ABSTRACT

SABATINI, STEPHANIE M. Assessing Spatial Thinking Across the Geology Curriculum: A Mixed Methods Study. (Under the direction of Dr. David McConnell).

Spatial thinking encompasses an array of fundamental skills geology students rely on to visualize and interpret complex geological features and processes. However, the explicit teaching of spatial thinking skills is often lacking in undergraduate geology programs. Before embarking on our mixed methods inquiry, we conducted a thorough literature review focusing on the psychometric approaches utilized to assess spatial thinking skills in undergraduate geology courses. We examined 11 studies, detailing their methodologies and outcomes. The analysis revealed a predominant focus on skills within the intrinsic dynamic category, with an observed trend of higher scores among students in upper-level courses. However, notable gaps emerged, including the absence of studies that comprehensively investigated skills across all four spatial categories and of evidence regarding mechanisms for student gains in courses lacking targeted intervention.

In our mixed methods study, we utilized qualitative and quantitative approaches to thoroughly explore the integration of spatial thinking in six undergraduate geology courses and its impact on students' skill development throughout the curriculum. Qualitatively, we developed and utilized the Spatial Thinking Observation Protocol (STOP) to observe lecture content and examine laboratory assignments. Findings from this qualitative strand highlighted distinct patterns in spatial skill emphasis, with certain courses (e.g., Mineralogy and Sedimentology and Stratigraphy) emphasizing intrinsic skills such as categorization and disembedding, while other courses (e.g., Physical Geology, Geomorphology) focused more on extrinsic skills like relations between objects. Quantitatively, we administered the Spatial Thinking Ability Test (STAT) pre- and post-tests to assess students' spatial thinking skills across all four spatial categories. The

STAT evaluates seven skills, providing a comprehensive measurement of students' spatial abilities. Our quantitative analysis revealed significant improvements in students' spatial thinking skills throughout the undergraduate geology curriculum, with upper-level courses generally yielding higher scores compared to introductory ones.

The synthesis of qualitative and quantitative findings yielded valuable insights into the development and integration of spatial thinking skills in geology education. Through joint displays and triangulation analysis, we identified instances of convergence and divergence between the qualitative and quantitative findings. Convergence occurred when significant gains were observed in skills that were emphasized in the course content. Conversely, divergence highlighted cases where expected skill improvements did not align with observed test scores, indicating potential methodological constraints or skill development occurring outside observed courses. These findings underscore the complex interplay between course content, instructional methods, and skill acquisition in undergraduate geology courses, emphasizing the need for further research to refine teaching approaches and assessment instruments.

Assessing Spatial Thinking Across the Geology Curriculum: A Mixed Methods Study

by
Stephanie M. Sabatini

A dissertation or thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Marine, Earth, and Atmospheric Sciences

Raleigh, North Carolina
2024

APPROVED BY:

____
Dr. David McConnell
Committee Chair

____
Dr. Jenn Ayscue

____
Dr. Nicole LaDue

____
Dr. Arianna Soldati

DEDICATION

To my son, Jack, whose vibrant and joyous spirit always brightened my days.

And to Taylor Swift, whose music kept me inspired and entertained during the writing process.

BIOGRAPHY

Stephanie was born and raised in St. Louis, Missouri. After reading a science-fiction novel entitled *8.4* by Peter Hernon, ninth-grade Stephanie became fascinated with earthquakes and the societal impact of natural hazards, which launched her journey into the geosciences. She completed her Bachelor of Sciences in Geology and Geophysics in 2013 at Missouri University of Science & Technology. During her undergraduate years, her interests shifted into mineralogy, where she was able to teach the lab and conduct undergraduate research. After two years working in the environmental consulting industry, Stephanie realized that teaching science was her true calling. She spent the next several years teaching middle and high school science during which she serendipitously discovered the field of geoscience education, a combination of both her passions. Stephanie met Dr. David McConnell in June 2018 while visiting friends in North Carolina and then attended her first Earth Educators' Rendezvous with David the next month. A year later, Stephanie, her husband (Tony), son (Jack – nearly two at the time), and two dogs (Belle and Thor) moved to Raleigh, North Carolina so Stephanie could pursue her doctorate degree.

After five years, a global pandemic, and two new pets (Luna – cat, Daisy – dog), Stephanie is wrapping up her PhD program. During her time, she has served as the teaching assistant for multiple geology courses, MEAS GSA co-president, member of faculty search committee, and has had many great experiences at NC State and in Raleigh and looks forward to her future endeavors.

ACKNOWLEDGMENTS

I would like to first acknowledge my family, whose support and kindness made my dream possible. I have been pursuing graduate degrees most of my son's life and he has been my biggest cheerleader. Thank you, Jack, for always being able to put a smile on my face. In moving to North Carolina, my husband sacrificed his blossoming career to help me realize my ambitions. Thank you, Tony, for always putting your family first. I started my PhD program with two dogs and I am ending it with the same two, plus another dog and a cat. Thank you, puppers and kitty, for always being happy to see me and special thank you to my Belle, who has been with me through three degrees. Finally, I want to thank my parents for their unending support throughout my life.

Next, I would like to acknowledge my advisor, Dr. David McConnell whose mentorship and encouragement was instrumental in my persistence through this program. David acted not only as a research mentor, but also as a guide into the world of academia, providing guidance, advice, and opportunities for growth. I feel very honored to have been a part of the Geoscience Learning Process Research (GLPR) group and to have trained under David's supervision. Thank you, David, for every conversation, every cup of coffee, and every piece of advice and I look forward to finishing this project.

Next, I would like to acknowledge several significant figures from my PhD journey. Thank you, Tiffany VanDerwerker, for being my friend and allowing me to be myself. Thank you, Joyce Smith, for commiserating and celebrating with me as we trekked this path together. Thank you, other past GLPR (Doug Czajka, Jason Jones, Laura Lukes, and Katherine Ryker) GLPR-adjacent (Rachel Atkins) members, for your guidance and friendship. Thank you, Spatial

Thinking Research Interest Group (STRIG) members, for your inspiration and collaboration, and special thanks to Tim Shipley for reviewing an earlier version of Chapter 2.

Finally, I would like to acknowledge my doctoral dissertation committee members for their dedication to my growth as a researcher. Thank you, Dr. Arianna Soldati, for your enthusiasm and commitment to bringing geoscience education research into your classroom. Thank you, Dr. Nicole LaDue, for providing rigorous guidance to this project. Thank you, Dr. Jenn Ayscue, for providing qualitative research expertise to this project.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 2

CHAPTER 3

CHAPTER 4

CHAPTER 1 - INTRODUCTION

Spatial thinking skills play a crucial role in various disciplines, including geology, where the ability to conceptualize and manipulate spatial relationships is fundamental. While the importance of spatial thinking in STEM fields has been recognized globally, its explicit integration into undergraduate curricula remains limited (Shea et al., 2001; Wai et al., 2009). Geoscience Education Research (GER) has identified spatial and temporal reasoning as significant challenges in geoscience education (Ryker et al., 2018). The growing emphasis on workplace skills and competencies in STEM education underscores the need to address spatial thinking within undergraduate geology programs (Bralower et al., 2008; Cooper et al., 2015).

Traditionally, undergraduate geology programs have focused on delivering course content without explicit attention to the development of students' spatial thinking skills (Gold et al., 2018). This lack of emphasis persists despite evidence suggesting that spatial thinking skills can be cultivated through targeted training and interventions (Uttal et al., 2013). Moreover, studies have shown that the gap between achievement in STEM fields and spatial thinking ability diminishes with expertise (Hambrick et al., 2012; Uttal & Cohen, 2012), highlighting the importance of early spatial training for novices in STEM disciplines (Uttal & Cohen, 2012).

In recent years, there has been increasing interest in assessing and improving spatial thinking skills within undergraduate geology programs. Various studies have investigated baseline spatial thinking skills, interventions for skill improvement, and the relationship between spatial skills and geological concepts (Titus & Horsman, 2009; Ormand et al., 2014; Gold et al., 2018). However, these studies have often focused on specific subsets of spatial skills or utilized aggregate pre/post-tests, leaving certain spatial thinking skills under-examined.

To address these gaps, this dissertation aims to examine how spatial thinking skills are integrated into undergraduate geology curricula. Drawing on a mixed methods approach, this study seeks to characterize the incorporation of spatial thinking instruction across a range of geology courses. Additionally, through the administration of spatial tests at multiple points in the curriculum, the quantitative component of this research aims to provide insights into the development and integration of spatial thinking skills in undergraduate geology education.

The dissertation is structured as follows: Chapter 2 provides a review of existing psychometric assessments of spatial thinking in undergraduate geology courses. This includes frameworks for categorizing spatial skills, methodologies used in previous research on spatial skill assessment and instruction, as well as the results from these studies. In Chapters 3 and 4, we present our mixed methods investigation. Chapter 3 focuses on the qualitative aspect of the mixed methods study, detailing the development of the Spatial Thinking Observation Protocol (STOP), its application to characterize spatial thinking instruction in undergraduate geology courses, and the discussion of results and implications arising from this qualitative analysis. Chapter 4 presents our quantitative analysis, including the administration of the Spatial Thinking Assessment Tool (STAT) at different stages of the curriculum and the integration of the qualitative and quantitative data stands to elucidate how course content influences skill development. Chapter 5 provides a summary of the findings drawn from this investigation.

CHAPTER 2 – A REVIEW OF EMPIRICALLY ASSESSED SPATIAL THINKING SKILLS IN UNDERGRADUATE GEOLOGY COURSES

Prepared for submission to *Journal of Geoscience Education*

INTRODUCTION

Undergraduate geology programs have traditionally focused on delivering course content without explicit efforts to track the acquisition and evolution of students' skills throughout a degree program. However, in recent years, various STEM disciplines have started to consider changes to the undergraduate experience that include placing more emphasis on the development of workplace skills and competencies (e.g., Bralower et al., 2008; Cooper et al., 2015). Spatial thinking has been identified as one of several key aspects of geoscience thinking (Kastens et al., 2009). In addition to its importance in academics, industry leaders also value spatial thinking. (Egger et al., 2019). Gold et al. (2018) noted that the lack of an ability to practice spatial thinking may serve as a roadblock to entry into the geosciences and STEM disciplines (see also Uttal & Cohen, 2012).

Despite the apparent ubiquity of spatial thinking in geology as a discipline, nearly half (47%) of faculty teaching courses for geoscience majors reported that their students never practice 3D spatial thinking or do so only once or twice during the course (Egger et al., 2019). Typically, spatial thinking is not formally taught (Gold et al., 2018) but is instead assumed to either already exist in the minds of students when they are admitted into a program or to intuitively develop as they progress through their coursework. While some may assume that students' capacity for spatial thinking is fixed, Uttal et al. (2013) found that spatial thinking skills can be improved with targeted training. Titus and Horsman (2009) noted that the lack of instructional support for the development of spatial thinking skills may discourage otherwise talented students and influence them to switch majors.

The geoscience education community identified three grand challenges related to the development of spatial and temporal thinking in the geosciences (Ryker et al., 2018). These included calls to identify the essential spatial skills and geologic tasks within and across geoscience disciplines and to discern whether current spatial thinking assessment strategies are appropriate for use in the geosciences. In the last two decades, there has been several studies that investigated undergraduate geology students' baseline spatial thinking skills (Titus & Horsman, 2009; Ormand et al., 2014, Hannula , 2019), whether specific skills could be improved through various interventions (Ormand et al., 2017; Gold et al., 2018; Giorgis, 2015), and whether particular spatial thinking skills were related to specific geological concepts (Kreager et al., 2022; Polifka et al., 2022). These studies typically focused on a limited set of spatial reasoning skills or relied on aggregate pre/post-tests to assess multiple skills. Regardless of the format, several types of spatial thinking skills have either been under-examined or excluded from these studies.

This review will address how researchers have assessed undergraduate geology students' spatial thinking skills by examining and comparing frequently examined spatial skills, the methods used to assess these skills, and the patterns of student performance that are apparent from these research studies. Examining prior research illustrates variations in the use of spatial thinking terminology, raises questions regarding whether some spatial thinking tools are suitable for application in the geosciences, and aids in pinpointing gaps in terms of both the courses and the specific skills that have been studied.

Spatial Thinking Typologies and Skills

Spatial thinking in geology involves a variety of specific skills (Table 2-1) that incorporate a range of spatial concepts (e.g., orientation, scale), different types of spatial representations (e.g., block diagrams, cross-sections, stereonets, geologic maps), and transformations through time and space (e.g., rates, gradients, flows; NRC, 2006). For this review, we will utilize the spatial thinking typology described by Newcombe & Shipley (2015; Table 2-1). Their typology utilizes a 2x2 matrix developed from research in psychology, neuroscience, and linguistics which represents four distinct ways in which people think and reason about space (Chatterjee, 2008).

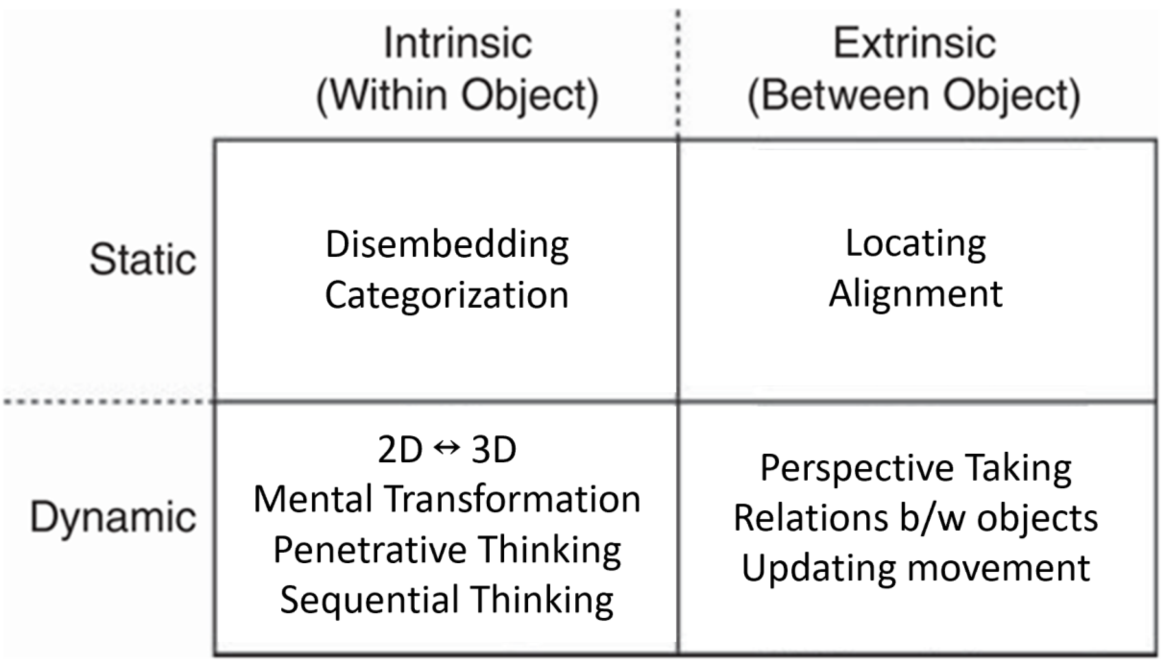
Table 2-1
A typology of spatial thinking skill categories (after Newcombe & Shipley, 2015)

Category	Characteristics	Examples of related geological skills
Intrinsic-Static	Spatial characteristics of objects (e.g., size, shape); relations between parts	Identifying fossils, minerals; interpreting outcrops, map patterns
Intrinsic-Dynamic	Transformations of objects, shapes; changing object orientations, positions without external reference; 2D/3D conversions	Interpreting crystal symmetry; cross-section construction; inferring change of object over time
Extrinsic-Static	Spatial relations between objects relative to an external frame of reference (e.g., map, horizontal or vertical)	Measure strike and dip; plot locations on a map
Extrinsic-Dynamic	Transformations of relations among moving objects and/or relative to surroundings	Interpreting how features (e.g., topography, geology) would appear from a different location and/or change with time

Newcombe and Shipley's (2015) typology offers a framework in which essential spatial skills in geology are mapped onto the four cognitive categories (Figure 2-1). Each matrix cell is defined by a combination of intrinsic (*within-object*), extrinsic (*between-objects*), static (*object(s) visualization)* and dynamic (*object(s) transformation)* properties (Table 2-1). Eleven distinct spatial thinking skills are encompassed by the matrix categories (Figure 2-1). The skills identified by Newcombe and Shipley (2015) also include skills identified in other geology-focused spatial thinking typologies which are discussed below. Geologists must apply multiple skills while in the field, classroom, or laboratory (Table 2-1) and some skills may be more significant for specific courses/disciplines than others. Students who excel in the geosciences will typically use a combination of both intrinsic and extrinsic skills.

Figure 2-1

Spatial thinking skills categorized using intrinsic/extrinsic and static/dynamic dimensions



Other spatial thinking frameworks were discussed by Kastens and Ishikawa (2006), Liben and Titus (2012), and Manduca and Kastens (2012), all of whom note the link between geoscience and cognitive science perspectives. Kastens and Ishikawa (2006) presented ten spatial tasks that are common across geology sub-disciplines and described each task from a novice and expert perspective. Their spatial tasks corresponded with the first three categories of Newcombe and Shipley's framework (Table 2-2).

Alternatively, Liben and Titus (2012) identified three broad areas within the geosciences that require spatial thinking: 1) Map Skills – including map reading and map navigation; 2) Orientations – including perceiving slopes and measuring orientations (e.g., strike and dip); and 3) Spatial Diagrams (e.g., cross sections and stereographic projections) and provided suggestions for teaching and future research directions in this area. Their three areas generally paralleled the latter three categories of the Newcombe & Shipley framework (Table 2-2).

Table 2-2

A comparison of spatial thinking frameworks

Categories	Newcombe & Shipley (2015)	Kastens & Ishikawa (2006)	Liben and Titus (2012)	Manduca & Kastens (2012)
Intrinsic-Static	Disembedding; Categorization	Describing an object's shape; Identifying/ Classifying an Object by its Shape; Ascribing Meaning to the Shape of an Object; Recognizing a Shape or Pattern Amid a Noisy Background		Disembedding
Intrinsic-Dynamic	2D ↔ 3D; Penetrative thinking; Mental transformation; Sequential thinking	Synthesizing 1D/2D Information to Create a 3D Mental Image; Envisioning the Process by which Objects Change Shape	Spatial diagrams	Mental rotation; Visual penetrative ability
Extrinsic-Static	Locating; Scaling; Space as Time proxy; Alignment	Recalling the Location and Appearance of Previously Seen Objects; Describing the Position and Orientation of Objects;	Map reading; Orientations	Object location
Extrinsic-Dynamic	Perspective taking; Relations among objects; Updating movement in space	Making and Using Maps; Envisioning the Process by which Objects Change Location	Map navigation	Mental animation; Perspective taking

Finally, Manduca and Kastens (2012) identified a framework that was a precursor to that of Newcombe and Shipley (2015) but identified fewer skills (Table 2-2).

