block Cross-sectioning Test (GBCT), Water Level Test (WLT), Mental Rotation

Table 2 Demographics of participants in the sample

Demographic		% of total sample
Gender	Man	62
	Woman	38
	Nonbinary	0
Race/ethnicity	White	81
	Black	1
	Asian	6
	Hispanic	4
	Middle Eastern	2
	Native American	1
Courses completed	Physical Geology	90
	Sedimentology and/or stratigraphy	78
	(Introduction to) Hydrogeology	92
	Advanced hydrogeology	54
	Groundwater modeling	53
	Aqueous or low-temperature geochemistry	50
	Hydrogeology field course	47
	Surface-water hydrology	47
Education level	Pursuing MA/MS	17
	Hold MA/MS	39
	Pursuing PhD	17
	Hold PhD	25
Licensed/certified		19
HAZWOPER Trained		58
Have you worked professionally as a hydrogeologist?	None	57
	<5 years	16
	5–10 years	6
	>10 years	21

Test (MRT), Map Planning Test (MPT), Hydrogeology Knowledge Test (HKT), Domain Experience Questionnaire (DEQ), and Hydrogeology Task (HT). A session typically took between 1.5 and 2 h, and the researchers compensated participants with $50 electronic gift cards. In total, 72 individuals participated in the study.

Data analysis and results

To analyze the data, the researchers used Statistical Package for Social Sciences (SPSS) Version 28. The researchers ran descriptive statistics (Table 3) on all variables and screened the data for outliers and influential cases, but finding none, analyzed the full data set. The statistical tests used assume normally distributed data and some of the data were not. (Figs. S1–S14 of the ESM provide histograms and probability-probability (P-P) plots that display data distribution for all variables.) In particular, the distribution of Water Level Test scores was bimodal between perfect (or near-perfect) and low. To avoid violating the assumptions of the tests, the researchers used bootstrapping to treat the data as a population from which smaller samples are taken and returned before a subsequent one is drawn (Field 2013).

As an extra measure of validity, the results of an independent samples *t*-test confirmed that the data aligned with expected patterns documented in the literature for mental rotation. This test confirms a sex difference (Newcombe and Stieff 2012) in which males had higher mental rotation scores (mean (M) = 10.78, standard deviation (SD) = 4.39) than did females (M = 8.48, SD = 4.24), t statistic $t(69) = 2.12$, p-value (p) = 0.037. The remaining measures did not exhibit a significant difference in scores by sex.

Table 4 presents results of the Pearson correlation (r). The p-values (provided in the following) indicate the likelihood that the data would occur by random chance with a significance level set at ≤0.05. Measures of hydrogeology knowledge and expertise strongly correlate ($r = 0.565$, $p < 0.001$), which confirms reasonable expectations and reinforces the validity of each measure.

Table 3 Descriptive statistics for each variable. *N* sample size, *M* mean, *SD* standard deviation

Test (what it measures)	*N*	Minimum	Maximum	*M*	SD	Skewness[a]	Kurtosis[b]
GBCT (visual penetrative ability)	72	2.00	16.00	10.40	3.86	-0.47	-0.65
WLT (frames of reference)	72	0.00	12.00	10.08	3.24	-1.99	3.07
MRT (mental rotation)	71[c]	0.00	19.00	9.87	4.47	0.05	-0.60
MPT (spatial scanning)	72	10.00	40.00	25.69	7.28	-0.12	-0.55
HKT (hydrogeology knowledge)	72	9.16	23.00	17.28	3.23	-0.275	-0.34
DEQ (expertise)	72	2.50	14.00	6.62	3.91	0.46	-0.91
HT (hydrogeology performance)	72	2.00	42.00	25.51	9.62	-0.61	-0.56

[a]Measures asymmetry of the data distribution

[b]Indicates peakedness of the data distribution

[c]One participant did not complete the Mental Rotation Test

The total HT score correlates strongly with both expertise ($r = 0.389$, $p < 0.001$), and hydrogeology knowledge ($r = 0.585$, $p < 0.001$), supporting that it is a valid measure of skill. All the spatial thinking tests correlate with each other, which has no theoretical basis, but other studies report the same (see, for example, Hambrick et al. 2012). The spatial thinking skills that positively correlate with performance on the HT are: (1) visual penetrative ability ($r = 0.562$, $p < 0.001$); (2) working with multiple frames of reference ($r = 0.483$, $p < 0.001$); and (3) spatial scanning ($r = 0.252$, $p < 0.005$).

In performing the hierarchical regression, the researchers entered hydrogeology knowledge in step 1 as a known predictor based on past work (Hambrick et al. 2012; McNeal et al. 2019) and the variable with the strongest correlation with the HT. At this point, hydrogeology knowledge predicted 34% of the variance in performance on the HT ($\Delta R^2 = 0.335$, $\Delta F(1, 67) = 35.270$, $p < 0.001$; see Table 5 for further explanation on each parameter). In step 2, the researchers experimented with adding the remaining predictors (Field 2013) and found that once they added visual penetrative ability and working with multiple frames of reference to the model, there was no additional change in the variance (no increase in R^2) upon adding spatial scanning, so they did not retain this variable in the model. Because visual penetrative ability and working with multiple frames of reference were highly correlated ($r = 0.442$, $p < 0.001$), the researchers collapsed them into a single variable by adding the two scores together after equally weighting them. The researchers entered this spatial score variable in step 2. The researchers present the results of the final model in Table 5 and show that after hydrogeology knowledge, spatial thinking skills (specifically visual penetrative ability and working with multiple frames of reference) account for an additional 15% of the variance in HT scores.

Having fitted the model, the researchers needed to check for bias in the model to ensure that it is accurate for the sample and generalizable to the population (Field 2013). The researchers provide plots of residuals and bivariate relationships that accompany this discussion in the ESM. They created a plot of standardized predicted values against standardized residuals (Fig. S15 of the ESM). The points are random and, for the most part, evenly dispersed. An ideal pattern of completely even dispersion would indicate that the data meet the assumptions of linearity and homogeneity of variance (Field 2013). Additionally, the researchers

Table 4 Table of correlations (Pearson's *r*)

Test (what it measures)	GBCT	WLT	MRT	MPT	HKT	DEQ	HT
GBCT[a] (visual penetrative ability)	1.00	0.42**	0.37**	0.41**	0.36**	4.07	0.56**
WLT[a] (frames of reference)		1.00	0.44**	0.31*	0.33**	0.13	0.44**
MRT[b] (mental rotation)			1.00	0.39**	0.04	–0.06	0.15
MPT[a] (spatial scanning)				1.00	–0.04	–0.16	0.26*
HKT[a] (hydrogeology knowledge)					1.00	0.57**	0.59**
DEQ[a] (expertise)						1.00	0.42**
HT[a] (hydrogeology performance)							1.00

*Correlation is significant at the 0.05 level (2 tailed). ** Correlation is significant at the 0.01 level (2 tailed)

[a]$N = 72$

[b]$N = 71$

Table 5 Final model

Variable	ΔR^{2a}	β^b	ΔF^c	df[d]
Step 1 Hydrogeology knowledge (HKT)	0.34	0.42	37.39**	1.70
Step 2 Spatial score (visual penetrative ability and frames of reference)	0.15	0.43	34.71**	1.69
Total adjusted R^2	0.49			

$N = 72$: ** $p < 0.001$

[a]Variance in the outcome: the proportion of the variance in the outcome variable that is shared with the predictor variable

[b]Beta: standardized regression coefficient, which is change in the outcome associated with a unit change in the predictor

[c]F-ratio: ratio of variability explained by model to average variability unexplained by model, which tests overall fit

[d]Degrees of freedom: the number of values that are free to vary, which has bearing on significance tests