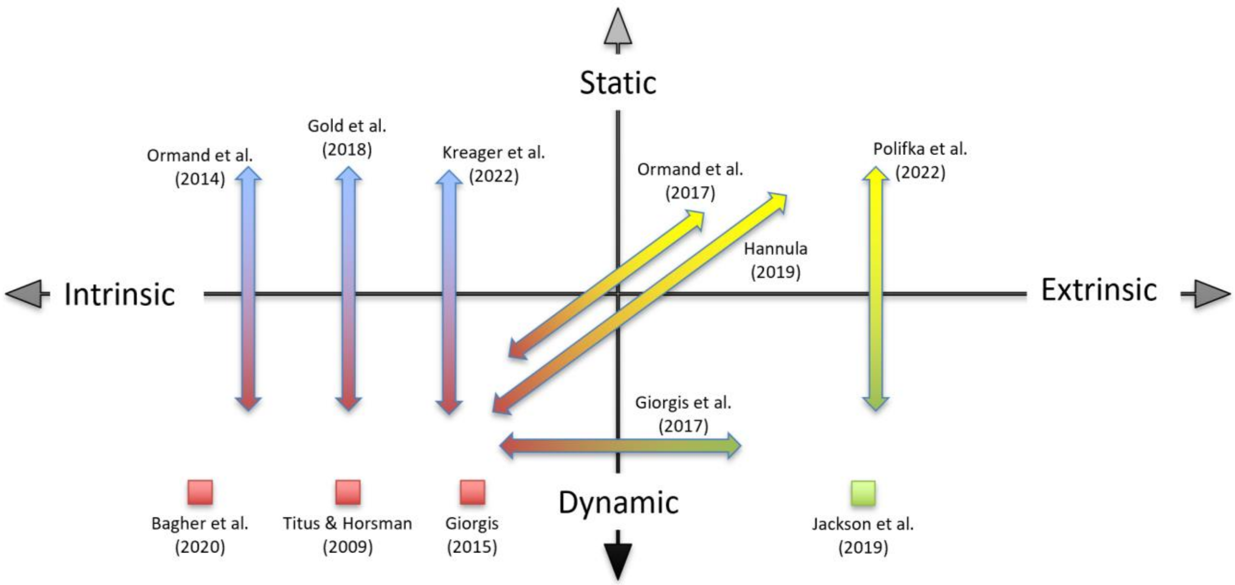
Previous Work

Studies selected for this review had to meet three criteria pertaining to the study population of interest, the spatial thinking skills investigated, and the use of psychometric spatial measures as a means of assessment. First, the population of interest for this review is undergraduate students enrolled in geology courses and will be referred to herein as geology students. While the study population includes geology majors, many of the studies described below also include non-major students in large-enrollment introductory geology courses (e.g., Titus & Horsman, 2009; Ormand et al., 2014; Gold et al., 2018; Polifka et al., 2022). In these studies, majors and non-majors were not analyzed as separate groups. Therefore, an exclusive focus on geology majors is not possible and would severely limit the possible research studies available for inclusion in this analysis.

Secondly, certain skills and skill categories comprise a majority of the research in this area (Figure 2-2). Consequently, this review will focus on studies investigating six spatial thinking skills (disembedding, penetrative thinking, 2D ↔ 3D manipulation, mental transformation, alignment, and perspective taking) that are discussed in multiple studies. An analysis of all the skills identified by Newcombe and Shipley (2015; Figure 2-1, Table 2-1) would significantly increase the length of this review. Firstly, we selected the most frequently examined spatial skills in the geosciences, which revealed three skills in the intrinsic dynamic category (penetrative thinking, mental transformation, and 2D ↔ 3D). To highlight the broad range of spatial thinking in the geosciences, the most frequently examined skills from each of the remaining categories were also selected (disembedding, alignment, perspective taking). We will discuss 11 studies investigating undergraduate geology students' spatial thinking skills that examined either one (4 studies) or two categories (7 studies) of Newcombe & Shipley's classification (Table 2-2).

Figure 2-2

Spatial thinking studies discussed in this paper by spatial thinking category



Finally, this analysis will focus on studies utilizing psychometric spatial thinking skill measures that have been developed and validated by cognitive scientists (e.g., Hidden Figures – Ekstrom et al., 1976; Water Level Test – Piaget & Inhelder, 1967), sometimes in collaboration with geoscience experts (e.g., Geologic Block Cross-Sectioning Test – Ormand et al., 2014). In one case, Giorgis et al. (2017) developed a topographic map assessment which has been included since it is very similar to the validated Topographic Map Assessment (Jacovina et al., 2014) and its questions assess students' perspective taking ability. In a recent systematic literature review, McLaughlin and Bailey (2022) identified several studies that indirectly assess undergraduate geology students' spatial thinking skills without the use of psychometric instruments. In these studies, researchers describe the use of instructional exercises, such as physical models, computer programs, and/or sketching and anecdotally discuss their effectiveness at improving students' spatial thinking. These studies would be classified as Practitioner Wisdom/Expert Opinion using the GER Strength of Evidence Pyramid (St. John & McNeal, 2018) and represent nearly 50% of McLaughlin and Bailey's studies. These examples will not be reviewed here due to their qualitative focus and absence of validated assessments.

Psychometric tests are essential in assessing cognitive ability and are based upon the assumption that cognitive processing functions exist within the brain as separate and measurable quantities (Eliot, 1987). A psychometric test isolates a singular spatial skill using unfamiliar, non-discipline specific items to reduce the cognitive bias from prior content knowledge. These instruments are readily administered with little training required and have demonstrated reliability to large numbers of subjects under controlled conditions to provide measurable outcomes (Schmidt & Hunter, 1998). Psychometric test results provide a snapshot of students' cognitive ability in a spatial skill.. This snapshot in turn informs the researcher how a student may perform on a broad array of problems that require the use of the skill (Eliot, 1987). Additionally, psychometric tests measuring various spatial thinking skills have been found to be predictive of success in STEM learning (Wai et al., 2009; Uttal et al., 2013).

SPATIAL THINKING IN UNDERGRADUATE GEOLOGY STUDENTS

We identified 11 studies which are classified as Case Studies and Cohort Studies according to the GER Strength of Evidence Pyramid (Table 2-3; St. John & McNeal, 2018). Case studies often focus on a single course and/or instructor while cohort studies may analyze a wider range of courses, institution types or student populations (St. John & McNeal, 2018). The Strength of Evidence Pyramid can be used to situate individual studies according to their level of data aggregation, study size, generalizability, as well as whether the information is unfiltered (original research) or filtered (analysis of published work). Studies situated at higher levels on the pyramid can be regarded as possessing more robust evidence (St. John & McNeal, 2018).

Table 2-3

Summary table of studies using spatial skill measures

Research article	Spatial Skills Assessed	Study Design	Strength of Evidence
Titus & Horsman (2009)	2D ↔ 3D Mental Rotation Penetrative Thinking	Pre-/Post- tests	Cohort Study
Ormand et al. (2014)	Disembedding Mental Rotation Penetrative Thinking	Pre-/Post- tests	Cohort Study
Giorgis (2015)	2D ↔ 3D Mental Rotation Penetrative Thinking	Experimental Intervention with Pre-/Post- tests	Case Study
Giorgis et al. (2017)	Mental Rotation Perspective Taking*	Experimental Intervention with Post- tests only	Case Study
Ormand et al. (2017)	Mental Rotation Penetrative Thinking Alignment	Intervention with Pre-/Post- tests	Cohort Study
Gold et al. (2018)	Disembedding Mental Rotation Penetrative Thinking	Experimental Intervention with Pre-/Post- tests	Cohort Study
Hannula (2019)	Penetrative Thinking Alignment	Pre-/Post- tests	Case Study
Jackson et al. (2019)	Perspective Taking*	Experimental Intervention with Post- tests only	Case Study
Bagher et al. (2020)	Penetrative Thinking	Experimental Intervention with Post- tests only	Case Study
Kreager et al. (2022)	Disembedding 2D ↔ 3D	Correlational analysis	Cohort Study
Polifka et al. (2022)	Alignment Perspective Taking	Correlational analysis	Case Study

Note. *Not assessed psychometrically.

In the following sections, we will discuss how researchers defined specific skills in each of the four spatial thinking categories, the methods used to assess students' aptitude with the skills, and the results of their investigations. This review will conclude with a short examination of the patterns seen across the four skill categories and will address gaps in the current body of knowledge on spatial thinking in undergraduate geology students.

Intrinsic Static Skills: Disembedding

Disembedding is one of two intrinsic-static spatial skills (Figure 2-1). Newcombe and Shipley (2015) defined disembedding as "isolating and attending to one aspect of a complex display or scene" (p.186). This skill is crucial in many geology tasks, such as interpreting seismic reflection data (Kreager et al., 2022), identifying rock and mineral properties in hand sample and thin section, and directing attention to important outcrop patterns in the field (Ormand et al., 2014). Essentially, disembedding is used when a geologist pays attention to important geologic features or patterns, while ignoring non-pertinent, distracting information.

Definitions

Each of the research studies summarized here chose to operationalize disembedding in different ways. Ormand et al. (2014) defined disembedding in the exact terms of Newcombe and Shipley (2015). Gold et al. (2018) identified disembedding as "the ability to isolate and attend to one aspect of a complex feature or landscape" (p. 2207). Kreager et al. (2022) applied a different term from the cognitive science realm: flexibility of closure, or the "ability to hold a given visual percept or configuration in mind as to disembed it from other well-defined perceptual material" (p. 755). The definition chosen by Newcombe and Shipley (2015) and Ormand et al. (2014) occupies a midpoint in its specificity by using the terms "complex display or scene" (p. 147), which zeroes in on the physical visualization piece and broadens the view to a range of scales.

Methods

Table 2-4 provides an overview of the three studies discussed in this section. Ormand et al. (2014) administered pre-/post-tests of spatial thinking measures to undergraduate geology students at different types of institutions across several different courses (Table 2-4) to provide a baseline of students' spatial thinking abilities. Gold et al. (2018) investigated the effects of spatial thinking practice modules using an intervention design and through collection of students' pre- and post- spatial abilities. Kreager et al. (2022) correlated what they defined as students' spatial thinking ability to their facility for completing a sequence stratigraphy task. Ormand et al. (2014) and Gold et al. (2018) compared spatial thinking of students enrolled in introductory and upper-level courses, whereas Kreager et al. (2022) focused on students enrolled in Sedimentology/Stratigraphy-type courses, one of which was at the graduate-level.

Table 2-4

Methods overview for disembedding studies

Research article	University type	Geology courses	Sample size	Disembedding instrument
Ormand et al. (2014)	LA	Introductory Geology	41	Hidden Figures[1], 16 items, no time limit specified
		Tectonics	15	
	RU	Introductory Geology	130	
		Structural Geology	17	
	LA	Sed/Strat	12	
Gold et al. (2018)	RU	Introductory Geology	E = 281 C = 224	Hidden Figures[1], 8 items, no time limit specified
		Unspecified Upper Level	E = 45 C = 42	
Kreager et al. (2022)	RUs (3)	Sed/Strat	78	Hidden Patterns[1], 200 items, 3 min

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, [1]Ekstrom et al., 1976

The other main difference between these investigations was in how the psychometric instruments were administered. Ormand et al. (2014) used the Educational Testing Service's (ETS) Hidden Figures test (Ekstrom et al., 1976; see Figure 2-3a for example) in its entirety, while Gold et al. (2018) used half of the same test items, and neither specified if students were given a time limit as directed by Ekstrom et al. (1976). Kreager et al. (2022) administered the Hidden Patterns test (Ekstrom et al., 1976; see Figure 2-3b) to the specifications of ETS but mislabeled the test as the Hidden Figures test.

Though each investigation served a different purpose (e.g., baseline study, Ormand et al. 2014; task correlation, Kreager et al. 2022), there are also some shared features. All three studies examined multiple, relatively large sample populations (different university types and geology courses) and provided descriptive analyses of these sub-populations. Each study also used a psychometric measure that was developed as part of the ETS's suite of Factor-Referenced Cognitive Tests (Ekstrom et al., 1976). Additionally, disembedding was not the only spatial skill examined in these studies (see Table 2-3), with two of the three studies also assessing mental rotation and penetrative thinking (Ormand et al., 2014; Gold et al., 2018).

Results

Ormand et al. (2014) reported Introductory Geology students pre-test scores on the Hidden Figures test (Ekstrom et al., 1976) in the low to moderate 40% range with consistently higher pre-test scores recorded in upper-level courses (Table 2-5). In all but one case, post-test scores were higher but often not to a statistically significant degree (Table 2-5). Gold et al. (2018) did not differentiate pre- and post-test scores among different populations in their experimental and control groups (Table 2-5), but their reported Hidden Figures scores for the combined population were lower than those reported by Ormand et al. (2014). This difference may be due to the relative difficulty of the subset of items chosen by the researchers compared to the full suite of questions. Student perception surveys reveal that 52% of students believed that the spatial thinking modules improved their spatial thinking abilities (Gold et al, 2018). Although it represented a different test, the Hidden Patterns scores obtained by Kreager et al. (2022) were relatively consistent with those of Ormand et al. (2014). Additionally, they found a low but significant (r=0.26) correlation between students' disembedding test scores and their performance on a sequence stratigraphy task (Kreager et al., 2022). In general, the gains were inconsistent across studies and within a single study (Ormand et al., 2014), were rarely statistically significant, and explain a small part of the variance in an authentic geologic task.

Table 2-5

Results overview for disembedding

Research article	Sample population	Ave. Pre-Test Score	Ave. Post-Test Score	Gain
Ormand et al. (2014)	Introductory Geology (LA)	44.8%	51.2%	6.4%
	Tectonics (LA)	59.2%	65.8%	6.7%
	Introductory Geology (RU)	41.4%	54.9%	13.6*
	Structural Geology (RU)	50.0%	58.1%	8.1%
	Sed/Strat (LA)	54.2%	46.4%	-7.8%
Gold et al. (2018)	Introductory Geology (E) Upper-Level (E)	18.9%	22.1%	2.9%
	Introductory Geology (C) Upper-Level (C)			3.4%
Kreager et al. (2022)	Sed/Strat	NA	49.4%	NA

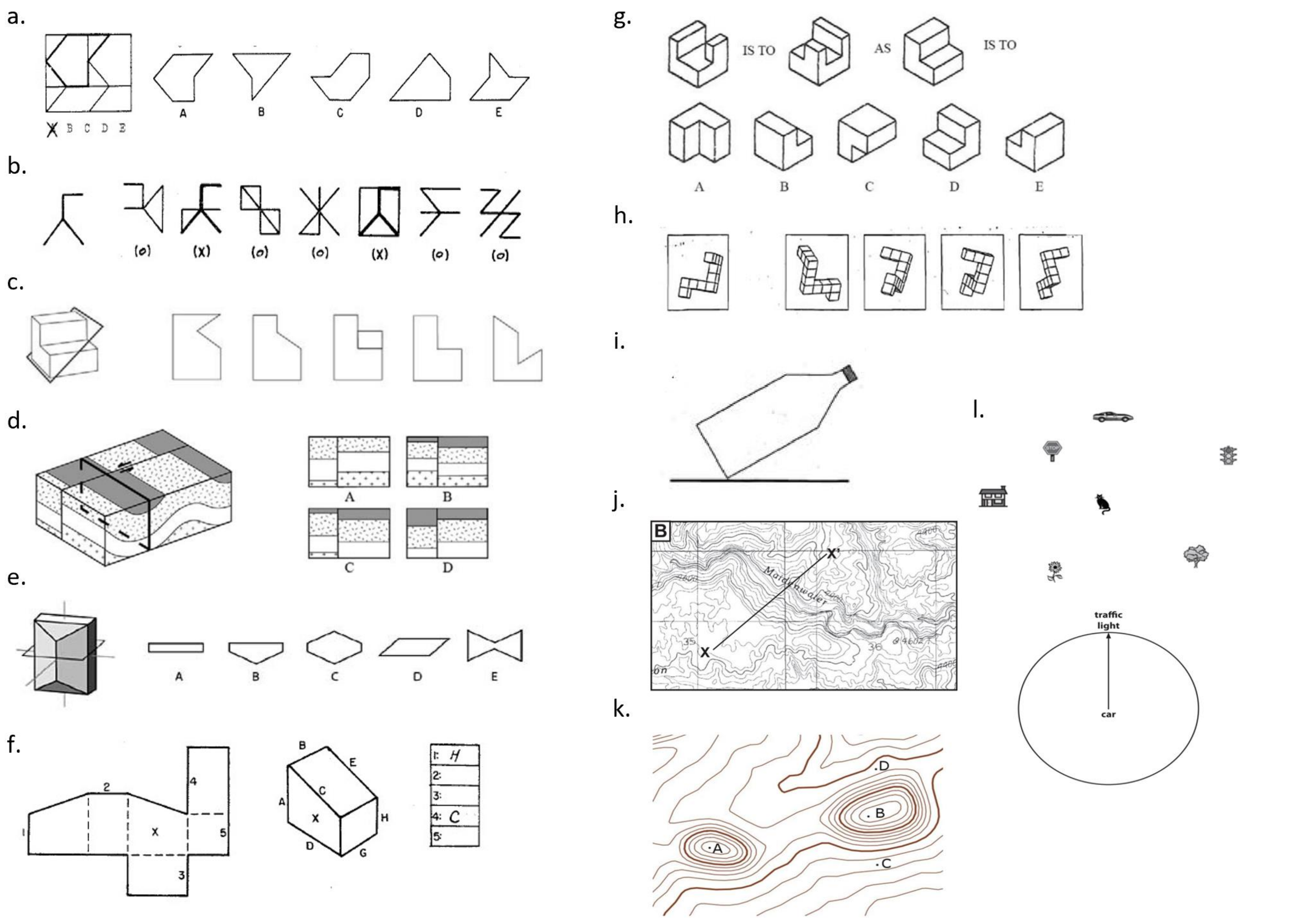
Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, *statistically significant

Intrinsic Dynamic Skills: Penetrative Thinking

Newcombe and Shipley (2015) described penetrative thinking as visualizing spatial relations inside an object (Figure 2-1; Table 2-2). This skill is frequently applied in geology to visualize geologic structures that are hidden from view beneath the ground surface. For example, geologists use penetrative thinking when interpreting and sketching cross-sections, visualizing the position of the groundwater table, and characterizing the general structure of the Earth (Ormand et al., 2014). Seven studies featuring penetrative thinking (Table 2-6) will be discussed below.

Figure 2-3

Examples of the types of questions from the spatial skills tests used in this study



Note. (a) The Hidden Figures Test asks students to identify which of five simple figures is present in a more complex pattern. In the example provided, figure A is present in the complex pattern to the right. (b) The Hidden Patterns test asks students to recognize a figure that is present in images made up of combinations of lines. For example, the figure on the left can be identified in two of the seven images to the right. (c) The Planes of Reference Test asks students to identify the correct shape of the intersection of the plane and the object. (d) The Geologic Block Cross-Sectioning Test asks students Subject to identify the correct geologic cross-section from the intersection of the plane and the block diagram. (e) The Crystal Slicing Test. Subjects are asked to identify the correct shape of the intersection of the plane and the crystal. (f) The Surface Development Test asks students to visualize how a piece of paper could be folded to form an object. Students match the numbered sides of the piece of paper with the letters corresponding to different sides in the completed object. (g) The Purdue Visualization of Rotations Test asks students to identify what the object at the top right would look like if rotated in the same fashion as the first object. (h) Mental Rotation Test asks students to mark the two objects that are the same as the one on the left. (i) The Water Level Test asks students to draw a line to show the top of the water surface assuming that the bottle is half full of water, (j) Giorgis et al. (2017) topographic map assessment asks students to identify topographic profile from X-X'. (k) Topographic Map Assessment asks students to imagine a position and indicate which letters are in their field of view. (l) Spatial Orientation Test asks students to draw a line indicating the direction a third item is located based on their location and direction. Figures a, b, and f and remaining figures are reproduced with permission from ETS and original publishing authors, respectively.

Definitions

As with disembedding, Ormand et al. (2014) used the same definition as Newcombe and Shipley (2015) and Ormand et al. (2017) described imagining a slice of an object. Several studies used different terminology in addition to penetrative thinking, such as visual penetrative ability (Titus & Horsman, 2009; Hannula, 2019), cross-sectioning (Bagher et al., 2020), penetrative visualization (Giorgis, 2015) and spatial visualization (Gold et al., 2018). Each definition referred to the mental imaging of an object's interior with specific callouts to objects' interior shapes (Hannula, 2019), the spatial relations inside an object (Bagher et al., 2020; Gold et al., 2018), or generally what is inside an object (Titus & Horsman, 2009; Giorgis, 2015). Bagher et al. (2020) added that the ability includes being able to transform three-dimensional data into a two-dimensional profile and Hannula (2019) emphasized the use of surficial clues in determining spatial relations within an object.