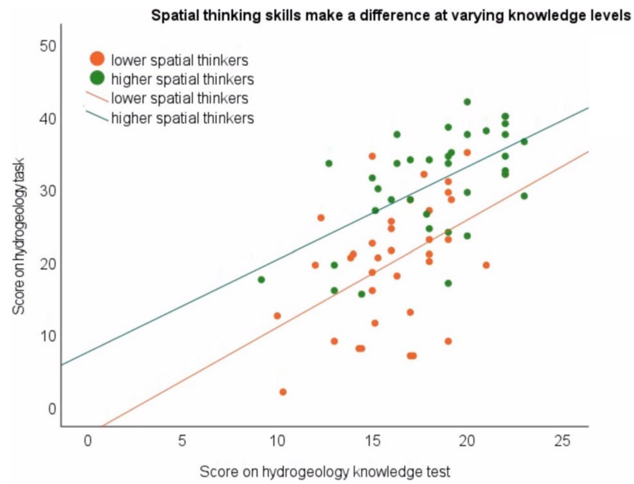

Fig. 3 Hydrogeology Task vs. Hydrogeology Knowledge Test by higher and lower spatial thinkers

created scatterplots of the residuals (difference between the model prediction and data points) of the HT and each of the predictor variables (Figs. S16 and S17 of the ESM). There are no extreme outliers, but some "slightly wandering points" could show influence on a predictor's regression coefficient (Field 2013). To test the normality of the residuals, the researchers created a histogram (Fig. S18 of the ESM) and a normal probability-probability (P-P) plot (Fig. S19 of the ESM). The residuals do not demonstrate a normal distribution, so as expected, the P-P plot exhibits some deviations from the diagonal line. This could indicate a lack of homogeneity of variance. Note, however, these plots must be created without the use of robust methods, and because the researchers recognized limitations in the data set, they used bootstrapping to build the model, which largely overcomes these problems (Field 2013). The sample size provides sufficient power to base claims on the model. The researchers conducted a post-hoc power analysis (using G*Power; Faul et al. 2009) for a linear multiple regression for a sample of 72 and found it was sufficient to detect a "medium" effect (Cohen's $f^2 = 0.15$; a measure of effect size used for multiple regression). This result aligns with comparable studies found in the literature (e.g., Hambrick et al. 2012, found sufficient power for a medium effect with 67 participants) and the sample size is typical in social science research involving data collection from human subjects (e.g., Linnenbrink-Garcia et al. 2012, $N = 94$; Resnick and Shipley 2013, $N = 35$). Specifically, this means that a sample size of 72 is sufficiently large and provides enough statistical power to support the claim that the spatial thinking skills identified (visual penetrative ability and working in multiple frames of reference) are significant predictors of performance on the HT. (Stated another way, with significance level set at ≤ 0.05, it can be expected that 80% of random samples from the same population will exhibit this relationship). The researchers are overall satisfied with the model's ability to describe the sample and to generalize to the larger population that the sample represents.

To investigate the interplay between spatial thinking and hydrogeology knowledge, the researchers performed a median split on the spatial thinking score (combined visual penetrative ability score and working with multiple frames of reference score) and found a significant difference in the performance of lower and higher spatial thinkers on the HT (independent sample *t*-test, $t(70) = 5.67$ $p < 0.001$). Higher spatial thinkers had higher HT scores ($M = 30.88$, SD $= 7.31$) than lower spatial thinkers ($M = 20.15$, SD $= 8.68$). A scatterplot of performance on the HKT regressed onto performance on the HT by lower and higher spatial thinkers (Fig. 3) illustrates the difference.

Discussion

The two research questions frame the discussion below.

What spatial thinking skills are essential to successfully completing a hydrogeology task in hydrogeology?

The researchers identified visual penetrative ability and working with multiple frames of reference as spatial thinking skills important for completing the hydrogeology task. The spatial score representing these combined factors was a statistically significant predictor of performance on the HT and accounted for 15% of the variance. Spatial scanning correlated with performance on the HT; however, the fact that it did not increase R^2 in the hierarchical regression analysis indicates that visual penetrative ability and working with multiple frames of reference already captured the advantage this skill affords in the model. Overall, this work quantifies an ability to move among a potentiometric surface, a cross-section, and a contamination gradient and be able to visualize groundwater flow and contaminant dispersion using information from these different reference frames simultaneously.

The Geologic Block-Cross-sectioning Test measures visual penetrative ability specifically in a geology context. The close association between the Geologic Block Cross-sectioning Test and the cross-section part of the HT could potentially bias the overall task score. However, even with the score for the cross-section portion of the task removed, spatial thinking and task scores still correlated ($r = 0.514$, $p < 0.001$), which indicates that skill with visualizing the structure of the subsurface from surface clues plays an important role across the task as a whole, not just in constructing a cross-section.

Alles and Riggs (2011) conducted problem-solving interviews with students representing low, medium, and high visual penetrative ability, which gives some insight into what these skills impart. These researchers worked with students as they attempted to draw two unknown surfaces of incomplete block diagrams after they provided them with features of one surface. Alles and Riggs (2011) probed the students' problem-solving difficulties and thought processes and found that students with high visual penetrative ability followed an ideal process in this order:

1. "Saw" or visualized a three-dimensional perspective that was represented on a 2-D medium.
2. Evoked a mental image of the external as well as internal structure using a 3-D perspective. This stage is rapid, dynamic, and involves the manipulation of tools or use of gestures.
3. Recognized how the mental model would project to the surface, which was necessary to draw the unknown surfaces. This stage was typically quick with little hesitation and relied on a strong mental model.
4. Reformed the mental image with the newly constructed face and combined all the previous information to generate a completely new cross section. This stage took the most time and was the most difficult.

Alles and Riggs (2011) emphasize that success in the activity is entirely dependent on a strong mental model and that the strength of the mental model "is realized when they move from an introspective mental image to an externalized manifestation by consolidating their thoughts on paper and verbally explaining their mental processes". Cohen and Hegarty (2007) analyzed the errors of 59 students on the Santa Barbara Solids Test (a nongeological test of visual penetrative ability) and concluded that sources of difficulty stemmed from failure to shift mental perspective from the current view and imagine the cutting plane from another orientation. Collectively these studies pointed to the importance of being able to visualize structures that are not apparent and have to be constructed from clues in other sources.

Working in multiple frames of reference adds a level of complexity. Of the Water Level Test, Liben et al. (2011) write that it "is thought to tap the respondent's ability to establish and use a systematic stable spatial reference system in the face of conflicting or distracting alternative referents". In other words, an individual needs to keep the seducing tilting of the bottle from causing them to draw a tilting water line as well. Vasta et al. (1994) found that increasing the angle of tilt produced larger errors, so the seductive effect is strong. While the researchers are continuing to investigate specifically how having skill with multiple frames of reference benefits individuals working on the HT, they can offer a few anecdotal observations from this study. First, in the three-point problem, the researchers asked participants to draw an arrow representing groundwater flow direction. When drawn correctly, there is 22° between the arrow and the right side of the planar surface, yet many participants straightened the arrow so that it was parallel to the right side. Additionally, and especially with students, the researchers noticed incidences where students conflated aspects of the surface with the subsurface, e.g., treating potentiometric surface elevation data as topographic surface elevation data, and assuming groundwater flow would mimic surface-water flow. It may be that weak spatial thinkers have difficulty distinguishing between these hydrologic settings and moving between them. The addition of the contaminant increases the complexity of the task, and drawn maps suggest that some of the participants lacked accurate conceptions of a contaminant plume as a 3-D entity. The researchers encountered open contours and other anomalies, suggesting that some participants failed to mentally visualize the contaminant plume, although they may have instead had a robust mental model but were unable to draw contours appropriately.

The HT required working with map (bird's eye) views and cross-sectional views to synthesize multiple types of data (elevation, potentiometric surface elevation, stratigraphic, contaminant concentration) and infer water flow direction and contaminant migration and dispersion. Individuals who were able to conjure mental models of the hydrogeologic environment using multiple perspectives and seamlessly move between perspectives demonstrated higher levels of success with the task.

What effect do these spatial thinking skills have on performance on a hydrogeology task?