Methods

The research of Ormand et al. (2014) and Gold et al. (2018; Table 2-6) were introduced above in *Disembedding Methods*. Three of the studies provided baseline penetrative thinking measures for undergraduate geology students across multiple courses (Titus & Horsman, 2009; Ormand et al., 2014; Hannula, 2019; Table 2-6). The remaining four studies examined the effect of a spatial thinking intervention on the development of penetrative thinking skills. Interventions ranged from onetime classroom exercises (Giorgis, 2015; Bagher et al., 2020) to semester-long training (Ormand et al., 2017; Gold et al., 2018). Giorgis (2015) implemented a Google Earth map interpretation exercise and Bagher et al. (2020) utilized an immersive Virtual Reality simulation to visualize where earthquakes are generated in the subsurface. Ormand et al. (2017) and Gold et al. (2018) both developed a robust set of exercises that were used throughout the courses with the explicit focus on spatial thinking training. Three of the studies utilized a pre-/post-test design (Giorgis 2015; Ormand et al., 2017; Gold et al., 2018) and one only post-tested students (Bagher et al., 2020).

Titus and Horsman (2009) developed the Planes of Reference (Figure 2-3c) test after Crawford and Burns (1946) and this test was used subsequently in five of the studies reviewed (Table 2-6). Ormand et al. (2014) were motivated by findings from Kali and Orion (1996) to create the Geologic Block Cross-Sectioning Test (GBCT, Figure 2-3d) that was administered in four studies (see Table 2-6). Ormand et al. (2017) created and used the Crystal Slicing test (see Figure 2-3e) as an alternative to the Planes of Reference test, as it displays common geologic shapes rather than abstract geometric shapes. Five studies administered one of these tests, two studies administered two tests, and one study administered all three tests; (Table 2-6); each test was fully implemented in all studies. A time limit was indicated for some studies (Titus &

Horsman, 2009; Giorgis, 2015; Hannula, 2019) but not for others (Ormand et al., 2014, 2017; Gold et al., 2018; Bagher et al., 2020). Other differences amongst these studies included scrutinizing whether there is a test-retest effect (Ormand et al., 2014; Hannula, 2019) and a comparison between low and high content knowledge groups (Giorgis, 2015).

Table 2-6

Methods overview for penetrative thinking

Research article	University type	Geology courses	Sample size	Penetrative thinking instrument(s)
Titus & Horsman (2009)	LA	Introductory Geology Structural Geology Tectonics	58 26 73	Planes of Reference
Ormand et al. (2014)	LA	Tectonics	15	Planes of Reference GBCT
	RU	Structural Geology	17	
Giorgis (2015)	LA	Structural Geology	E = 51 C = 24	Planes of Reference
Ormand et al. (2017)	RU	Mineralogy Structural Geology	58 97	Planes of Reference GBCT Crystal Slicing
Gold et al. (2018)	RU	Introductory Geology	E = 281 C = 224	Planes of Reference
		Unspecified Upper Level	E = 45 C = 42	
Hannula (2019)	LA	Field Methods Structural Geology	83 51	GBCT
Bagher et al. (2020)	RU	Introductory Geology	E = 11 C = 10	GBCT

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group

Results

An overview of the results from the studies assessing penetrative thinking in undergraduate geology students is presented in Table 2-7 and is delineated by study and penetrative thinking instrument to facilitate comparisons between groups.

Results show consistent patterns across the penetrative thinking assessment studies. Except for Giorgis (2015), students enrolled in upper-level courses (e.g., Structural Geology, Tectonics) performed better than students enrolled in Introductory Geology courses. Planes of Reference test scores for Mineralogy students were closer to those of Introductory students than for students in upper-level courses (Ormand et al., 2017; Table 2-7). Planes of Reference and Crystal Slicing tests showed roughly similar gains for students (Ormand et al., 2017). Students in a Field Methods course showed intermediate results compared to Mineralogy and Structural Geology GBCT scores (Hannula, 2019).

Gold et al. (2018) results showed the effect of spatial training modules on pre/post-Planes of Reference test scores. The experimental group showed significantly higher gains than the control group (11.1% vs. 5.8%), suggesting training provided a greater boost to skill development than coursework alone. Giorgis (2015) also found that students in the experimental group with low and high prior content knowledge improved their Planes of Reference test scores compared to the control group whose scores decreased. Bagher et al. (2020) found that students who participated in a one-time virtual reality experience performed better on the GBCT than their control group counterparts (Table 2-7).

Test-retest analyses were performed in two studies with mixed results. Hannula (2019) found no test-retest effect using the GBCT; students who took the post-test in Field Methods (n=83) and subsequently completed a similar pre-test in Structural Geology (n=51) showed no statistically significant difference in their scores. Ormand et al. (2014), however found a statistically significant test-retest effect for the Planes of Reference test of 8.6% using a separate group of undergraduate psychology students. These results (n=27) were similar to those of the target population (see Table 2-7) which demonstrated a moderate effect size.

Table 2-7

Results overview for penetrative thinking studies

Research article	Test	Geology Courses	Ave. Pre-Test Score	Ave. Post-Test Score	Gain
Titus & Horseman (2009)	Planes of Reference	Introductory Geology (LA)	35-39%	42-49%	7-10%*
		Structural Geology (LA)	52-54%	63-64%	9-10%*
		Tectonics (LA)	53-59%	60-68%	7-9%*
Ormand et al. (2014)		Tectonics (LA)	59.5%	69.3%	9.8%*
		Structural Geology (RU)	57.5%	67.5%	10%*
Giorgis (2015)		Structural Geology (LA, E)	34%	45.3%	11.3%*
		Structural Geology (LA, C)	38.7%	35.3%	3.4%
Gold et al. (2018)		Introductory Geology (RU, E) Upper-Level (RU, E)	35.7%	44.4%	11.1%*
		Introductory Geology (RU, C) Upper-Level (RU, C)			5.8%*
Ormand et al. (2017)		Mineralogy (RU)	34.7-36%	46.7-51.3%	12-15.3%*
		Structural Geology (RU)	51.3-54.7%	64-64.7%	10-12.7%*
Ormand et al. (2014)	GBCT	Tectonics (LA)	56.6%	64.3%	7.7%
		Structural Geology (RU)	73.1%	74.4%	1.3%
Ormand et al. (2017)		Mineralogy (RU)	22.1-30%	38.6-39.3%	9.3-16.5%*
		Structural Geology (RU)	63.6-64.3%	70-72.4%	6.4-8.1%*
Hannula (2019)		Field Methods (LA)	38%	52%	14%*
		Structural Geology (LA)	49%	52%	3%
Bagher et al. (2020)		Introductory Geology (RU)	NA	56.8%	NA
Ormand et al. (2017)	Crystal Slicing	Mineralogy (RU)	44.7-49.3%	63.3-66%	16.7-18.6%*
		Structural Geology (RU)	69.3-72%	76-77.3%	5.3-6.7%*

Note. LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, *statistically significant

Intrinsic Dynamic Skills: 2D ↔ 3D

Newcombe and Shipley (2015) referred to the mental manipulation between two-dimensional space and three-dimensional space as “visualizing 3D from 2D” or vice versa. They describe this mental process as using information from a 2D or 3D image of an object to understand the spatial relations of that same object in a different dimensional configuration. Geology students employ this skill any time they are viewing a geologic figure in 2D and visualizing how the object would appear in 3D. For example, visualizing the 3D shape of a landform from a 2D topographic map or determining mineral habit when viewing a 2D drawing. The following studies will be reviewed in the next section: Titus and Horsman et al. (2009), Giorgis (2015), and Kreager et al. (2022).

Definitions

Titus and Horsman (2009) and Giorgis (2015) used the term spatial manipulation to describe visualizing 3D from 2D, which referred to the ability to mentally manipulate an image into another arrangement. This definition comes from Ekstrom et al.'s (1976) concept of spatial orientation. Similarly, Kreager et al. (2022) referred to spatial manipulation in the context of folding and unfolding.

Methods

An overview of the methods for the three 2D ↔ 3D studies are provided in Table 2-8. The studies discussed here were previously introduced in the *Disembedding and Penetrative Thinking Methods* sections.

Table 2-8

Methods overview for 2D ↔ 3D studies

Research article	University type	Geology courses	Sample size	2D ↔ 3D instrument
Titus & Horsman (2009)	LA	Introductory Geology Structural Geology Tectonics	58 26 73	Surface Development Test[1], 4 items, 3 min
Giorgis (2015)	LA	Structural Geology	E = 51 C = 24	Surface Development Test[1], 4 items, 3 min
Kreager et al. (2022)	RUs (3)	Sed/Strat	78	Surface Development Test[1], 6 items, 6 min

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, [1]Ekstrom et al., 1976

There are several similarities among the three studies. All three studies employ questions from the Surface Development Test (Figure 2-3f) and evaluate more than one skill in their assessments. However, the focus of all three studies remains solely on the intrinsic side of the spatial thinking typology matrix. Specifically, Titus and Horsman (2009) and Giorgis's (2015) studies are exclusively situated in the intrinsic dynamic cell of the matrix while Kreager et al. (2022) spans into the intrinsic static category (Figure 2-2). Additionally, Kreager et al. (2022) and Giorgis (2015) use the full suite of questions from the ETS test, while Titus and Horsman (2009) only utilize two-thirds of the questions as part of an aggregate test.

Results

Titus and Horsman (2009) observed that students in Structural Geology courses showed larger gains, while gains across other courses were about half as large but still statistically significant (Table 2-9). Giorgis (2015) and Titus and Horsman (2009) used the Surface Development test in the same course (Structural Geology) and post-test results were in a consistent range (68.8-79%), however Giorgis (2015) found smaller gains between the pre- and post-tests compared to the gains found by Titus and Horsman (2009).

Table 2-9

Results overview for 2D ↔ 3D

Research article	Sample population	Ave. Pre-Test Score	Ave. Post-Test Score	Gain
Titus & Horsman (2009)	Introductory Geology (LA)	37-39%	44-52%	7-13%*
	Structural Geology (LA)	49-54%	72-73%	19-23%*
	Tectonics (LA)	49-55%	60-63%	8-11%*
Giorgis (2015)	Structural Geology (LA, E)	60%	68.8%	8.5%*
	Structural Geology (LA, C)	71.5%	79%	7.5%
Kreager et al. (2022)	Sed/Strat (RU)	NA	58.7%	NA

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, *statistically significant

Additionally, Kreager et al. (2022) found the correlation between their sequence stratigraphy task and Surface Development test results to be moderately strong, with a positive correlation coefficient of 0.49. Giorgis (2015) noted that the control group scored higher than the experimental group both in pre- and post-assessments, however, the difference in gains observed between the groups might not have practical significance (Table 2-9).

Intrinsic Dynamic Skills: Mental Transformation

Newcombe and Shipley (2015) defined mental transformation as visualizing how an object will change over time. This definition encompasses changes in the size, shape, or orientation of an object. In geology, mental transformation occurs when students imagine the folding of rock layers, how a crystal might look from different orientations, or when they envision how a volcano would change after an explosive eruption. Six studies will be reviewed in this section: Titus and Horsman (2009), Ormand et al., (2014, 2017), Giorgis (2015), Giorgis et al. (2017), and Gold et al. (2018).

Definitions

Prior studies have focused almost exclusively on one specific aspect of mental transformation: mental rotation. Titus & Horsman (2009) and Giorgis (2015) explored what was formerly referred to as "spatial relations" (Shepard & Cooper, 1982), now recognized as mental rotation, particularly focusing on the ability to rotate an object around its center. Ormand et al. (2014, 2017) explored mental rotation as the process of visualizing the orientation of an object after it has been rotated. Giorgis et al. (2017) also referenced mental rotation but did not explicitly define it. Instead, they utilized it to measure an aspect of 3D visualization skills as part of a topographic map visualization task. Meanwhile, Gold et al. (2018) defined mental rotation as visualizing geological objects as they are rotated, such as during a compressional folding event. Each of these studies investigates mental transformation as change in the orientation of an object.

Table 2-10

Methods overview for mental transformation

Research article	University type	Geology courses	Sample size	Mental transformation instrument(s)
Titus & Horsman (2009)	LA	Introductory Geology Structural Geology Tectonics	58 26 73	PVRT[1], 10 items, 3 min
Ormand et al. (2014)	LA	Introductory Geology Upper-level courses	41 63	PVRT[1], 10 items, no time limit specified
	RU	Introductory Geology Structural Geology	130 17	
	LA	Sed/Strat	12	
Giorgis (2015)	LA	Structural Geology	E = 51 C = 24	PVRT[1], 10 items, 3 min
Ormand et al. (2017)	RU	Mineralogy Structural Geology	58 97	MRT-A[2], 24 items, no time limit specified
Giorgis et al. (2017)	LA	Introductory Geology	E = 94 C = 82	PVRT[1], 10 items, no time limit specified
Gold et al. (2018)	RU	Introductory Geology	E = 281 C = 224	PVRT[1], 10 items, no time limit specified
		Unspecified Upper Level	E = 45 C = 42	

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, [1]Guay et al. (1976), [2]Vandenberg & Kuse (1978)

Methods

Table 2-10 provides an overview of the methods for the six studies investigating mental transformation. All the study designs have been previously summarized in the preceding *Methods* sections except for Giorgis et al. (2017), which examined the effect of an augmented reality sandbox exercise on students' perspective taking abilities. Perspective taking is an extrinsic dynamic spatial skill, but the researchers administered a mental rotation test to identify students with high and low spatial visualization scores (Giorgis, et al., 2017).

Five studies implemented the same psychometric measure to assess mental rotation: the Purdue Visualization Test of Rotation (PVRT; Guay et al. 1976; Figure 2-3g). All five studies chose to use a subset of 10 questions from the full 30-question survey and cited time constraints as a concern for using an edited sample. It is unclear whether the same 10 items were chosen among the five studies. Ormand et al. (2017) did not use the PVRT (in contrast to their 2014 study) but instead chose the 24-item Vandenberg and Kuse (1978) Mental Rotation Test (MRT-A; Figure 2-3h) describing it as highly reliable especially regarding its test-retest effect. Several studies examined mental rotation in Introductory Geology and Structural Geology courses.

Results

The data from the various mental transformation studies reveals a consistent pattern of improvement in participants' skills from pre- to post-test. Structural Geology courses consistently showed higher gains in mental rotation abilities compared to other geology courses (Table 2-11). The results of the Mental Rotations Test (MRT) and the Purdue Visualization of Rotations Test (PVRT) were similar across the studies and courses. Giorgis et al. (2017) did not report descriptive statistics for their PVRT scores but separated high and low scorers into groups to compare their performance on a topographic map task.

Extrinsic Static Skills: Alignment

In Newcombe and Shipley's (2015) typology, alignment is defined as "reasoning about the spatial and temporal correspondence" (p.186) and is one of two extrinsic-static skills (Figure 2-1). From a cognitive science perspective, alignment refers to the use of analogical reasoning or the correspondence between two domains (Cheek et al., 2018). Geologists use alignment when taking the strike and dip of bedding or fault planes or when reasoning about the original horizontality of rock strata. In these examples geologists must align the current orientation of the rock with its initial orientation when the rock was formed, the horizontal reference frame. Three studies (see Table 2-12) will be discussed below.

Table 2-11

Results overview for mental transformation studies

Research article	Test	Geology Courses	Ave. Pre-Test Score	Ave. Post-Test Score	Gain
Titus & Horseman (2009)	PVRT	Introductory Geology (LA)	27-31%	34-40%	7-9%*
		Structural Geology (LA)	39-60%	52-70%	10-13%*
		Tectonics (LA)	42-57%	52-67%	10%*
Ormand et al. (2014)		Introductory Geology (LA)	41.5%	50.2%	8.8%*
		Upper-level courses (LA)	60.2%	69.7%	9.5%*
		Introductory Geology (RU)	49.2%	56.1%	6.9%*
		Structural Geology (RU)	60.0%	63.5%	3.5%
		Sed/Strat (LA)	37.5%	50.8%	13.3%*
Giorgis (2015)		Structural Geology (LA, E)	40%	46%	6%*
		Structural Geology (LA, C)	41%	46%	5%
Giorgis et al. (2017)		Introductory Geology (LA, E)	Scores not reported		
		Introductory Geology (LA, C)			
Gold et al. (2018)		Introductory Geology (RU, E) Upper-Level (RU, E) Introductory Geology (RU, C) Upper-Level (C)	39.2%	47.5%	6.7-9.6%*
Ormand et al. (2017)	MRT-A	Mineralogy (RU)	32.5-36.7%	42.1-50.8%	9.6-14.1%*
		Structural Geology (RU)	37.1-45.4%	52.6-63.8%	15.5-18.4%*

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group, *statistically significant

Definitions

Each study used different terminology for this extrinsic-static spatial skill and two studies highlighted the importance of distinguishing or perceiving horizontal from non-horizontal orientations. Ormand et al. (2017) also used the term alignment and described it as the act of recognizing horizontality within a tilted reference frame. Hannula (2019) used the “concept of horizontal" to denote alignment and provided no formal definition or description. She also noted the term "spatial perception" had been used previously to categorize this skill (Linn & Petersen, 1985; Voyer et al., 1995). Polifka et al. (2022) didn't refer specifically to alignment but more generally defined extrinsic-static spatial information as "the information between objects that does not change" (p.6). Given the interests of these three studies, we will narrow the scope of Newcombe & Shipley's (2015) definition to focus specifically on reasoning about horizontality.

Methods

These three studies shared two things in common: they occurred at research universities and utilized the Water Level Task (Table 2-12), which was originally developed by Piaget and Inhelder (1946). In this test, there is a drawing of a tilted bottle sitting on a horizontal surface and respondents are asked to imagine and draw the water line in the bottle (Figure 2-3i). The number of items differed in each study with Ormand et al. (2017) using the same number (12) as the original Piaget and Inhelder (1946) test, Hannula (2019) using 50% of items, and Polifka et al. (2022) using 25% of items. The alignment assessment was either untimed or a time was not specified.

Table 2-12

Methods overview for alignment studies

Research article	University type	Geology courses	Sample size	Penetrative thinking instrument(s)
Ormand et al. (2017)	RU	Mineralogy Structural Geology	58 97	Water level task, 12 items, time limit not specified
Hannula (2019)	LA	Field Methods Structural Geology	83 51	Water level task, 6 items, untimed
Polifka et al. (2022)	RU	Introductory Geology	87	Water level task, 3 items, untimed

Note: RU = Research University, LA = Liberal Arts College

Results

Ormand et al. (2017) demonstrated that Structural Geology students scored higher on the water level task than Mineralogy students (Table 2-13). Both groups though show little to no gain between pre- and post-test scores. There is however a 10% gain in the Water Level Task score in the Field Methods course (Hannula, 2019). While this finding is not statistically significant, a 10% gain may be practically significant, especially within the context of a course where the goal is to develop practical field skills, such as using a Brunton compass to take strike and dip of a tilted planar surface. Polifka et al. (2022) reported that Introductory Geology students earned relatively low scores on the Water Level Task, and that the task explained very little of the variance (0.003) in a plate tectonics knowledge activity.

Table 2-13

Results overview for alignment

Research article	Test	Sample population	Ave. Pre-Test Score	Ave. Post-Test Score	Gain
Ormand et al. (2017)	Water level task	Mineralogy (RU) Structural Geology (RU)	75.8-78.3% 90-94.2%	76.7-85% 88.3-93.3%	0.9-6.7% -1.7-$^{-}$0.9%
Hannula (2019)		Field Methods (LA)	55%	65%	10%
Polifka et al. (2022)		Introductory Geology (RU)	NA	50.3%	NA

Note: RU = Research University, LA = Liberal Arts College

Extrinsic Dynamic Skills: Perspective Taking

Perspective taking is “visualizing the appearance of a scene from a different vantage point” (p. 186; Newcombe & Shipley, 2015) and comprises part of the extrinsic-dynamic set of spatial skills (Figure 2-1). This skill emerges frequently during map interpretation, especially using topographic maps, where geologists must plot a navigable path from one location to another or imagine moving to a new perspective to visualize the appearance of landforms represented by topographic patterns. Geologists often practice perspective taking when they are not able to physically visit a location; they must visualize its appearance using a variety of data sets (e.g., topographic maps, satellite imagery). This skill is also applied during the determination of left-lateral or right-lateral strike-slip fault movement, where a geologist must imagine the relative displacement of features from different sides of the fault. Three studies have been reviewed below for their assessment of perspective taking.