Together with hydrogeology knowledge, the model accounts for 49% of the variance on the HT, which is commensurate with the results of similar studies in social science research (see, for example, Hambrick et al. 2012; McNeal et al. 2019; Pugh et al. 2010). The spatial score variable (representing visual penetrative ability and working with multiple frames of reference) accounts for 15% of the variance. The researchers illustrate the effect of the variability in Fig. 3, which shows that performance on the HT increased with higher levels of hydrogeology knowledge. However, when they divided the sample in half by spatial score and created groups of lower and higher spatial thinkers, the difference that these spatial thinking skills make becomes apparent. At low levels of hydrogeology knowledge, the model suggests that spatial thinking skills account for more than a 10-point (or >25%) difference on the 45-point task. The advantage decreases with higher levels of hydrogeology knowledge, which aligns with other studies. For example, Hambrick et al. (2012) found that novice geologists with lower levels of geology knowledge relied more on spatial thinking skills to complete a bedrock mapping task than expert geologists. Because experts with lower spatial thinking skills performed as well as those with high skills, Hambrick et al. (2012) reasoned that their higher levels of mapping expertise allowed them to circumvent limits they may have with spatial thinking skills. In contrast to the Hambrick et al. (2012) study, this study found that the performance gap diminishes but still persists at high levels of knowledge. Geoscience education researchers cite the Hambrick et al. (2012) study as justification for including spatial thinking training in introductory geoscience courses because, at lower levels of knowledge, dependence on spatial thinking skills may be necessary to overcome barriers in curricula that are problematic for some students. This study suggests the same implications for hydrogeology. Students in introductory hydrogeology courses without good visual penetrative ability or who lack skill with working in multiple frames of reference are at risk of poor performance and may not persist in the discipline. Targeted spatial training (or minimal recognition by the instructor) could help alleviate challenges, advance content understanding, increase retention, and ultimately provide the workforce with more diverse thinkers and problem solvers.

Limitations

Dozens of spatial thinking skills, and tests for measuring them, exist. It is quite possible that the researchers missed an important spatial thinking skill because they did not measure it. Reasonably, they could only administer a limited number of tests that would not mentally fatigue the participants. They relied on the expertise of an advisory board and multiple professional hydrogeologists to select the administered tests; nevertheless, the tests did not begin to cover the complete domain of spatial thinking. A validated concept inventory for hydrogeology did not exist at the time the researchers conducted the research. While it is possible that the HKT was not a completely reliable instrument, the researchers took great effort to develop it according to established protocol for development of concept inventories. Work toward complete validation is ongoing. Finally, the HT asks participants to solve problems at the macro-scale, whereas practitioners may interrogate many flow and transport problems (e.g. Gao et al. 2022) at the pore-scale, where complex pore structures and geometries are involved. The researchers do not know if the spatial thinking skills required at this scale are different from those observed. This opens additional questions that they and others can investigate.

Conclusions, implications for hydrogeology instruction, and future research

This study provides evidence that two spatial thinking skills are important in hydrogeology: visual penetrative ability and working in multiple frames of reference. On the surface, the need for spatial thinking skills might appear obvious, and it is likely that experienced hydrogeologists would come to this conclusion. However, the researchers argue that identifying what specific spatial skills have application in hydrogeology is a valuable endeavor that can inform education, training, and further investigation. This work suggests that visual penetrative ability and working with multiple frames of reference are particularly important skills that predict performance on typical hydrogeology problems required of undergraduate students. Especially at the introductory level, these skills appear to increase facility with visualizing the hydrogeologic environment using multiple perspectives and moving seamlessly between frames of reference. The researchers, therefore, encourage instructors to be more aware of the possible limitations of individuals with lower spatial thinking skills. Not everyone is able to easily visualize multiple perspectives simultaneously and this (addressable) limitation may not be immediately apparent for those for whom this comes easily. Furthermore, targeted spatial thinking training may yield positive results for both weak and strong spatial thinkers. Hydrogeology instructors may want to consider assessing the current state of students' spatial thinking skills by having students complete tests of visual penetrative ability and working in multiple frames of reference (Ormand et al. 2014; Piaget and Inhelder 1956). The National Association of Geoscience Teachers (NAGT) provides access to instruments and surveys on its website (NAGT 2023). Additionally, The Spatial Thinking Workbook (Ormand et al. 2017) includes teaching activities and instructional strategies that can be accessed through the website hosted by the Science Education Resource Center at Carlton College (SERC 2023).

Although development of curricula is outside the scope of this project, the research team is currently documenting how students work through the HT to capture instances of spatial thinking and to better understand how students use (or do not use) these skills. The researchers are particularly interested in the strategies that low spatial thinkers may employ when approaching problems, as this may give insight into pedagogical scaffolds that they can develop to assist struggling students. The researchers found an effective investigative method to be observing students in groups that combine low and high spatial thinkers because instances where students explain concepts to each other are especially revealing of the nature of student thinking.

As the presumably first study that has investigated spatial thinking in hydrogeology, this study provides primary empirical evidence of the importance of visual penetrative ability and working in multiple frames of reference in this discipline. The researchers hope that it illuminates the nature of these skills for hydrogeology practitioners and educators. The goal for future work is to combine these findings with qualitative research to investigate student use of these skills. This will represent an important next step toward understanding the development of complex skill and expertise in hydrogeology.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10040-024-02843-6.

Acknowledgements The researchers would like to thank Leilani Arthurs, Craig Nichols, Kamini Singha, and Tim Shipley for advising the project. They appreciate the contributions of Patrick Burke, who helped with data collection and the feedback of the reviewers, who greatly improved the manuscript. The researchers additionally thank Holly Michael, Miguel Valencia, and Chris Zahasky, who assisted with recruiting. Finally, they are immensely grateful to the project participants.

Funding This material is based upon work supported by the National Science Foundation under Grants Numbers DUE-2043616 and DUE-2225637. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

Declarations

Conflicts of Interest The authors have no conflicts of interest to disclose.

References

Alles M, Riggs EM (2011) Developing a process model for visual penetrative ability. Geol Soc Am Spec Pap 474:63–80. https://doi.org/10.1130/2011.2474(06)

Arthurs L, Hsia JF, Schweinle W (2015) The oceanography concept inventory: a semicustomizable assessment for measuring student understanding of oceanography. J Geosci Educ 63(4):310–322. https://doi.org/10.5408/14-061.1

Baker KM, Petcovic H, Wisniewska M, Libarkin J (2012) Spatial signatures of mapping expertise among field geologists. Cartogr Geogr Inf Sci 39(3):119–132. https://doi.org/10.1559/15230406393119

Bédard J, Chi MT (1992) Expertise. Curr Direct Psychol Sci 1(4):135–139. https://doi.org/10.1111/1467-8721.ep10769799

Buckley J, Seery N, Canty D (2018) A heuristic framework of spatial ability: a review and synthesis of spatial factor literature to support its translation into STEM education. Educ Psychol Rev 30(3):947–972. https://doi.org/10.1007/s10648-018-9432-z

Carroll JB (1993) Human cognitive abilities: a survey of factor-analytic studies, 1st edn. Cambridge University Press. https://doi.org/10.1017/cbo9780511571312.012

Cheng YL, Mix KS (2014) Spatial training improves children's mathematics ability. J Cogn Dev 15(1):2–11. https://doi.org/10.1080/15248372.2012.725186

Cohen CA, Hegarty M (2007) Sources of difficulty in imagining cross sections of 3D objects. In: Proceedings of the Annual Meeting of the Cognitive Science Society, vol 29, no. 29. Nashville, TN, August 2007

Cohen CA, Hegarty M (2012) Inferring cross sections of 3D objects: a new spatial thinking test. Learn Individ Differ 22(6):868–874. https://doi.org/10.1016/j.lindif.2012.05.007

Coren S, Girgus JS, Erlichman H, Hakstian AR (1976) An empirical taxonomy of visual illusions. Percept Psychophys 20(2):129–137. https://doi.org/10.1037/e666602011-052

Davenport CE, French AJ (2020) The fundamentals in meteorology inventory: validation of a tool assessing basic meteorological conceptual understanding. J Geosci Educ 68(2):152–167. https://doi.org/10.1080/10899995.2019.1629193

Faul F, Erdfelder E, Buchner A, Lang A-G (2009) Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behavior Res Methods 41(4):1149–1160. https://doi.org/10.3758/brm.41.4.1149

Fetter CW (2013) Applied hydrogeology, 4th edn. Pearson, Upper Saddle River, NJ

Field A (2013) Discovering statistics using IBM SPSS statistics. Sage, Thousand Oaks, CA

French JW, Ekstrom RB, Price LA (1963) Manual for kit of reference tests for cognitive factors (revised 1963). Educational Testing Service, Princeton, NJ. https://doi.org/10.21236/ad0410915