Definitions

The investigators did not use the term perspective taking in the write-up of their studies. As with their treatment of alignment, Polifka et al. (2022) used the broad category “extrinsic dynamic spatial ability” to constitute perspective taking. Giorgis et al. (2017) and Jackson et al. (2019) did not name this dynamic skill directly and only the latter used a validated instrument for their assessment.

Table 2-14
Methods overview for perspective taking studies

Research article	University type	Geology courses	Sample size	Penetrative thinking instrument(s)
Giorgis et al. (2017)	LA	Introductory Geology	E = 94 C = 82	Self-created topographic map assessment, 15 items
Jackson et al. (2019)	LA	Field Methods Structural Geology	E = 295 C = 435	Topographic Map Assessment, 18 items, not delineated by skill
Polifka (2022)	RU	Introductory Geology	87	Spatial Orientation Test, 12 items, 5 minutes

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group

Methods

All three studies were conducted in introductory courses and Polifka et al. (2022) was the only one to administer an instrument that directly assessed perspective taking: the Spatial Orientation Test (Kozhevnikov & Hegarty, 2001; Figure 2-3l; Table 2-14). Giorgis et al. (2017)

and Jackson et al. (2019) utilized topographic map assessments to indirectly assess students' ability of this skill. While Giorgis et al. (2017) did not use the already developed and validated Topographic Map Assessment (Jacovina et al., 2014), they used similar items that were directly related to perspective taking (Figure 2-3j). These items included identifying landforms, choosing the correct topographic profile, and practicing line of sight judgment. Jackson et al. (2019) used the formal Topographic Map Assessment (Figure 2-3k; Jacovina et al., 2014), but did not differentiate between perspective taking items and non-perspective taking items. Because of this, the results from Giorgis et al. (2017) and Jackson et al. (2019) are difficult to evaluate in the context of this review. Both studies also only used their topographic map assessments as a post-test to assess the effectiveness of a one-time augmented reality sandbox activity.

Table 2-15

Results overview for perspective taking

Research article	Test	Sample population	Ave. Pre-Test Score	Ave. Post-Test Score	Difference
Giorgis et al. (2017)	self-TMA	Introductory Geology (LA)	NA	E = 55.3% C = 56%	0.7%
Jackson et al. (2019)	TMA	Introductory Geology (LA)	NA	E = 57.8% C = 59.2%	1.4%
Polifka (2022)	SOT	Introductory Geology (RU)	NA	72.04%	NA

Note: LA = Liberal Arts College, RU = Research University, E = Experimental group, C = Control group

Results

Findings from these studies included that a one-time augmented reality intervention produced no statistical difference between experimental and control groups (Giorgis et al., 2017; Jackson et al., 2019; Table 2-15). In both cases, the control groups scored slightly higher on the topographic map assessments (Table 2-15), but results were not statistically significant. Polifka et al. (2022) showed that very little variance in their plate tectonics knowledge task was explained by the Spatial Orientation Test.

DISCUSSION

The following section provides a comparison across the four spatial skill categories that were reviewed here including a discussion of the limitations. As a reminder, at least one skill was chosen from each cell of the 2x2 matrix presented by Newcombe and Shipley (2015): intrinsic-static (disembedding), intrinsic-dynamic (penetrative thinking, 2D ↔ 3D**,** mental transformation), extrinsic-static (alignment), and extrinsic-dynamic (perspective taking).

Terminology

This review highlights the significant diversity in how researchers have approached and defined spatial skills, leading to variations in terminology and operationalization across studies. These trends underscore the complexity and multidimensionality of spatial skills research within the field of geoscience education. Newcombe & Shipley's (2015) spatial thinking typology, specifically designed for geoscience education researchers, has not been fully utilized in studies published after 2015, particularly in recent research. This suggests a potential missed opportunity for comprehensive and standardized assessments of spatial skills within geology courses.

One notable issue is the tendency toward overgeneralization or specificity in some studies. For instance, Polifka et al’s (2022) analysis could be interpreted to suggest that extrinsic static and extrinsic dynamic categories were represented by two skills. One challenge in doing spatial thinking research is the ability to thoroughly characterize and assess all aspects of a particular skill. For example, most studies investigating mental transformation focused on mental rotation, one aspect of mental transformation and did not consider other aspects such as mental deformation (size and shape changes) (e.g., Resnick & Shipley, 2013). In a similar vein, Kreager et al. (2022) focused on the folding and unfolding aspect of 2D ↔ 3D spatial manipulation, which does not include other manipulations such as visualizing the shape of a landform from a topographic map or visualizing a 2D map projection from a 3D globe. Given the limited number of studies available for review, and the relatively narrow scope of some of them, certain aspects of some skills could not be fully assessed.

To address these concerns, future research could benefit from employing Newcombe & Shipley's (2015) spatial thinking typology more consistently and leveraging it to its full extent. This has the potential to lead to more comprehensive and comparable investigations of spatial skills within geology education. Additionally, promoting the use and refinement of a standardized spatial skills typology in geoscience education research has the potential to foster greater consistency and advancement in the understanding of spatial skills in the field. This also highlights the need for more investigations on spatial thinking skill development in geoscience classrooms. A larger dataset, especially one that includes skills that have so far received limited attention, could reveal which psychometric tests are appropriate for this population and provide more robust evidence on which courses/interventions lead to higher spatial thinking gains.

Data Patterns

Curricular-Patterns

There are several data patterns that can be discerned from these analyses. First, students in upper-level courses tended to perform better than students in introductory courses in disembedding, penetrative thinking, 2D ↔ 3D, mental transformation, and alignment (although there were some exceptions, noted in the analyses above). This determination cannot be evaluated for perspective taking as only introductory-level students were included in the three studies. Additionally, this review highlights that gains in spatial skills tend to be higher in intermediate and upper-level geology courses compared to introductory ones. It is, however, unclear why spatial skills exhibit similar or different levels of improvement in specific geology courses. For example, penetrative thinking gains are not always consistent in structural geology courses. These inconsistencies may arise for several reasons. In the instance of structural geology, the course content may be largely similar across institutions, however individual instructors may choose to omit or include topics or place a greater emphasis on certain topics or activities, thus affecting students' exposure to spatial thinking instruction. Instructional strategies, such as using active learning, may also play a role as they may unintentionally improve spatial thinking without using explicit interventions. The experimental design of interventions may also have been a factor, including variation in the quality of intentional and incidental training and testing procedures (i.e., difficulty of chosen test items, time limits, etc.). This lack of clarity could have curricular implications, as the goal of any undergraduate program is to provide students with adequate knowledge and training and we currently do not fully understand how and when students are gaining different spatial skills during their coursework.

Why do we see these patterns? One idea is that students improve different spatial skills in different geology courses. For example, we may see higher penetrative thinking gains in Structural Geology compared to Mineralogy due to the focus of subsurface geological structures and frequent cross-sectioning practice that often occurs in Structural Geology courses. Ormand et al. (2017) demonstrated that Structural Geology students scored higher on the Water Level Task than Mineralogy students (Table 2-13). These results hold merit in that there is anecdotally more extrinsic course content in Structural Geology compared to Mineralogy. Consequently, instructors may have deployed activities in courses that incidentally enhanced some spatial thinking skills over others.

Course sequence may also be playing a role, such as with the moderate penetrative thinking scores in Field Methods (Hannula, 2019) compared to the lower scores in Mineralogy and higher scores in Structural Geology (Ormand et al., 2017). This intuitively makes sense as students take the field course between Mineralogy and Structural Geology (Hannula, 2019). While we see higher scores in upper-level courses, in some cases, the gains lessen as students advance through the curriculum (i.e., penetrative thinking scores in Mineralogy and Structural Geology; Ormand et al., 2017). This could be attributed to a potential ceiling effect later in a student's academic experience as they accumulate more experience with spatial visualization. Polifka et al. (2022) showed that Introductory Geology students score the lowest on the Water Level Task, likely due to their novice status in geology and relatively little practice with the concept of horizontality in previous coursework.

Further, the type and frequency of intervention utilized in each study potentially played a significant role in shaping the observed gains. For example, one-time interventions, such as in Giorgis (2015) led to practically insignificant gains, whereas semester-long training led to greater improvements (Ormand et al., 2017). However, it is noted that post-test scores for students in Structural Geology courses that did not feature specific interventions (Titus & Horsman, 2009, Ormand et al, 2014) were often comparable with those of students in courses with interventions (Ormand et al, 2017, Gold et al, 2018; see Tables 2-7 and 2-11). It is also possible that selection bias is confounding some of these results. The one-time training exercises were typically implemented in introductory courses with non-majors, whereas the longer duration interventions took place in upper-level geology courses. However, selection bias is less likely to play a role in most upper-level courses which are populated with majors. This diversity in approaches calls for a systematic and consistent assessment of course content and instruction and how it relates to multiple spatial skills across various geology courses to gain a comprehensive understanding of spatial skill development. Such a unified approach could contribute to better comparisons and advancements in spatial thinking research in geoscience education. For example, “Grand Challenge #1” of the Community Framework for Geoscience Education Research’s section on spatial and temporal reasoning (Ryker et al., 2018) asks, *What skills and tasks are essential to the different specialties within the geosciences? What spatial and temporal reasoning skills map onto these specific tasks?* The geoscience community could benefit from the development of a framework that maps geoscience tasks with spatial skills and course work.

Instrumentation Patterns

Notably, statistically significant results were only obtained in pre/post-tests in most sample sets for the intrinsic dynamic instruments (Geologic Block Cross-Sectioning Test, Planes of Reference Test, Crystal Slicing Test, Surface Development Test, Purdue Visualization of Rotation Test, Mental Rotation Test) but not for intrinsic static, extrinsic static, or extrinsic dynamic tests (Hidden Figures Test, Water Level Task, and Topographic Map Assessment). This finding may be interpreted to indicate that the lack of gains were related to poor training exercises or, alternatively, that the tests used were not suitable for this population. Uttal et al. (2013) found similar positive results for tests in the intrinsic domain (e.g., mental rotation, 2D ↔ 3D) with most studies in their meta-analysis showing gains with moderate to large effect sizes. Their meta-analysis of studies using the Water Level Task revealed variable outcomes showing gains with small to large effect sizes compared to the geoscience-focused investigations in this review. While most of the gains in the three disembedding studies were not statistically significant, Pallrand and Seeber (1984) found statistically significant gains during their experimental intervention study in physics classrooms using the Hidden Figures Test. The ubiquity of disembedding in geology and the apparent surface validity of the Hidden Figures test suggests that geology students should show improvement over time. Similarly, past work has shown the Water Level Task as appropriate to measure college students' alignment skills (Liben et al., 2011), especially in a geologic context. Research on perspective taking in non-geology settings also supports the Spatial Orientation Test as a suitable psychometric instrument (Atit et al., 2016; Tarampi et al., 2018). These findings suggest the need for additional investigations for the non-intrinsic dynamic skill sets, including but not limited to, a careful examination of how frequently specific spatial skills are featured in individual courses.

There may also be issues related to the selection of test items and assessment strategy. For example, Gold et al. (2018) reported lower Hidden Figures scores than those reported by Ormand et al. (2014). This difference may be due to the relative difficulty of the subset of items chosen by researchers compared to the full suite of questions. Researchers may have administered these spatial thinking tests under different conditions, such as timed/untimed or may have differed in ways the test instructions were presented to students. Polifka et al. (2022) used a virtual reality training exercise and task items that did not include students needing to identify a horizontal frame of reference, which may have resulted in the task having a low correlation with the Water Level Task scores. These inconsistencies show the need for agreement on reliable test items and fidelity to testing procedures. Often decisions related to testing procedures were logistical, but these decisions may have had unintended consequences. This highlights the need for delineating the spatial skills needed to complete authentic geologic tasks, as noted by Ryker et al.'s (2018) Grand Challenge #1.

Addressing Gaps

None of the studies discussed here assessed skills from all four categories of Newcombe and Shipley's (2015) matrix (Figure 2-2). Several studies were focused on a single category (Titus & Horsman, 2007; Giorgis, 2015; Jackson et al., 2019; Bagher et al., 2020). Studies that examined at least two categories sampled exclusively from the intrinsic side of the matrix (Ormand et al, 2014; Gold et al., 2018; Kreager et al, 2022), from the dynamic pair of categories (Giorgis et al., 2017, Hannula, 2019) or occasionally from the extrinsic pairing (Polifka et al., 2022) or a combination of extrinsic and intrinsic factors (Ormand et al., 2017). In general, there are few studies that span across the intrinsic/dynamic divide in the undergraduate geology student population and even fewer studies that utilize psychometric instruments.

Further, there are no studies in geology of five of the 11 skills identified by Newcombe and Shipley (2015, Figure 2-1: categorization, sequential thinking, locating, relations among objects, and updating movement). Further, two aspects of alignment (scaling, space as proxy for time) were also not represented. This indicates that researchers often investigate skills that have previously been assessed in geoscience literature, perhaps due to the social nature of science or the available instrumentation. There are some recent exceptions, such as Czajka and McConnell (2018) and Resnick et al. (2017), which investigate space as proxy for time and scaling, respectively. Future studies should investigate multiple skills across the 2x2 matrix, with particular attention given to skills that have not been represented in the literature, especially categorization and updating movement (Table 2-2), as they are commonly taught and practiced skills in undergraduate geology courses.

Future research should also focus on employing a mixed methods design that quantitatively measures students' spatial thinking ability via validated instruments, but also qualitatively through student and instructor interviews and course observations. This will provide a deeper examination of what spatial thinking skills are taught and which skills are more likely to be impacted by different courses. This research would also inform instructors where students are at in their skill development and has the potential to influence how tasks are taught and performed, Qualitative data could also provide valuable insight into developing instruments to assess the other spatial skills identified by Newcombe and Shipley (2015), such as categorization and locating. Finally, researchers seeking to use aggregate skill tests, composed of elements of multiple instruments, are encouraged to assess spatial skills from all four quadrants of the Newcombe and Shipley (2015) typology to fully describe undergraduate geology students' spatial thinking abilities.

CHAPTER 3 – CHARACTERIZING SPATIAL THINKING INSTRUCTION IN UNDERGRADUATE GEOLOGY COURSES USING AN OBSERVATION PROTOCOL

Prepared for submission to *Geosphere*

INTRODUCTION

There has been a global emphasis in developing the science, technology, engineering, and math (STEM) workforce over the last several decades. One strong predictor of success in STEM is spatial thinking ability (Wai et al., 2009; Shea et al., 2001). The Geoscience Education Research (GER) community also identified spatial and temporal reasoning as one of its "Grand Challenges" that requires a sustained research effort (Ryker et al., 2018). Spatial thinking encompasses a variety of distinct skills to describe, visualize, and transform the spatial characteristics (e.g., scale, shape, orientation) of objects and spatial relations between objects over a range of space and time scales (NRC, 2006). Explicit instruction in spatial thinking does not typically occur in undergraduate geoscience classrooms. Instructors may unconsciously assume that spatial thinking is an innate ability that will organically blossom with progression through their coursework (Gold et al., 2018). While the latter may be at least partially true, Titus and Horsman (2009) noted that the lack of instructional support for the initial development of spatial thinking skills may discourage otherwise talented students with an interest in the geosciences and influence them to switch majors.

Fortunately, extensive research has shown that spatial thinking ability is not fixed in individuals and that it can be improved by targeted training and interventions (Uttal et al., 2013). Several studies have also found that the gap between STEM achievement and spatial thinking ability decreases with expertise in STEM fields (Hambrick et al., 2012; Uttal & Cohen, 2012). Therefore, spatial training would be most effective if it was targeted at novices in STEM fields where students must rely more heavily on intuitive spatial ability and less on domain-specific content knowledge. Such strategies have the potential to help increase the number of STEM undergraduates who can succeed in their chosen field (Uttal & Cohen, 2012).

Hegarty et al. (2012) convened a special meeting to discuss the role of spatial thinking across the college curriculum. They highlighted the need to continue investigating, developing, and refining best-practice spatial thinking training and instruction. In the last two decades, there have been several studies that have investigated improving undergraduate STEM students' spatial thinking skills through a variety of methods, such as isolated classroom exercises (Giorgis et al., 2017 – geoscience; Stieff et al., 2016 – chemistry), use of spatially-engaging technology or models (Kozhevnikov & Thornton, 2006 – physics; Jo et al., 2016 – geography), spatial thinking instructional exercise sets or modules (Ormand et al., 2017 – geoscience; Carlisle et al., 2015 – chemistry; Hoyek et al., 2009 – anatomy) or full courses dedicated to or emphasizing aspects of spatial thinking (Sorby, 2009 – engineering; Hannula, 2019 – geoscience). However, while recognizing the need for improved spatial thinking instruction, there have been no studies that have explicitly attempted to assess how spatial thinking skills are featured in a range of courses across a standard undergraduate curriculum. Establishing a baseline that identifies which spatial skills are currently emphasized in courses versus those that receive relatively little attention, can help us better identify where to spend time and resources as we seek to enhance undergraduate students spatial thinking skills. This manuscript will focus on the baseline incorporation of spatial thinking instruction in an undergraduate geology curriculum using an observation-based protocol.

BACKGROUND

Spatial Thinking in Geology

Undergraduate students exhibit a wide range of natural abilities and therefore often begin their academic career in geology with a variety of spatial thinking skills (Kali & Orion, 1996; Ormand et al., 2014). Students who score well on tests of one type of skill may score poorly on others, even in the same category (e.g., mental rotation vs. penetrative thinking; Ormand et al., 2014). Efforts to assess and/or improve geoscience spatial skills have focused on characterizing students' baseline spatial skills in various courses and evaluating effectiveness of interventions aimed at improving students spatial thinking (e.g., Ormand et al., 2014; 2017; Gold et al., 2018).

Results from previous work found that students enrolled in upper-level courses (e.g., Structural Geology) earned higher scores than introductory geology students (Titus & Horsman, 2009; Ormand et al., 2014). While these patterns existed across studies, the amount of improvement varied among institutions for a given geology course. For example, using the same test, Ormand et al. (2014) recorded higher mental transformation scores for introductory geology and structural geology courses than Titus & Horseman (2009). Differences in the timing and mechanisms through which students acquire various spatial skills throughout their coursework may be due to contrasts in course content and instructional strategies (i.e., use of active learning). We therefore aim to gain clarity on this phenomenon by characterizing the instruction of spatial thinking in an undergraduate geology curriculum. In doing so, we hope to capture a picture of which spatial thinking skills are taught and practiced in various courses to better understand how geology curricula support students in developing this essential competency.

Research Questions

This investigation aims to characterize how spatial thinking is incorporated into undergraduate geology courses across the curriculum. To achieve this goal, we identified two research questions:

1. How do undergraduate geology instructors incorporate references to spatial thinking and the use of spatial thinking skills in their instruction and course materials?
2. How does spatial thinking content vary across the undergraduate geology curriculum?

By answering these two questions we seek to identify when and how often students are exposed to a range of spatial thinking skills during a suite of courses that are typically included in an undergraduate geology curriculum. This research seeks to inform instructors as they work to support the development of spatial thinking skills.

Observations to Characterize Instruction

Our research questions required that we carefully examine how spatial thinking was represented by instruction across a range of undergraduate geology courses. Consequently, we adopted a qualitative methodological approach that incorporated direct classroom observations of instruction. Observation is a common practice to characterize instruction as it occurs in-situ where the focus of the observer is on the instructor and their teaching (and sometimes students) without disturbing the natural environment in the classroom (Johnson & Turner, 2003). This qualitative practice allows the researcher to see what an instructor does and says without relying on potentially inaccurate self-report information. The data collected from observations provides context and description of instructional content and strategies. However, as with all qualitative methods, the observer may introduce bias into the collection and analysis of data and may be more time consuming than alternative quantitative methods (Johnson & Turner, 2003).