Gao H, Tatomir AB, Karadimitriou NK, Steeb H, Sauter M (2022) Effect of pore space stagnant zones on interphase mass transfer in porous media, for two-phase flow conditions. Transp Porous Media 146(3):639–667. https://doi.org/10.1007/s11242-022-01879-0

Gleeson T, Allen DM, Ferguson G (2012) Teaching hydrogeology: a review of current practice. Hydrol Earth Syst Sci 16(7):2159–2168. https://doi.org/10.5194/hess-16-2159-2012

Hambrick DZ, Libarkin JC, Petcovic HL, Baker KM, Elkins J, Callahan CN, Turner SP, Rench TA, LaDue ND (2012) A test of the circumvention-of-limits hypothesis in scientific problem solving: the case of geological bedrock mapping. J Exp Psychol Gen 141(3):397–403. https://doi.org/10.1037/a0025927

Hegarty M, Canham MS, Fabrikant SI (2010) Thinking about the weather: how display salience and knowledge affect performance

in a graphic inference task. J Exp Psychol Learn Mem Cogn 36(1):37–53. https://doi.org/10.1037/a0017683
Hestenes D, Wells M, Swackhamer G (1992) Force concept inventory. Phys Teacher 30(3):141–158. https://doi.org/10.1119/1.2343497
Kali Y, Orion N (1996) Spatial abilities of high-school students in the perception of geologic structures. J Res Sci Teach 33(4):369–391. https://doi.org/10.1002/(sici)1098-2736(199604)33:4%3c369::aid-tea2%3e3.0.co;2-q
Kell HJ, Lubinski D, Benbow CP, Steiger JH (2013) Creativity and technical innovation: spatial ability's unique role. Psychol Sci 24(9):1831–1836. https://doi.org/10.1177/0956797613478615
LaDue ND, Ackerman JR, Blaum D, Shipley TF (2021) Assessing water literacy: undergraduate student conceptions of groundwater and surface water flow. Water 13(5):622. https://doi.org/10.3390/w13050622
Lee, Fetter, McCray (2003) Hydrogeology laboratory manual, 2nd edn. Pearson, Upper Saddle River, NJ
Libarkin JC, Anderson SW (2005) Assessment of learning in entry-level geoscience courses: results from the geoscience concept inventory. J Geosci Educ 53(4):394–401. https://doi.org/10.5408/1089-9995-53.4.394
Liben LS, Kastens KA, Christensen AE (2011) Spatial foundations of science education: the illustrative case of instruction on introductory geological concepts. Cogn Instr 29(1):45–87. https://doi.org/10.1080/07370008.2010.533596
Linnenbrink-Garcia L, Pugh KJ, Koskey KLK, Stewart VC (2012) Developing conceptual understanding of natural selection: the role of interest, efficacy, and basic prior. J Exp Educ 80(1):45–68. https://doi.org/10.1080/00220973.2011.559491
McGee MG (1979) Human spatial abilities: psychometric studies and environmental, genetic, hormonal, and neurological influences. Psychol Bull 86(5):889. https://doi.org/10.1037/0033-2909.86.5.889
McGrew KS (2009) CHC theory and the human cognitive abilities project: standing on the shoulders of the giants of psychometric intelligence research. Intelligence 37(1):1–10. https://doi.org/10.1016/j.intell.2008.08.004
McNeal PM, Petcovic HL, LaDue ND, Ellis TD (2019) Identifying significant cognitive factors for practicing and learning meteorology. J Operation Meteorol 7(1):1–26. https://doi.org/10.15191/nwajom.2019.0701
National Association of Geoscience Teachers (NAGT) (2023) Instruments and surveys. National Association of Geoscience Teachers Web Site. https://nagt.org/nagt/geoedresearch/toolbox/instruments/index.html. Accessed 13 Dec 2023
National Research Council, Singer SR, Nielsen NR, Schweingruber HA (2012) Discipline-based education research: understanding and improving learning in undergraduate science and engineering. National Academies Press, Washington, DC, pp 6–11. https://doi.org/10.17226/13362
Newcombe NS, Shipley TF (2014) Thinking about spatial thinking: new typology, new assessments. Studying visual and spatial reasoning for design creativity. pp 179–192. https://doi.org/10.1007/978-94-017-9297-4_10