Additionally, some information is not able to be gleaned from observations, such as whether the students are truly engaged in learning at a specific moment and if the instruction is effective (Smith et al., 2013).

Many observational protocols have been developed to characterize different aspects of instruction (e.g., teacher beliefs, self-efficacy). Two observation protocols (COPUS, GTO) that have been used in undergraduate STEM classrooms are particularly relevant to our study. The Classroom Observation Protocol for Undergraduate STEM (COPUS; Smith et al., 2013) takes a systematic approach in characterizing teachers' and students' actions during a class period. The observer collects data every two minutes, in which they document the occurrence of one or more instructional activities coded as either "student is doing" or "instructor is doing". Examples of student activities include taking notes, asking/answering a question, or participating in a small group or whole class discussion. Instruction activities include traditional lecturing, providing feedback, or showing a demonstration (Smith et al., 2013). Both code categories also include "other" as an option, in which the observer can note the activity. The COPUS requires that the observer understands and can identify the various classroom actions and can be relatively straightforward to use (Smith et al., 2013).

Also relevant to the present study is an observation protocol developed to characterize how students use teamwork skills in undergraduate geology classrooms. Nyarko (2021) developed the Geoscience Teamwork Observation (GTO) protocol using the taxonomy of teamwork skills described by Marks et al. (2001) and several additional skills identified by geoscience employers. Redundant skills were consolidated, and their typology was reviewed by a combination of qualitative researchers who were either experts in teamwork research or field camp coordinators. The resulting protocol consisted of nine individual teamwork skills within three skill categories (transition skills, action skills, and interpersonal skills; Nyarko, 2021). Observers were to classify how often students practiced each skill in a class period using a Likert scale from one to four and record brief descriptions of how the skills were exhibited. Similarly to the COPUS, the observer must be knowledgeable about the associated teamwork skills and be able to identify their occurrence.

The GTO was specifically created to assess teamwork skills (Nyarko, 2021). Similarly, a spatial thinking observation protocol did not exist prior to this project, hence we developed our own to suit our research objectives using aspects from the COPUS and GTO. Our observation protocol used the Newcombe and Shipley (2015) spatial skill typology as an a priori coding scheme much like how the GTO used the skill taxonomy described by Marks et al. (2001). Like the COPUS, our instrument requires the observer to document the presence of spatial thinking instruction or practice every two minutes. A more thorough description of the development of our observation protocol can be found in the Methods section. Our goal was to provide a systematic method of characterizing spatial thinking in courses across the undergraduate geology curriculum.

METHODS

Context

This study took place at a public research university located in the southeastern U.S. and specifically within five required undergraduate geology courses (Physical Geology, Historical Geology, Mineralogy, Sedimentology/Stratigraphy, Structural Geology) and one popular elective course (Geomorphology). While the location of this study was chosen out of convenience to the researchers, the courses were purposefully chosen since they are generally required courses in most undergraduate geology curricula (Klyce & Ryker, 2022).

Table 3-1
Course descriptions

Course Name	Course Description
Physical Geology	Systematic consideration of processes operating on and below the earth's surface and the resulting features of landscape, earth structures, and earth materials. Occurrences and utilization of the earth's physical resources. Includes a related lab course.
Historical Geology	Utilization of the principles of geology to reconstruct and understand the earth's history. Geologic events that cause modification of the earth's crust, emphasizing North America. History of life and the environmental significance of changes in animal and plant life through geologic time. Includes a related lab course.
Introduction to Mineralogy	Introduction to the basics of Mineralogy (crystallography, morphology, crystallochemistry, optics, and systematics), with an emphasis on mineral identification both at the macro (hand sample) and micro (thin section) scale. Includes labs.
Introductory Sedimentology and Stratigraphy	Properties and classification of sediments and sedimentary rocks, geologic occurrences and origin of minerals and rocks formed by physical, chemical, and biologic processes at and near the Earth's surface. Principles of the division of stratified terrains into natural units, the correlation of strata and associated data, the interpretation of depositional environments, facies, and sequences, description of burial histories, and sedimentary basin analysis.
Structural Geology	Basic principles of geometric, kinematic and dynamic analysis as applied to fractures, shear zones, folds, and fabrics of deformed rock bodies. Considers both brittle and ductile realms of the crust from microscale to regional tectonics. Includes labs.
Geomorphology: Earth's Dynamic Surface	Landforms and the processes responsible for their origin. Emphasis on the geologic principles involved in interpreting the origin and evolution of various landforms, and discussion of North American geomorphic process. No labs.

Students typically take these courses in sequence, beginning with Physical Geology and Historical Geology as freshmen, moving into Mineralogy and Sedimentology/Stratigraphy as sophomores, and ending with Structural Geology and Geomorphology as juniors. (Some students may take Geomorphology during their Sophomore year.) Table 3-1 presents brief descriptions for each course. The Institutional Review Board approved this study (IRB 24198). The instructors of these courses were recruited to participate in this research and provided consent via a paper form.

Observation Protocol Development

We developed an observation protocol to extract the spatial thinking content and practice that is inherently embedded within the courses. We began by selecting a spatial skill typology that encompassed the breadth of skills that are essential in geoscience (i.e., Newcombe & Shipley, 2015). This typology is based upon work in cognitive psychology, neuroscience, and linguistics and provides a framework for how people think and reason about space (Chatterjee, 2008). The typology includes thirteen unique skills that can be categorized as intrinsic (within-object) or extrinsic (between-object) and static (visualization) and dynamic (transformation; see Table 3-2). Newcombe and Shipley's (2015) original typology specifies eleven skills but includes reference to two "important cases" under alignment. We have chosen to split "Alignment" into three cases that appear distinct on a surficial level (Alignment, Scaling, Space as a Proxy for time; Table 3-2).

Following the adoption of Newcombe and Shipley's (2015) typology, we sought to develop a way to systematically characterize spatial thinking instruction in our courses. The COPUS (Smith et al., 2013) provided a framework for our Spatial Thinking Observation Protocol (STOP), as it required observations to be recorded every two minutes thereby providing a comprehensive and semi-quantitative assessment of activities during a given class period. Thus, every two minutes the activity in the classroom was assigned a code for the presence of spatial thinking content.

Table 3-2
Spatial Skill Typology (Newcombe & Shipley, 2015)

Skill Category	Skill Name	Skill Description	Geoscience Example
Intrinsic Static	Disembedding	Isolating and focusing on a single aspect of a complex array	Detecting faults in a seismic reflection profile
	Categorization	Classifying objects or characteristics of objects by their spatial properties	Identifying grain size type in clastic sedimentary rocks
Intrinsic Dynamic	2D/3D	Visualizing the 3-D spatial relations of objects shown in 2-D and vice versa	Using a stereographic projection to visualize crystal faces
	Mental Transformation	Visualizing a change in an object's shape, size, or orientation	Envisioning the change in rock as it undergoes ductile deformation
	Penetrative Thinking	Visualizing the shape or spatial relations inside an object	Drawing a cross-sectional profile using a geologic map
	Sequential Thinking	Visualizing and remembering a series of object transformations	Deducing the tectonic processes responsible for a complex rock outcrop
Extrinsic Static	Locating	Identifying the past or present location of objects on maps or in space	Using Google Earth to pinpoint the location of the San Andreas fault
	Scaling	Reasoning about the size of objects and their properties	Comparing the heights of composite, shield, and cinder cone volcanoes
	Space Proxy for Time	Reasoning about how spatial relations correspond to temporal relations	Recognizing the order of events in a block diagram using relative time principles
	Alignment	Perceiving horizontal from a non-horizontal reference frame	Measuring the strike and dip of a bedding place using a Brunton compass
Extrinsic Static	Perspective Taking	Visualizing the appearance of an area from different points of view	Evaluating whether two map points are visible to each other in the field
	Relations among objects	Visualizing the spatial relations between multiple objects on maps or in space	Understanding the sequence of topographic features across an ocean basin
	Updating movement	Visualizing the movement of object(s) across space and time	Imagining the movement of tectonic plates at plate boundaries

If spatial thinking was present, codes were also ascribed for whether that content was presented as instruction or practiced by the students, and which skill category and specific spatial skill was the focus. Figure 3-1 provides an image of a sample data collection sheet used in a lecture session and Table 2-3 displays the list of codes. Laboratory assignments were coded using the same set of codes.

Figure 3-1
Example template of observation protocol data collection sheet

Course: Mineralogy Date: 9/23/21 Class #: 12 Class Type: Lecture Observer Name: Sabatini

Time	Spatial Skill Present? (Y/N)	Spatial Skill	Spatial Skill Type	Teacher Demo or Student Practice	Describe
1:30	N				
1:32	N				
1:34	Y	2D to 3D Categorization	ID/IS	Practice	draw & identify type of optical indicatrix
1:36	Y	2D to 3D Categorization	ID/IS	Practice	draw & identify type of optical indicatrix
1:38	Y	2D to 3D Categorization	ID/IS	Practice	draw & identify type of optical indicatrix
1:40	Y	2D to 3D Categorization	ID/IS	Practice	draw & identify type of optical indicatrix
1:42	N				
1:44	Y	Scaling	ES	Instruction	hand sample vs. thin section scale
1:46	Y	Mental Transformation	ID	Instruction	pleochroism- color changes as mineral rotates
1:48	Y	Penetrative Thinking	ID	Instruction	cut of mineral determines whether pleochroism occurs
1:50	Y	Relations among objects	ED	Instruction	orientations of polarizers in microscope
1:52	Y	Relations among objects	ED	Instruction	comparing relief of mineral to surrounding materials
1:54	Y	Mental Transformation	ID	Instruction	Becke line moves towards object with higher refractive index
1:56	Y	Mental Transformation	ID	Instruction	Becke line moves towards object with higher refractive index
1:58	Y	Relations among objects	ED	Instruction	Extinction angles compared to polarizers
2:00	N				
2:02	N				
2:04	Y	Categorization	IS	Instruction	parallel association vs. twinning
2:06	Y	Categorization	IS	Instruction	optical twinning types: polysynthetic, tatrtan, carlsbad
2:08	N				
2:10	N				
2:12	N				

Data Collection

The first author attended every lecture session for each of the six courses during the Fall 2021 (Mineralogy and Structural Geology), Spring 2022 (Historical Geology, Sedimentology/Stratigraphy, and Geomorphology), and Fall 2022 (Physical Geology) semesters. Each course's data was recorded in a separate Google Sheet document with tabs for every lecture period. For each class period, data collection began at the official class start time and proceeded every two minutes until the official end of class, unless the class period ended early. During that time spatial thinking content was recorded using the STOP (Figure 3-1; Table 3-3).

Table 3-3
STOP codebook

Code Category	Code	Code meaning
Spatial Thinking Present?	Y	Spatial thinking content present
	N	Spatial thinking content not present
Instruction or Practice?	Instruction	Spatial thinking content presented as course instruction
	Practice	Students using spatial thinking skills
Spatial Category	IS	Spatial thinking content focused on intrinsic static skill category
	ID	Spatial thinking content focused on intrinsic dynamic skill category
	ES	Spatial thinking content focused on extrinsic static skill category
	ED	Spatial thinking content focused on extrinsic dynamic skill category
Spatial Skill	Disembedding	Spatial thinking content focused on disembedding skill
	Categorization	Spatial thinking content focused on categorization skill
	2D/3D	Spatial thinking content focused on 2D/3D skill
	Mental Transformation	Spatial thinking content focused on mental transformation skill
	Penetrative Thinking	Spatial thinking content focused on penetrative thinking skill
	Sequential Thinking	Spatial thinking content focused on sequential thinking skill
	Locating	Spatial thinking content focused on locating skill
	Scaling	Spatial thinking content focused on scaling skill
	Space Proxy for Time	Spatial thinking content focused on space proxy for time skill
	Alignment	Spatial thinking content focused on alignment skill
	Perspective Taking	Spatial thinking content focused on perspective taking skill
	Relations among objects	Spatial thinking content focused on relations among objects skill
	Updating movement	Spatial thinking content focused on updating movement skill

The number of lecture sessions observed, scheduled lecture duration, and resulting number of observations can be found in Table 3-4. These values were tabulated using Stata. Exam days, project workdays, presentations, and field trips were not counted as lecture sessions, as the class time experience would likely be very different for each student and the instructor. In total, we collected over 5,000 observations and spent more than 170 hours in classrooms.

Table 3-4
Summary of data collected in lectures and labs.

Course Name	Lecture session duration	# of lecture sessions	# of individual lecture observations	# of lab assignments
Physical Geology	75 minutes	24	875	11
Historical Geology	50 minutes	38	869	13
Mineralogy	75 minutes	18	628	10
Sedimentology/ Stratigraphy	75 minutes	26	971	10
Structural Geology	50 minutes	33	794	9
Geomorphology	75 minutes	39	958	6*
Totals		**178**	**5,095**	**59**

Note: *Homework assignments were used in lieu of lab assignments

Each observation was coded in situ during the class. This process was not always straightforward. However, we used a set of heuristic rules and information gleaned from a review of the literature on the skills in Newcombe and Shipley's (2015) typology to establish an accurate and reliable procedure. First, we decided whether spatial thinking was occurring or not. If yes, a brief description of what the instructor was explaining or what the students were doing was recorded (Figure 3-1). This helped to focus on the main point of the lesson and the object(s) of interest at that time. Next, we indicated whether the spatial thinking content was presented as instruction or if students were practicing the skill. Student practice often included things like clicker questions or in-class activities. Finally, the spatial category and spatial skill were determined. Several questions helped to guide this determination, such as, "Was the main focus on one or more objects?", "Was the main focus on the characteristics of an object or its relationship to another object?", or "Was there mental animation of the object(s)' characteristics or relations"? We assumed that the code assigned to a specific time was representative of the entire 2-minute period and thus assigned two minutes to the specific skill.

For example, the Historical Geology instructor displayed a photograph of a rock sequence in the Grand Canyon and highlighted the presence of the "Great Unconformity" in the scene by describing that the different characteristics of the rocks above and below the unconformity, implying a gap in the rock record at this location. From this anecdote, we can surmise that spatial thinking is occurring from the vocabulary used. We can also conclude that this portion of the lesson can be coded as instruction, as the students are not actively engaged in an activity. Next, we established that the focus is on isolating the unconformity in the photo. In making this determination, we narrowed down that only one object (i.e., the unconformity) is the focus and there is not a transformation of the object, leading us to code this snippet of instruction as "intrinsic static". From there, we can utilize the definitions of the two intrinsic static skills to narrow down the specific spatial skill to disembedding.

Data was collected from lab assignments in a different way. Since laboratory sessions are much more dependent on individual and group work, the collection of data every two minutes was not possible. Instead, we extracted spatial thinking content from the lab assignment documents that students were asked to complete. The first author used the same a priori coding scheme to characterize the presence of spatial thinking in the lab assignments in all the courses. An exception was made for Geomorphology where six homework assignments were used instead, as the course did not have a corresponding lab. To code the labs, we carefully read through each of the lab activities and assigned codes to the instructions and questions within the document. Lab instructions made up a small portion of the activities compared to the larger component of the lab in which students were expected to either perform a task and/or answer questions. For lab instructions, codes were typically applied to each paragraph or a few sentences within a paragraph if the topic changed. For the student-directed question/answer portion, each question, sub-question, or change in procedure was assigned a code.

Data Analysis

The lecture and lab document content were coded using the set of a priori codes listed in Table 3-3. The codes were then compiled and quantitized using the data transformation strategy described by Onwuegbuzie and Teddlie (2003) in which qualitative codes are counted numerically for use in statistical analysis. Using Stata 15.1, code frequencies were tallied for each of the code categories within the six courses. By multiplying the number of codes by two, we determined the total amount of class time spent on each category or skill. Additionally, we calculated the proportion of lab assignments and lecture time dedicated to specific categories or skills by dividing the number of codes for each category or skill (e.g., disembedding, extrinsic dynamic) by the total number of observations for the course's lab or lecture.

Validity and Reliability

Given the qualitative nature of this research, preserving the credibility and confirmability was of the utmost importance (Creswell, 2003). Content validity of our coding scheme was achieved through the application of a priori codes derived from the robust spatial thinking typology developed by Newcombe and Shipley (2015). Their typology has been employed in numerous studies in the geosciences (Ormand et al., 2014; 2017, Gold et al., 2018, McLaughlin & Bailey, 2022) and was developed by cognitive scientists with expertise in spatial thinking and expert geoscientists. Consequently, the use of the typology in assessing the presence of spatial thinking in an undergraduate curriculum is valid.

The consistency of the observation protocol coding scheme was established through inter-coder reliability analysis. Three videos of instructors involved in this study teaching equivalent lectures were selected and shared with a second coder, a graduate student whose master's thesis research involved assessing undergraduate meteorology students spatial thinking skills. The selected videos exhibited a range of spatial thinking skills and modalities to ensure the integrity of the analysis. The first author provided a training session on how to employ the STOP using the first 30 minutes of a Historical Geology video. After the training session, the remaining portion of the recording and additional videos from the Mineralogy and Sedimentology /Stratigraphy courses were coded independently. Inter-coder reliability expressed as percent agreement values are shown in Table 3-5 and subdivided by the code categories presented in Table 3-3.

Table 3-5
Inter-coder reliability values

Code Category	Percent Agreement
Overall	85%
Spatial Thinking Present?	88%
Instruction or Practice?	88%
Intrinsic/Extrinsic	88%
Static/Dynamic	82%
Spatial Skill	79%

The first four categories are dichotomous: Is spatial thinking present: yes or no? If there is spatial thinking, is it presented as instruction or are students practicing it? Is the spatial thinking content focused on one or more objects? (i.e., intrinsic or extrinsic, respectively) Does the observed spatial thinking involve visualizing the characteristics of object(s) or mentally transforming those characteristics? (i.e., static or dynamic, respectively) The final code category represents the individual spatial skill that is being taught or practiced and can be narrowed down using the intrinsic/extrinsic and static/dynamic categories. For example, if the spatial thinking content is coded as intrinsic and static, then the spatial skill will be either disembedding or categorization. All inter-coder values are higher than Multon (2010) accepted range of 70%, indicating acceptable reliability of the STOP.

We anticipated the Spatial Skill category to have the lowest agreement as its determination partly relies on the coding of the Intrinsic/Extrinsic and Static/Dynamic categories and requires the greatest amount of specificity. Some class content/activities are easier to pin down than others. For example, in the Mineralogy course video, there was an activity in which students practiced identifying the crystal forms of crystal models. Both coders assigned this as the intrinsic static skill of categorization. Conversely, the extrinsic dynamic category appears more challenging. In the Sedimentology/Stratigraphy video the instructor was explaining how the petroleum exploration industry uses the relationship between the different sequence tracts to find oil deposits. Both coders agreed the instruction fell into the extrinsic dynamic category, but the first author coded this as relations among objects and the second coder assigned it the updating movement code. Other coding disagreements also took place at the static/dynamic category level. For example, the Mineralogy video showed the instructor describing the different Bravais lattice structures (i.e., primitive triclinic versus primitive monoclinic). The first author coded this instruction as the intrinsic static category of categorization, while the second coder assigned it the intrinsic dynamic code of 2D/3D. While both codes could be applied, the first author's declaration better matched to the objective of the instruction, which was to differentiate between the 14 Bravais lattice structures.

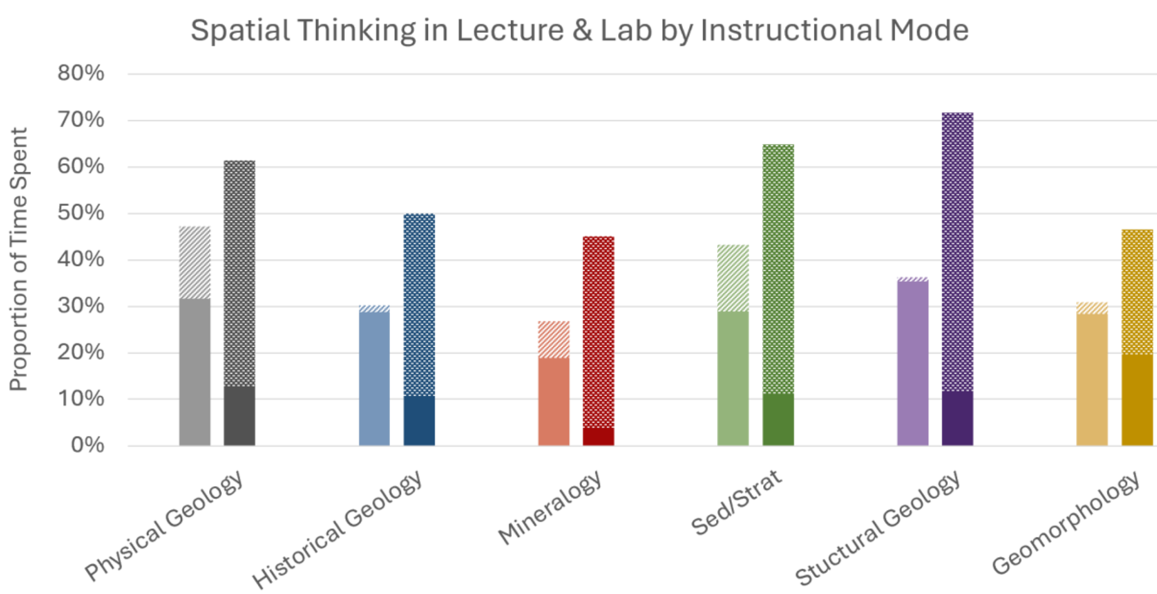
Given the broad agreement among the coders, we interpreted the protocol to be reliable and the rest of the STOP data presented here was solely reviewed by the first author. Given agreement values shown in Table 3-5, we inferred that deciphering between intrinsic and extrinsic (88% percent agreement) is a simpler process for the observers than between static and dynamic (82% percent agreement) spatial content.

RESULTS

Spatial thinking content is present in lectures from a minimum 27% of the time in Mineralogy to a maximum of 47% in Physical Geology (Figure 3-2). There is a greater amount of spatial thinking content present in the laboratory assignments for each course ranging from 45% in Mineralogy to 72% in Structural Geology (Figure 3-2). Uttal et al. (2013) asserted that spatial thinking skills could be enhanced through explicit training. Therefore, we were interested in how often students practice spatial thinking compared to listening to passive instruction. In lectures, students are actively engaged in spatial thinking practice between approximately 1% of the time in Structural Geology and 15% of the time in Sedimentology/Stratigraphy (Figure 3-2). In contrast, practice in lab ranges up to 60% of the time for Structural Geology.

Figure 3-2

Spatial thinking in lecture and lab by instructional mode



Note: The light-colored bar (left) for each course represents lecture. The dark-colored bar (right) represents labs. The solid portion of the bar represents spatial thinking content presented as instruction and the patterned portion represents student practice of spatial thinking. As a reminder, the lab data for Geomorphology represents course homework assignments.

Students' exposure to spatial thinking content can be further broken down by spatial category (Figure 3-3) and by individual spatial skills (Figure 3-4). Overall, students in lectures are exposed to more content from the intrinsic static category (~19 hours) than from any of the other categories (Figure 3-3). Intrinsic static lab assignments also average the largest proportion of spatial thinking tasks in the labs (~20%). Intrinsic static spatial thinking is most prevalent in the lectures and labs of Sedimentology/Stratigraphy (approximately 6.5 hours in lecture and 45% of lab assignments) and Mineralogy (approximately 3.5 hours in lecture and a 20% of lab assignments). It also features significantly in the Historical Geology lab (approximately 25%) and Structural Geology lecture (approximately 3 hours; Figure 3-3).

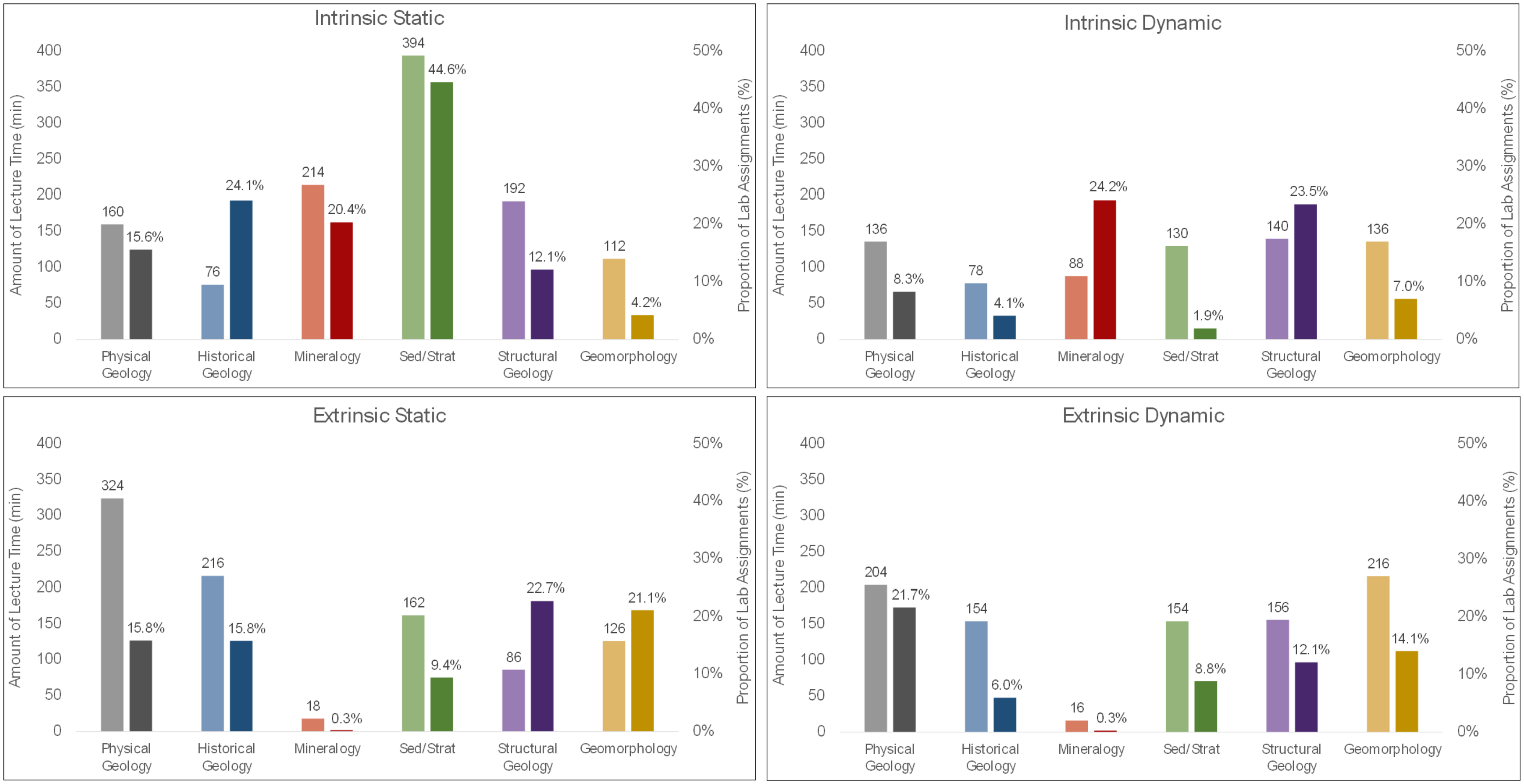
In contrast, the intrinsic dynamic category represents the least amount of overall lecture time (~9.6 hours) and averages about 11.5% of lab assignments (Figure 3-3). Mineralogy and Structural Geology boast the highest proportion of the intrinsic dynamic category (Figure 3-3). Overall, Historical Geology lecture and lab and the Sedimentology/Stratigraphy lab are the least likely settings to feature intrinsic spatial thinking skills (Table 3-3).

In contrast, almost all courses with the exception of Mineralogy feature aspects of the extrinsic domains (Figure 3-3). Extrinsic static is the second most featured category with more than 15 hours of lecture instruction and averaging ~14% of lab activities (Figure 3-3). Extrinsic static skills are present in approximately 5.5 hours of Physical Geology classes, equivalent to over two weeks of lecture material (Figure 3-3). They also appear in ~3.5 hours of Historical Geology lectures and Structural Geology labs and Geomorphology assignments contain approximately 20% of extrinsic static content.

The extrinsic dynamic category is present in lecture content just a little less than extrinsic static category with a total of approximately 15 hours (Figure 3-3).

Figure 3-3

Spatial thinking exposure in lecture and lab by spatial categories



Note: Results showing how time spent on spatial thinking was distributed among spatial thinking categories either through instruction or practice. The light-colored bars for each course represent spatial thinking exposure in lecture. The dark-colored bars represent spatial thinking exposure in labs. As a reminder, the lab data for Geomorphology represents course homework assignments.

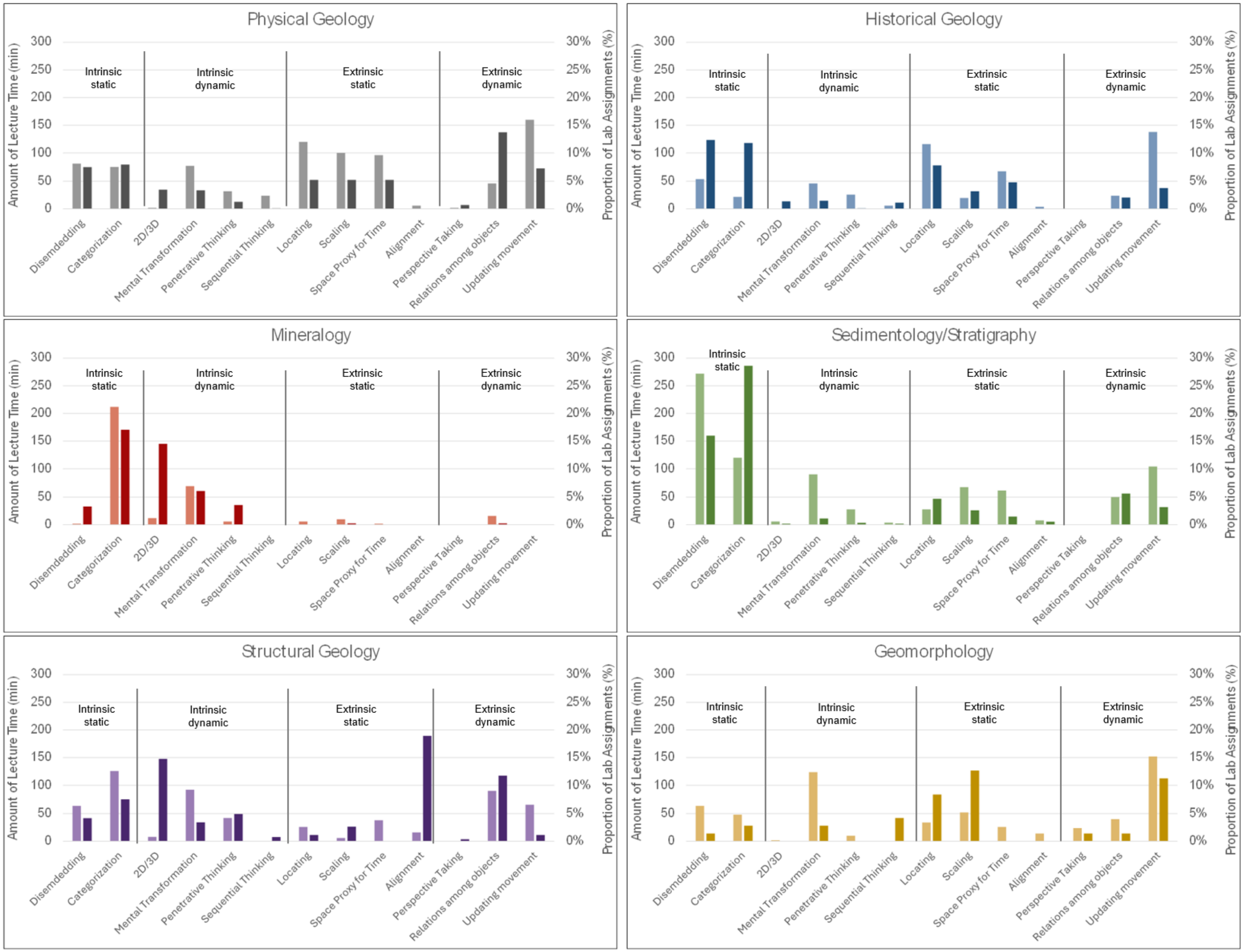
At least 2 hours of extrinsic dynamic skills are presented in every class except Mineralogy (Figure 3-3); representing a little over a week's worth of instruction in Physical Geology, Historical Geology, Sedimentology/Stratigraphy, Structural Geology, and Geomorphology. This category represents the least proportion of lab assignments among the various spatial thinking tasks, averaging a little over 10% of the semester.

As noted above, intrinsic static skills are most frequently observed in Mineralogy and Sedimentology/Stratigraphy; Figure 3-3. The individual skills in this category (categorization, disembedding) are the dominant skills present in the Sedimentology/Stratigraphy lecture and lab (Figure 3-4). Disembedding was present in nearly 4.5 hours of lecture time and over 15% of the lab activities and categorization represented over 25% of lab activities and featured in approximately 2 hours of lectures. While disembedding was barely emphasized in Mineralogy, categorization content made up nearly 2.5 hours of lecture time and over 15% of the lab activities.

Not all intrinsic dynamic skills were featured equally (Figure 3-4). For example, mental transformation was the most frequently observed intrinsic dynamic skill in course lectures ranging from approximately 45 minutes in Historical Geology to 2 hours in Geomorphology, with most lectures containing approximately 1 hour of this skill. . Mineralogy and Structural Geology labs spent approximately 15% of their time on the 2D/3D skill, while the lectures for these courses contain less than 15 minutes of this skill.. Penetrative thinking and sequential thinking were less frequently observed across all six courses (Figure 3-4).

Figure 3-4

Spatial skill exposure in lecture and lab by course



Note. Results showing the relative proportions of time students were exposed to each spatial thinking skill. The light-colored bars for each course represent lecture; the dark-colored bars represent labs. As a reminder, the lab data for Geomorphology represents course homework assignments.

In the extrinsic static category, several skills (e.g., locating, scaling, space as time proxy) appear relatively frequently in Physical Geology, Historical Geology, and Geomorphology. All three skills appear in similar proportions in Physical Geology lectures (approximately 1.5 – 2 hours) and labs activities (5%). Locating occurs for approximately 2 hours in Historical Geology lectures and scaling makes up 10%-15% of lab activities. Locating and space proxy for time also occur in fair amounts in Geomorphology and Historical Geology, respectively. One extrinsic dynamic skill, alignment, dominates this category for Structural Geology lab activities, (~ 20% of lab time) but barely features in the other courses (Figure 3-4)..

Updating movement and relations among objects were the skills most likely to feature in the extrinsic dynamic category (Figure 3-4). The updating movement skill was recognized frequently (~2.5 hours) in the lecture discussions of Physical Geology, Historical Geology, and Geomorphology. The only practicum that contained a large proportion of the updating movement skill was the Geomorphology homework assignments (Figure 3-4). Relations among objects occurred largely in Physical Geology lab (~15%) and Structural Geology lecture (~1.5 hours) and lab (~ 10%). One skill from this category, perspective taking, was rarely observed in any of the courses (Figure 3-4).

DISCUSSION

RQ1: How do undergraduate geology instructors incorporate references to spatial thinking and the use of spatial thinking skills in their instruction and course materials?

We attribute the observed patterns of spatial categories and skills to course content. For example, Mineralogy focuses on small-scale, individual objects (e.g., crystals, minerals) and their characteristics, thus it was anticipated that the most frequently observed skills would fall into the intrinsic (within object) domain. Whereas Geomorphology places more emphasis on large-scale landscape characteristics and transformations, thus we expected a larger proportion of our observations to be within the extrinsic (between objects) category. The results shown in Figure 3-3 support this interpretation. Some courses show a greater emphasis on intrinsic skills (e.g., Mineralogy, Sedimentology/Stratigraphy), in others, extrinsic skills are dominant (e.g., Physical Geology, Geomorphology), while Historical Geology and Structural Geology show a mix of skills across most categories (Figures 3-3, 3-4). While similar skills are represented in the lectures and labs, we also see some differences that are attributable to the contrasting formats of these settings. The following section discusses these variations by spatial category which are also summarized in Figure 3-5.

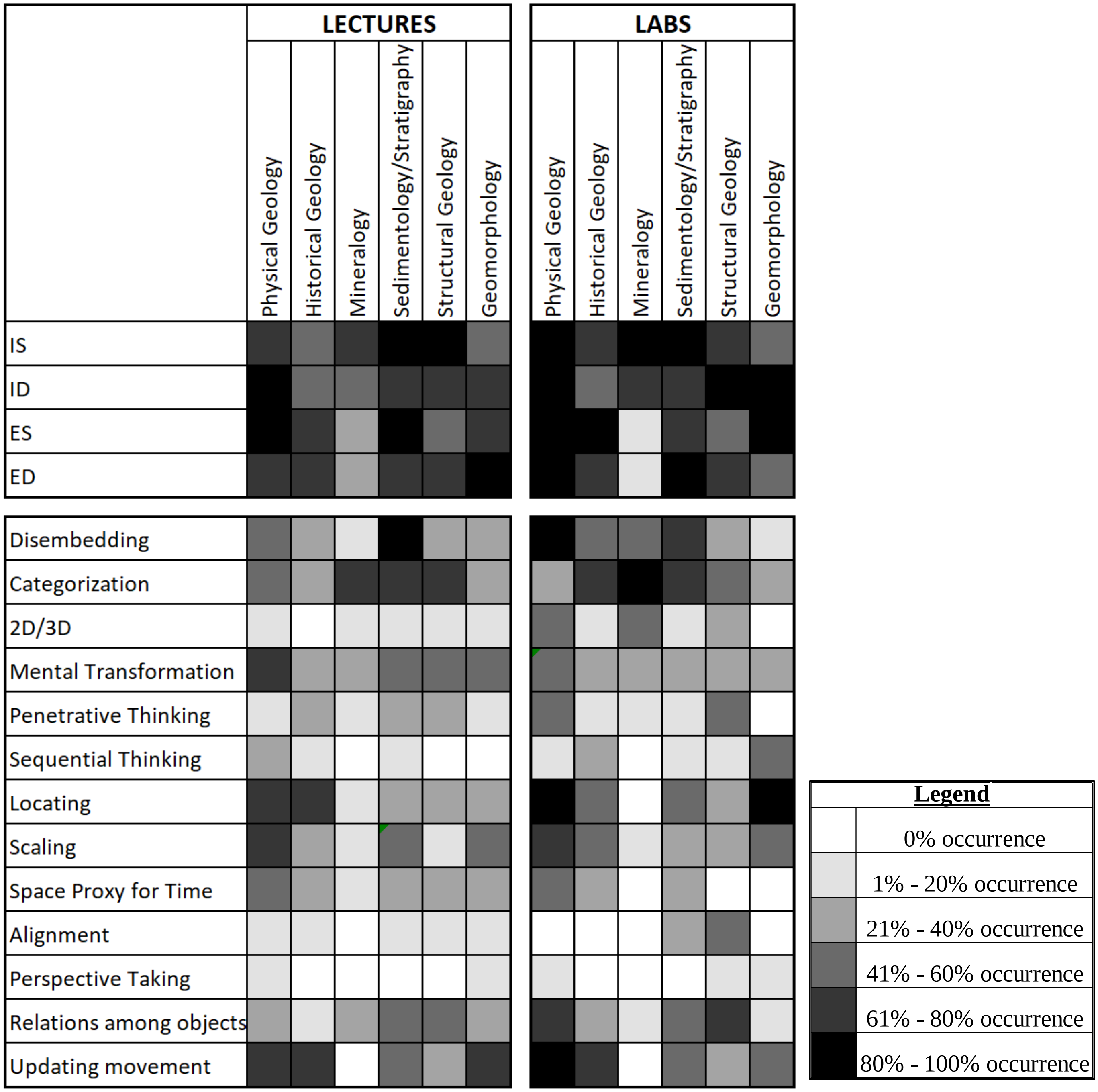
Intrinsic Static

We observed different patterns in how intrinsic static content was presented across several of the courses. Some courses (Physical Geology, Mineralogy) had essentially similar proportions of class time dedicated in lab and lectures for one or both of categorization and disembedding skills. For example, in Mineralogy lecture, students applied categorization skill identifying crystal systems, mineral habits, and cleavage type, and the same skill were also featured in several of labs which focused on crystallography and mineral properties (Figure 3-4)..