Newcombe NS, Stieff M (2012) Six myths about spatial thinking. Int J Sci Educ 34(6):955–971. https://doi.org/10.1080/09500693.2011.588728
Ormand CJ, Manduca C, Shipley TF, Tikoff B, Harwood CL, Atit K, Boone AP (2014) Evaluating geoscience students' spatial thinking skills in a multi-institutional classroom study. J Geosci Educ 62(1):146–154. https://doi.org/10.5408/13-027.1
Ormand CJ, Shipley TF, Tikoff B, Dutrow B, Goodwin LB, Hickson T, Atit K, Gagnier K, Resnick I (2017) The spatial thinking workbook: a research-validated spatial skills curriculum for geology majors. J Geosci Educ 65(4):423–434. https://doi.org/10.5408/16-210.1
Petcovic HL, Ormand CJ, Krantz RW (2016) Earth, mind, and paper: field sketches as expert representations of the Hat Creek fault zone. In: Krantz RW, Ormand CJ, Freeman B, eds (eds) 3-D structural interpretation: earth, mind, and machine. AAPG Memoir 111, American Association of Petroleum Geologists, pp 173–189. https://doi.org/10.1306/13561992m1113675
Piaget J, Inhelder B (1956) The child's conception of space. Routledge, London. https://doi.org/10.4324/9781315006239
Pugh KJ, Linnenbrink-Garcia L, Koskey KL, Stewart VC, Manzey C (2010) Motivation, learning, and transformative experience: a study of deep engagement in science. Sci Educ 94(1):1–28. https://doi.org/10.1002/sce.20344
Rebelsky F (1964) Adult perception of the horizontal. Percept Mot Skills 19(2):371–374. https://doi.org/10.2466/pms.1964.19.2.371
Resnick I, Shipley TF (2013) Breaking new ground in the mind: an initial study of mental brittle transformation and mental rigid rotation in science experts. Cogn Process 14(2):143–152. https://doi.org/10.1007/s10339-013-0548-2
Schneider WJ, McGrew KS (2012) The Cattell-Horn-Carroll model of intelligence. In: DP Flanagan, PL Harrison (eds) Contemporary intellectual assessment: theories, tests, and issues. Guilford Press, New York, pp 99–144
Science Education Resource Center at Carlton College (SERC) (2023) Spatial thinking workbook. Science Education Resource Center at Carlton College Web Site. https://serc.carleton.edu/spatialworkbook/index.html. Accessed 13 Dec 2023
Scribner ED, Harris SE (2020) The mineralogy concept inventory: a statistically validated assessment to measure learning gains in undergraduate mineralogy courses. J Geosci Educ 68(3):186–198. https://doi.org/10.1080/10899995.2019.1662929
Shea DL, Lubinski D, Benbow CP (2001) Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. J Educ Psychol 93(3):604. https://doi.org/10.1037/0022-0663.93.3.604
Sorby SA (2007) Developing 3D spatial skills for engineering students. Australas J Eng Educ 13(1):1–11. https://doi.org/10.1080/22054952.2007.11463998
Uttal DH, Meadow NG, Tipton E, Hand LL, Alden AR, Warren C, Newcombe NS (2013) The malleability of spatial skills: a meta-analysis of training studies. Psychol Bull 139(2):352. https://doi.org/10.1037/a0028446
Vandenberg SG, Kuse AR (1978) Mental rotations, a group test of three-dimensional spatial visualization. Percept Mot Skills 47(2):599–604. https://doi.org/10.2466/pms.1978.47.2.599
Vasta R, Liben LS (1996) The water-level task: an intriguing puzzle. Curr Dir Psychol Sci 5(6):171–177. https://doi.org/10.1111/1467-8721.ep11512379
Vasta R, Belongia C, Ribble C (1994) Investigating the orientation effect on the water-level task: Who? When? and Why? Dev Psychol 30(6):893. https://doi.org/10.1037/0012-1649.30.6.893
Wai J, Lubinski D, Benbow CP (2009) Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. J Educ Psychol 101(4):817. https://doi.org/10.1037/a0016127
Wiley J (1998) Expertise as mental set: the effects of domain knowledge in creative problem solving. Mem Cognit 26(4):716–730. https://doi.org/10.3758/bf03211392