Elsewhere, either the lecture or the lab was the principal source of skill development for categorization or disembedding. The Historical Geology lecture contained a few examples of these skills but they were more prevalent in the labs as students spent several lab periods identifying rocks, sedimentary structures, and fossils (Figure 3-4). Intrinsic static skills were the most frequently observed category in Sedimentology/Stratigraphy in both the lecture and lab.

Figure 3-5

Summary grid chart showing frequency of spatial category/skill in lecture and lab



Note. This grid chart shows the relative occurrence of spatial categories and skills in the six courses. White-colored boxes represent categories/skills that were present in 0% of the lecture sessions/lab assignments and black boxes represent that the category/skill was present in all (100%) of the lecture sessions/lab assignments, with higher saturation grays being more frequently occurring and lighter-colored grays less frequently occurring, Disembedding was the dominant spatial skill that was incorporated into the lecture materials , as there were several days dedicated to stratigraphic correlation. The lab also included this skill in activities about making observations of sedimentary rocks and thin sections. Categorization was most significantly represented in the Sedimentology/Stratigraphy lab as students examined thin sections of sandstones and carbonates using a petrographic microscope to identify and classify sedimentary rock features such as grain shape, sorting, and cementation.

Similarly, the Structural Geology lecture focused on the identification of geologic structures and interpreting their features (categorization and disembedding, respectively; Figure 3-4). The labs of these courses featured these skills as well in similar activities but with fewer instances.

Intrinsic Dynamic

Mineralogy and Structural Geology had the highest proportions of time spent overall using intrinsic dynamic skills, with the lab having a higher proportion of related activities than the lecture in both cases (Figures 3-3, 3-4). These lecture/lab differences mostly arose from the practice of using 2D/3D skills in lab (Figure 3-4). There were several activities in Mineralogy and Structural Geology, where 2D stereonet diagrams were used to represent 3D geologic features (e.g., crystal faces, bedding planes, faults, etc.). These tasks occurred almost exclusively in lab settings. Mental transformations were practiced in Mineralogy lecture when students visualized the effects of crystal symmetry operations (i.e., rotation, reflection) on an iPhone. Mental transformation was also present in several Mineralogy labs where students were asked to visualize crystal symmetry operations of several wooden blocks.

Intrinsic dynamic skills were present in other courses but represented a smaller proportion of spatial thinking tasks. For example, sequential thinking was demonstrated in

Physical Geology where students were shown pictures of four outcrops and asked to determine which two were most similar. Students must disentangle and remember the sequence of events the outcrops underwent to compare their similarities. In Geomorphology, students were shown a sequence of DEM photos to explain how landslide morphology changes over time (e.g., mental transformation). Penetrative thinking represented less than 10% of the spatial thinking present in all the courses examined (Figure 3-4). Cross-sectioning is a common practice among geologists and has been assessed in many spatial thinking studies in the geosciences (Ormand et al., 2014; 2017; Gold et al., 2018). We anticipated that penetrative thinking would be frequently taught and practiced in undergraduate geoscience courses and were surprised by its relative absence in the curriculum. Physical Geology and Structural Geology offered students some opportunities to draw and annotate cross-sections of plate boundaries and geologic maps, respectively.

Extrinsic Static

The extrinsic static category was prevalent in the Physical Geology and Historical Geology lectures and labs. The lectures of both courses featured instruction on locating skills with students identifying the past and present locations of continents, ocean basins, plate boundaries, and other geologic features. Relative dating principles, an example of space as a proxy for time, were taught and practiced in both courses as well. Space as a proxy for time was also shown, in Physical Geology, when discussing hot spots, students answered clicker questions to determine the direction of plate motion given the ages of volcanic islands. Scaling was present in Physical Geology lectures in activities such as ranking the relative sizes of various geologic objects (e.g., types of volcanoes, sediment grains). These skills were also featured in lab assignments, but to a lesser extent as more time was devoted to student practice in other spatial categories.

Geomorphology assignments and Structural Geology labs also emphasized specific extrinsic static skills. Five out of the six course assignments in Geomorphology required students to locate geologic features on maps or in Google Earth and several assignments involved students measuring the size of objects, such as cinder cone volcanoes and hillslope gradients. In Structural Geology, one lab was devoted to the alignment skill with students measuring the strike and dip of geologic structures using a Brunton compass throughout the assignment and in subsequent labs. These skills were also present in the lecture in both courses but were emphasized much less. Extrinsic static skills rarely appeared in Mineralogy but were featured on occasion in Sedimentology/Stratigraphy, especially in relation to differentiating between sedimentary clast sizes and judging the relative proportions of matrix to clasts (i.e., scaling).

Extrinsic Dynamic

Several skills from the extrinsic dynamic category were prominently featured in Physical Geology (Figure 3-4). In the lecture, there was discussion of processes, such as plate tectonic, earthquakes, and groundwater migration, that involved updating movement as students visualized how the different components moved in relation to one another. For example, students may envision how continental and oceanic plates interact at a convergent plate boundary. Along with these topics, students spent several lab sessions interpreting and using various types of maps. For example, in one Physical Geology lab activity, students used a topographic map and groundwater data to determine if a lake was natural or artificial by comparing the relative elevations of the land and groundwater surfaces (i.e., relations among objects). Despite its infrequent occurrence, the Physical Geology lab uniquely provided students with deliberate practice in perspective taking (Figure 3-4).

Sedimentology/Stratigraphy, Structural Geology, and Geomorphology lectures similarly discussed large-scale processes, such as sediment transport, global tectonics, and landscape evolution, respectively. The relations among objects and updating movement skills were embodied by these processes in labs as well but were less emphasized than skills from other spatial categories. In Structural Geology, the instructor described the directional relationship between normal and shear stress using a diagram to illustrate how compressional forces act on a geologic object (e.g., bedding; relations among objects). Updating movement was represented by lectures in a Sedimentology/Stratigraphy and Geomorphology on sediment transport involving the movement and redistribution of sediment.

RQ2: How does spatial thinking content vary across the undergraduate geology curriculum?

In addition to characterizing spatial thinking within courses, we were also interested in describing the prevalence of spatial thinking categories and skills across the curriculum. To accomplish this task, we tabulated the number of occurrences of each category and skill in each lecture session/lab assignment for each course. Conditional formatting was then applied to this tabulation using Excel, creating a grid representing the prevalence of spatial categories and skills across each semester (Figure 3-5). Grid squares were colored white if the category/skill was not present and each progressively darker colors indicated a 20% increase in occurrence, with black cells representing a skill occurred in 80% to 100% of the class sessions or lab activities. In the summary grid, white or lightly shaded squares indicate the category or skill never or rarely appeared the lectures or labs for that course, whereas darker colors indicate that the category or skill was consistently featured .

Some skills from all four categories were consistently represented in five of our courses. The exception was mineralogy which lacked skills from the extrinsic categories.

Among individual skills, disembedding and categorization were generally well-represented across the curriculum. Mental transformation, locating, scaling, relations among objects, and updating movement were also commonly observed skills across the semester. However, four skills - penetrative thinking, sequential thinking, alignment, and perspective taking — appeared less frequently in both lecture and lab settings. The distribution of skills between lectures and labs is relatively consistent with the exception of the 2D/3D skill, which was nearly absent in the lecture portion of the course, but moderately featured in lab assignments. Hypotheses and suggestions for addressing these gaps will be discussed in the Limitations section below.

Summary

Using the data presented in Figures 3-4 and 3-5, we can identify which skills are most typical for each of the courses represented in our study (Table 3-6). Eight of the thirteen spatial skills that were described in Table 3-2 are relatively common within the six courses we observed (Table 3-6). Locating was also fairly prevalent in several courses (Figure 3-4) but did not make the cutoff for inclusion in Table 3-6. However, the remaining four skills — penetrative thinking, sequential thinking, perspective taking, space as a proxy for time — are observed less often, with the first three particularly rare. Space as a proxy of time does feature in Physical Geology and Historical Geology courses but is rarely observed in later courses (Figures 3-4, 3-5). One concern we have regarding the lack of these skills in the curriculum, particularly penetrative thinking, is students' preparedness for encountering these skills at field camp where students may struggle with map reading and generating accurate cross-sections.

Table 3-6
Typical skills associated with courses

Course	Skills (listed in order of prevalence)
Physical Geology	Updating movement, *Relations among objects*, **Scaling**
Historical Geology	**Updating movement**, *Disembedding*, *Categorization*
Mineralogy	Categorization, *2D/3D*
Sedimentology/Stratigraphy	Disembedding, Categorization
Structural Geology	**Categorization**, *2D/3D*, *Relations among objects*, *Alignment*
Geomorphology	Updating movement, *Scaling*, **Mental Transformation**

Note. A skill was included in this list if the skill was present in 120 minutes or more in lecture sessions or more than 10% of the lab assignments. **Bolded skills** indicate the skill only met the lecture criteria for inclusion and *italicized skills* indicate the skills only met the lab criteria for inclusion.

Limitations

Content

One of the primary limitations of this study is the breadth of data collected. While using the STOP was relatively straightforward for lecture classes and lab assignments. It is much more difficult to apply to more student-centered activities, such as projects and field trips. In these types of activities, students have more autonomy to direct their learning, leading to more individualized experiences for students. While these activities were not included in our data collection efforts, we feel confident that the lecture and lab data are representative of the course content.

Further, there are a number of courses which we did not have the time to observe, and which may emphasize similar or different collections of spatial skills. While courses such as paleontology and petrology may echo many of the skills featured in Historical Geology and Mineralogy, respectively, they may also provide opportunities for under-emphasized skills (e.g., sequential thinking and space as a proxy for time in paleontology). Elsewhere, courses such as geophysics may enhance penetrative thinking skills and field courses may include activities that support the development of perspective taking skills.

Content may also vary for a course across institutions, depending on the preferences and strengths of the instructor. For example, the Mineralogy course in this study covered lecture topics, such as gemology and extraterrestrial mineralogy, which may not be universally taught compared to physical properties of minerals. We estimate that most of the content in study courses is similar across institutions, but there are likely some differences.

Procedural

One major difference between our observation protocol and the COPUS (Smith et al., 2013), was that the COPUS allowed for the selection of more than one activity for a given observation whereas our use of the STOP did not. For instance, during the two-minute period in which the observation was collected, the instructor may have been describing the processes occurring at an ocean-continent convergent plate boundary using a cross-section diagram. Using our selected procedure, this observation would have been coded as “the instruction of the extrinsic dynamic skill updating movement” as the principal focus was on the dynamic processes. However, if multiple codes had been permitted in our procedure, penetrative thinking would have also been identified as the description was accompanied by a cross-section through the plate boundary. We believe this procedural choice may have reduced the observed frequency of some skills in the lecture observations. Fortunately, using data from lectures and labs provided an additional line of evidence to capture overlooked data points.

We have been careful to differentiate between spatial thinking instruction, practice, and exposure (which includes both modes of content delivery). Research on best-practices in education asserts that students achieve higher learning gains from instructors employing active learning techniques compared to direct instruction (Freeman et al., 2014; Wieman, 2007). While students may have been present for instruction on spatial thinking in their courses, they may not have been engaged in spatial thinking themselves, which could result in unequal knowledge and skill acquisition.

CONCLUSIONS

Based on the data and discussion above, we reached the following conclusions:

1. A relatively typical sequence of geology courses include significant spatially intensive content and offer students numerous occasions to practice a range of spatial thinking skills.
2. Approximately a quarter to half of lecture time and half to three-quarters of lab time across six common geology courses features the discussion or application of spatial skills.
3. Intrinsic spatial skills are more likely to feature in Mineralogy and Sedimentology/Stratigraphy while extrinsic skills are more often observed in courses such as Physical Geology, Geomorphology and Structural Geology.
4. Some spatial thinking skills (penetrative thinking, sequential thinking, alignment, perspective taking) were infrequently observed or nearly absent in the undergraduate geology courses we analyzed.
5. The methodology described in this article may be appropriate to characterize other skillsets, if the skillset can be adequately delineated (i.e., typology).

Our results provide a baseline that other researchers may use to compare and contrast the characteristics of spatial thinking in geology courses. Overall, our analyses suggest that a student progressing through a geology curriculum would be likely to encounter examples of both intrinsic and extrinsic skills in their introductory courses (Physical Geology, Historical Geology). There would be a greater emphasis on intrinsic spatial skills during their second year in courses such as Mineralogy and Sedimentology/stratigraphy. During their third year, some of these skills would be reinforced and additional exposure to extrinsic skills in courses such as Structural Geology and Geomorphology would be expected. A student enrolled in the six courses we observed would be exposed to nearly 60 hours of spatial thinking focused lectures during their first three years in a geology program, equivalent to nearly two (~1.8) full lecture courses. A little over half (56%) of their lab activities are also focused on some aspect of spatial thinking. Consequently, perhaps it should not come as a surprise that previous research in spatial thinking has often shown that student spatial thinking skills steadily improve as they progress through the curriculum (e.g., Ormand et al., 2014, 2017; Gold et al, 2018).

Previous research in spatial thinking (e.g., Ormand et al., 2014; Gold et al, 2018) has emphasized some skills (Disembedding, Mental Transformation) that are clearly observed in multiple courses in our analysis. However, the skill that has received the most attention is penetrative thinking (e.g., Titus & Horseman, 2009; Ormand et al., 2014, 2017; Gold et al, 2018; Hannula, 2019) which, while present, is not a major feature of most courses. In contrast, there is little to no research on the development of several skills that are seen frequently seen in undergraduate geology courses (categorization, updating movement, locating, scaling, relations among objects) either due to lack of measurement instruments or perhaps a lack of recognition that these skills are widely present. Further, the apparent absence of some skills (penetrative thinking, sequential thinking, alignment, perspective taking) from the courses we observed may present an opportunity for other researchers to investigate if these skills are emphasized elsewhere in the undergraduate curriculum.

CHAPTER 4 – CHARACTERIZING SPATIAL THINKING INSTRUCTION IN UNDERGRADUATE GEOLOGY COURSES USING AN OBSERVATION PROTOCOL

Prepared for submission to *Journal of Geoscience Education*

INTRODUCTION

Spatial thinking represents a foundational set of skills woven throughout undergraduate geology curricula. For example, Physical Geology introduces students to the dynamic changes in the configurations of Earth's plates. In Mineralogy, students classify crystal systems and interpret mineral habits, honing their ability to discern subtle structural differences. Conversely, in Sedimentology/Stratigraphy and Structural Geology, the emphasis shifts to understanding processes on a broader scale through tasks like correlating stratigraphic layers and interpreting geological structures. Each course provides students with multiple opportunities for cultivating several different spatial thinking skills. Students may potentially be exposed to approximately a dozen different spatial thinking skills as they navigate their way through a typical geology curriculum (Newcombe & Shipley, 2015; Kastens & Ishikawa, 2006).

Researchers have understood the role spatial thinking plays in recruitment, achievement, and retention for STEM (science, technology, engineering, mathematics) disciplines since the early 2000s (Shea et al., 2001; Wai et al., 2009). Despite this awareness and demand for an increased workforce capacity in STEM fields, K-12 and higher education does not include explicit spatial thinking instruction or assessment (NRC, 2006). The absence of spatial thinking instruction persists into undergraduate education and may influence otherwise talented students to abandon or avoid spatially intensive majors (e.g., geosciences, chemistry, engineering; Shea et al., 2001).

A recent study (Sabatini & McConnell, in preparation) found that undergraduate geology students encounter spatial thinking content in nearly 30% of their core coursework (e.g., Physical Geology, Mineralogy, Structural Geology). Fortunately, spatial thinking is malleable and can be improved with targeted training (Uttal et al., 2013; Ormand et al., 2017; Gold et al., 2018). As various spatial thinking skills are incorporated into geology courses, instructors have the opportunity to provide training and practice in the application of these skills. We sought to understand how and when geology students develop various spatial thinking skills as they progress through the curriculum and if these skills were sustained. These findings could then help inform researchers and instructors in the creation of effective spatial thinking training and practice activities in ways that are synergistic to an instructors' established course content and teaching methods.

Background

Spatial thinking is a multi-faceted set of skills that people use to describe and visualize the spatial characteristics, relationships, and transformations of and between objects at a range of space and time scales (NRC, 2006). Experts in cognitive science and geoscience have developed several frameworks to identify geoscience-specific spatial skills (Kastens & Ishikawa, 2006; Manduca & Kastens, 2012; Newcombe & Shipley, 2015). We chose to utilize the typology developed by Newcombe and Shipley (2015) which consists of thirteen spatial skills experts deemed essential across the geoscience sub-disciplines. These spatial skills are mapped onto a framework developed by Chatterjee (2008) that represents how people think and reason about space along two dimensions: intrinsic/extrinsic and static/dynamic. The intrinsic/extrinsic dimension refers to the focus on either the characteristics of a singular object (intrinsic) or the relationship of multiple objects (extrinsic; Table 4-1).

Table 4-1
Spatial Skill Typology (Newcombe & Shipley, 2015)

Skill Category	Skill Name	Skill Description	Examples from real geology courses
Intrinsic Static	Disembedding	Isolating and focusing on a single aspect of a complex array	Correlating stratigraphic sections using lithology (Sedimentology/Stratigraphy)
	Categorization	Classifying objects or characteristics of objects by their spatial properties	Identifying crystal habit of mineral specimens (Mineralogy)
Intrinsic Dynamic	2D/3D	Visualizing the 3-D spatial relations of objects shown in 2-D and vice versa	Using a stereographic projection to visualize the orientation of joints (Structural Geology)
	Mental Transformation	Visualizing a change in an object's shape, size, or orientation	Envisioning the effect of symmetry operations using everyday items (Mineralogy)
	Penetrative Thinking	Visualizing the shape or spatial relations inside an object	Drawing a cross-sectional profile of a plate boundary (Physical Geology)
	Sequential Thinking	Visualizing and remembering a series of object transformations	Deducing the depositional/ erosional processes responsible for dune formation (Sedimentology/ Stratigraphy)
Extrinsic Static	Locating	Identifying the past or present location of objects on maps or in space	Using a world map to identify the location of continental cratons (Historical Geology)
	Scaling	Reasoning about the size of objects and their properties	Comparing the scaling of remote sensing imagery (Geomorphology)
	Space Proxy for Time	Reasoning about how spatial relations correspond to temporal relations	Recognizing the order of events in a block diagram using relative time principles (Historical Geology)
	Alignment	Perceiving horizontal from a non-horizontal reference frame	Measuring the strike and dip of a bedding plane using a Brunton compass (Structural Geology)
Extrinsic Static	Perspective Taking	Visualizing the appearance of an area from different points of view	Evaluating whether two map points are visible to each other in the field (Physical Geology)
	Relations among objects	Visualizing the spatial relations between multiple objects on maps or in space	Understanding the relationship between rock fractures and principal stress directions (Structural Geology)
	Updating movement	Visualizing the movement of object(s) across space and time	Imagining the migration of an inlet due to longshore currents (Geomorphology)

Note. Newcombe and Shipley's (2015) typology outlined eleven skills, with "Alignment" calling out two significant cases. We have subdivided "Alignment" into three distinct skills for clarity: Alignment, Scaling, and Space as a Proxy for Time.

The static/dynamic dichotomy emphasizes the contrast in mental processes, with static tasks involving visualization and interpretation, and dynamic tasks encompassing manipulation and transformation of the characteristics/relationships of object(s) (Table 4-1). The two dimensions result in four categories for the 13 different skills (Newcombe & Shipley, 2015; Table 4-1). For example, the intrinsic static category includes the skill of categorization which a student would use to interpret the features of minerals to identify a rock sample. In contrast, the extrinsic dynamic category includes the skill of perspective taking which allows a student to look at a topographic map and interpret how a landform would appear from different locations in the field. Table 4-1 provides additional information and examples or the application for each of the skills.

Undergraduate students entering the field of geology possess diverse spatial thinking abilities (Kali & Orion, 1996; Ormand et al., 2014). Proficiency in one type of spatial skill does not necessarily translate to success in others, even within the same category (e.g., mental rotation vs. penetrative thinking; Ormand et al., 2014). Efforts to assess and/or improve geoscience spatial skills have focused most heavily on skills in the intrinsic dynamic category (Table 4-1; e.g., Titus & Horsman, 2009; Ormand et al., 2014, 2017; Gold et al., 2018). Measuring spatial thinking skills has typically been done using a pre/post-test that is a composite of two, three or four tools or instruments that represent only the intrinsic dynamic category (Titus & Horsman, 2009; Giorgis, 2015; Bagher et al., 2020), both the intrinsic static and intrinsic dynamic domains (Ormand et al., 2014; Gold et al., 2018; Kreager et al., 2022), the intrinsic dynamic and extrinsic static categories (Ormand et al., 2017; Hannula, 2019), or the extrinsic static and extrinsic dynamic categories (Polifka et al., 2022). At the time this investigation was conducted, there were no studies which examined skills in all four categories. Additionally, most of these investigations assessed students' baseline spatial thinking skills in one (Polifka et al, 2013), two (Ormand et al., 2014; 2017; Gold et al., 2018), or three courses (Titus & Horsman, 2009) at the same institution.

These past investigations have shown that students enrolled in upper-level geology courses, such as Structural Geology, tend to outperform those in introductory courses (Titus & Horsman, 2009; Ormand et al., 2014). The degree of improvement may vary between similar courses taught at different institutions and may stem from differences in course content, instructional approaches, and the components of the assessment instrument used in each study.

In summary, no study has utilized a composite test with skills from more than two spatial skill categories and in more than three courses at the same institution. While this provides insight on certain skills in some contexts, it does not impart any information on spatial thinking as a whole within a typical geology curriculum at an institution. A critical influence not discussed in prior studies is that of the instructor and the course content. Titus and Horseman (2009) hinted that geological tasks in advanced courses may require students to use spatial thinking more frequently than introductory students, but did not provide an explanation on this relationship. It is likely the that course content and how instructors deliever that content (passive lecture vs. active learning) plays a significant role in students' spatial thinking development and how their skills improve over time.

Research Questions

The purpose of this mixed methods study was to determine how spatial thinking skills are represented and developed throughout a traditional undergraduate geology curriculum. Previous research has predominantly centered on exploring intrinsic dynamic spatial skills, particularly within introductory and structural geology courses. This emphasis has been mirrored in the interventions that have been tested and examined within these contexts. Our study examines spatial skills across all four spatial categories from the Newcombe and Shipley (2015) typology, encompassing several required courses in the geology curriculum and one popular elective. Moreover, we examined how instruction may be promoting students' development of spatial thinking. The following research questions guided our investigation.

1. How are spatial thinking skills developed by students at multiple points in a traditional undergraduate geology curriculum? (QUAN)

2. To what extent do undergraduate students' scores on spatial thinking skills tests converge with their instructors' integration of spatial thinking skills in their geology courses? (QUAN + QUAL)

We have provided a summary of the methods and findings of the qualitative strand for Research Question #2 in the next section to provide context for the mixed methods results discussed here. A full explanation of the methods and an interpretation of the qualitative results have been described in Sabatini and McConnell (in preparation).

Summary of Qualitative Methods and Findings

We developed the Spatial Thinking Observation Protocol (STOP) for the qualitative portion of our mixed methods study to identify spatial thinking content embedded within the lectures and labs of six courses. This qualitative protocol was based on Newcombe and Shipley's (2015) typology of spatial skills and modeled after the Classroom Observation Protocol for Undergraduate STEM **(**COPUS; Smith et al., 2013) and Geoscience Teamwork Observation (GTO) protocol (Nyarko, 2021). The first author attended all lecture sessions of the selected courses across multiple semesters, recording spatial thinking content using a set of predefined codes indicating the presence of spatial thinking, mode of content delivery (instruction vs. practice), the skill category (e.g., intrinsic static), and the spatial skill (e.g., disembedding). Data collection occurred every two minutes throughout each lecture session, resulting in over 5,000 observations. Lab assignments were coded using a similar protocol. Data analysis involved compiling and quantifying the coded content for statistical analysis.

The results of the study (Sabatini & McConnell, in preparation) shed light on the incorporation of spatial thinking content and skills across undergraduate geology courses, revealing variations in the emphasis on different spatial categories and skills. While some courses emphasized intrinsic skills, like Mineralogy focusing on individual object identification, others, such as Physical Geology, emphasized extrinsic skills for understanding larger-scale spatial relationships. Certain skills, like mental transformation and locating, were frequently addressed whereas others, such as sequential thinking, were rarely observed. We concluded that undergraduate geology courses offer significant opportunities for students to engage with spatial thinking.

METHODS

Study Design

This research study employed an embedded convergent mixed methods design (QUAN + QUAL), in which qualitative and quantitative data are collected in parallel, analyzed separately, and then merged (Creswell & Plano-Clark, 2018). Both datasets are of equal importance. This type of design is needed to both accurately capture and describe students' understanding and instructors' incorporation of spatial thinking in undergraduate geology courses. This manuscript will mainly focus on the quantitative results from this project and provide interpretations about the convergence of the mixed methods in the Discussion section. Please see Sabatini and

McConnell (in preparation) for complete description and discussion of the qualitative portion of this mixed methods research study.

Study Setting and Participants

This study took place at a public research university located in the southeastern U.S. and specifically within several required undergraduate geology courses and one popular elective course. While the location of this study was chosen out of convenience to the researchers, the courses were purposefully chosen since they are commonly required courses in most undergraduate geology curricula (Klyce & Ryker, 2023). Students typically take these courses in sequence, beginning with Physical Geology and Historical Geology as freshmen, moving into Mineralogy and Sed/Strat as sophomores, and ending with Structural Geology and Geomorphology (sometimes taken as a sophomore) as juniors. Undergraduate students enrolled in these courses were recruited to participate in this research.

Over 100 students and six instructors were invited to participate in the study. Physical Geology and Historical Geology students were recruited from in-person lab sections taught by graduate teaching assistants. Each section had 10-20 enrolled students most of whom were not geology majors. Students enrolled in the Physical Geology lab sections were typically also enrolled in the lecture portion of the course, which was taught by two instructors. The remaining courses had enrollments between 10-20 students and each course was primarily taught by the instructor of record, with a teaching assistant aiding in the course's lab component (if applicable). Data was collected over two semesters for several of these courses (Table 4-2). All students were recruited to participate in the quantitative strand of the study. We sought to recruit a diverse sample with respect to gender, race, major (introductory courses only), and ability.

Table 4-2 provides the number of participants and when quantitative and qualitative data were collected.

Table 4-2
Qualitative and quantitative data collection summary

Course	QUALITATIVE			QUANTITATIVE	
	# of Lecture Sessions	# of Lab Assignments	Data Collected	Sample Size	Data Collected
Physical Geology	24	11	Fall 2022	60	Fall 2021
Historical Geology	38	13	Spring 2022	27	Spring 2022
Mineralogy	18	10	Fall 2021	25	Fall 2021 Fall 2022
Sedimentology/ Stratigraphy	26	10	Spring 2022	25	Spring 2022 Spring 2023
Structural Geology	33	9	Fall 2021	24	Fall 2021 Fall 2022
Geomorphology	39	6*	Spring 2022	23	Spring 2022 Spring 2023

Note. *Homework assignments used in lieu of lab assignments.

Data Collection

Instrumentation

Several instruments were used in this study. Students completed a spatial thinking aggregate pre- and post-test and demographic survey. The aggregate test had seven subscales which correspond to seven different spatial thinking skills: disembedding, visualizing 2D ↔ 3D, penetrative thinking, scaling, alignment, perspective taking, and updating movement. Course observation data was collected for the qualitative portion of the study.

Spatial Thinking Aggregate Test (STAT). We designed the Spatial Thinking Aggregate Test (STAT) to encompass seven spatial thinking skills with at least one skill from each spatial thinking category (Table 4-3). Each subscale contained questions from previously established spatial thinking instruments, resulting in a 31-question test which can be completed in 20 minutes. Table 4-3 details the composition and timing of the STAT. The STAT-Pre was used to assess students' baseline spatial thinking skills prior to engaging in their enrolled geology course's content. To determine if students spatial thinking skills improved during their geology course, the STAT-Post was administered at the end of the semester. The post-test was structured in the same way as the pre-test, with different questions, but similar in difficulty for most of the subsections (Table 4-3). A copy of the STAT-Pre can be found in Appendix A.

Sample test items from each subsection of the STAT are provided in Figure 4-1. Below, we will outline the instructions for completing each question type for each STAT subsection. A Hidden Figures (Figure 4-1a; Ekstrom et al. (1976) test item requires students to select the correct shape that is obscured within an image. In the Surface Development test (Figure 4-1b; Ekstrom et al., 1976) students must match the numbered edges of a 2-D shape to the lettered edges of the corresponding 3-D shape. For the Geologic Block Cross-section Test (GBCT, Figure 4-1c; Ormand et al., 2014), students must select the correct geologic cross-section indicated by a slice through a block diagram. The Zoom Assessment (Figure 4-1d; Jones et al., 2010) involves students making judgements about the size of magnified items via multiple-choice questions. The Water Level Test (Figure 4-1e; Piaget & Inhelder, 1967) requires students to draw a line to indicate the water level in a half-full tilted bottle. For the Spatial Orientation Test, students are shown an array of objects and asked to imagine they are standing at one object and facing another object, then draw a line pointing in the direction of a third object (Figure 4-1f;

Kozhevnikov & Hegarty, 2001). Finally, in the Topographic Map Assessment (TMA) items (Figure 4-1g; Jacovina et al., 2014) ask students to identify a specified hiking path or river on a topographic map and justify their choice (i.e., drawing the least strenuous path between two points).

Table 4-3

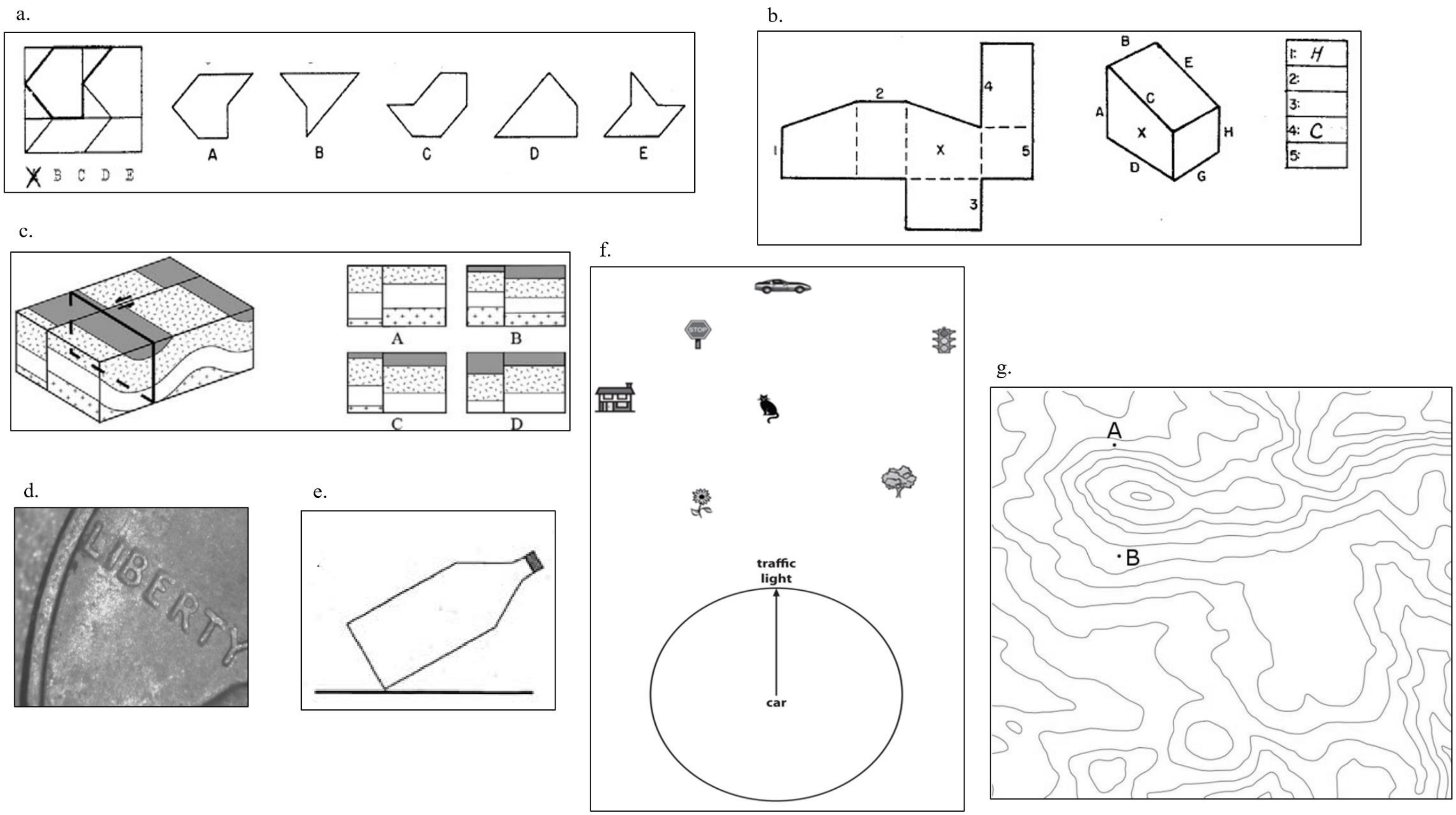
Spatial Thinking Aggregate Test

Spatial Test	Typology Category	Spatial Skill	# of Questions	Time Limit (min)	Question Type	Questions Same on Pre- and Post-Test
Hidden Figures[1]	Intrinsic Static	Disembedding	6	4	Multiple-choice	No
Surface Development[1]	Intrinsic Dynamic	2D ↔ 3D	3	3	Complete figure	No
GBCT[2]		Penetrative Thinking	6	3	Multiple-choice	No
Zoom Assessment[3]	Extrinsic Static	Scaling	6	3	Multiple-choice	Yes
Water Level Test[4]		Alignment	3	1	Complete figure	Slightly modified
Spatial Orientation Test[5]	Extrinsic Dynamic	Perspective Taking	4	2	Complete figure	No
Topographic Map Assessment[6]		Updating Movement	3	3	Complete figure & Short answer	Slightly modified

Note. [1]Ekstrom et al., 1976, [2]Ormand et al., 2014;2017, [3]Jones et al., 2010, [4]Piaget & Inhelder, 1967, [5]Kozhevnikov & Hegarty, 2001, [6]Jacovina et al., 2014. Note : A second set of questions for the Zoom Assessment was not available.

Demographic Questionnaire. The demographic questions included personal and school-related characteristics, such as age, gender, race/ethnicity, college major, number of enrolled semesters, and current or previously taken geology courses.

Figure 4-1
Sample STAT test items



Note. (a) Hidden Figures, (b) Surface Development Test, (c) Geologic Block Cross-Sectioning Test, (d) Zoom Assessment, (e) Water Level Test, (f) Spatial Orientation Test, (g) Topographic Map Assessment

Procedures

This study was approved by our Institutional Review Board (IRB #24198). Student participants were recruited during their first in-person lab session, usually within the first two weeks of the semester. A scripted recruitment statement detailed the purpose of the research project and notified students of the benefits and risks involved in participating in the research. After being read the script, students received a consent form where they indicated if they would like to participate in the research study.

After collecting the consent forms, all students immediately received and completed the STAT-Pre. The lead researcher administered the timed paper test by sub-section; students were told not move onto the next sub-section until directed. All students took the STAT regardless of consent status; this was because the tests were administered during class time as part of the course requirements, though they did not receive a grade for their participation. Students completed the Demographic Survey after the STAT-Pre to minimize stereotype threat (Steele & Aronson, 1995). At the end of the semester during the last lab period, students completed the STAT-Post in the same manner as the STAT-Pre. If students were enrolled in more than one participating course in a semester, only their first attempt of the STAT-Pre and/or STAT-Post was recorded. The STAT was administered over two iterations for Mineralogy, Sedimentology/Stratigraphy, Structural Geology, and Geomorphology in an attempt to attain an acceptable sample size for statistical significance.

Data Analysis

Data collection was followed by scoring the STAT-Pre and STAT-Post. Each question was worth one point, with partial credit allowed on the Surface Development Test, Water Level Test, Spatial Orientation Test, and Topographic Map Assessment. Multiple-choice questions

(Hidden Figures, GBCT, Zoom Assessment) earned one point for a correct response and no points for an incorrect choice. In the Surface Development test, students had to identify five corresponding edges for each question, with each match valued at 0.2 points. A student who correctly identified all five edge matches received one point, with 0.2 points deducted for each incorrectly matched edge pairing. In the Water Level Test, students were required to draw a horizontal line to indicate the water level. A deviation of within 5° from horizontal earned one point, while a deviation between 5° and 10° earned half a point, and deviations over 10° did not receive any points. Similarly, the Spatial Orientation Test was scored based on the magnitude of misalignment between the actual direction and students' answers. A deviation of up to 5° clockwise or counterclockwise earned one point, while a deviation between 5° and 15° earned half a point, and deviations over 15° resulted in no points. In the Topographic Map Assessment, students' answers included a drawn path and a justification. A full point was awarded if the justification matched the path and satisfied the prompt. If the justification did not match the path but the path satisfied the prompt, students received half credit. If neither criterion was met, no points were earned. Scores for individual questions, test-subsections, and skill categories were recorded and tabulated using Excel. Students' data was not included in the analysis if they did not provide consent. After the data was cleaned (i.e., missing/incomplete data issues addressed), the data was imported and analyzed using Stata 15.1. We began by conducting descriptive statistics (means, standard deviations, etc.) to describe the data set. Inferential statistics were also conducted, such as t-tests and multiple linear regressions.

To address Research Question #2, the findings from both data sources were integrated into a joint display and explained in narrative format to determine where convergence/divergence occurs. This allowed for themes and means to be compared across data sources.

Validity and Reliability

Each of the spatial thinking tests that make up STAT has previously established validity and reliability, making them suitable for use in our study. The validity and reliability of the Hidden Figures Test and Surface Development Test was established through a series of studies by the Educational Testing Service for use with sixth grade students (Ekstrom et al., 1976). The Water Level Test is a valid and reliable instrument that was developed by cognitive scientists (Ormand et al., 2017 after Piaget & Inhelder, 1967). Kozhevnikov and Hegarty's (2001) study support the validity of the Spatial Orientation Test as a test of perspective-taking tests by means of reorientation within an array, and with an internal reliability of 0.82. Zoom Assessment items were developed and validated by two science education researchers, an educational psychologist, and three biology instructors, and reliability was determined using a Kuder Richardson Formula 20 (KR20=.5392; Jones et al., 2011). Ormand et al. (2017) purport the GBCT has a high level of surface validity; meaning it does in fact test a person's ability to choose the correct slice through a 3D object. Furthermore, Newcombe et al. (2015) found that the reliability for the TMA was very high ($\alpha = .76$) and displays a wide range of performance according to item response theory analyses, which suggests that the TMA is useful for assessing topographic map reading ability and associated spatial skills. In general, most of these instruments are accepted and utilized by the spatial thinking research community.

The primary source of legitimation or inference quality for the mixed results of this study emerged from the triangulation between the STAT scores and observations. This triangulation of multiple data sources greatly enhances the validity and reliability of the integrated findings (Mathison, 1988). This triangulation was primarily established through a joint display and rich narrative-type descriptions.

RESULTS

Demographics

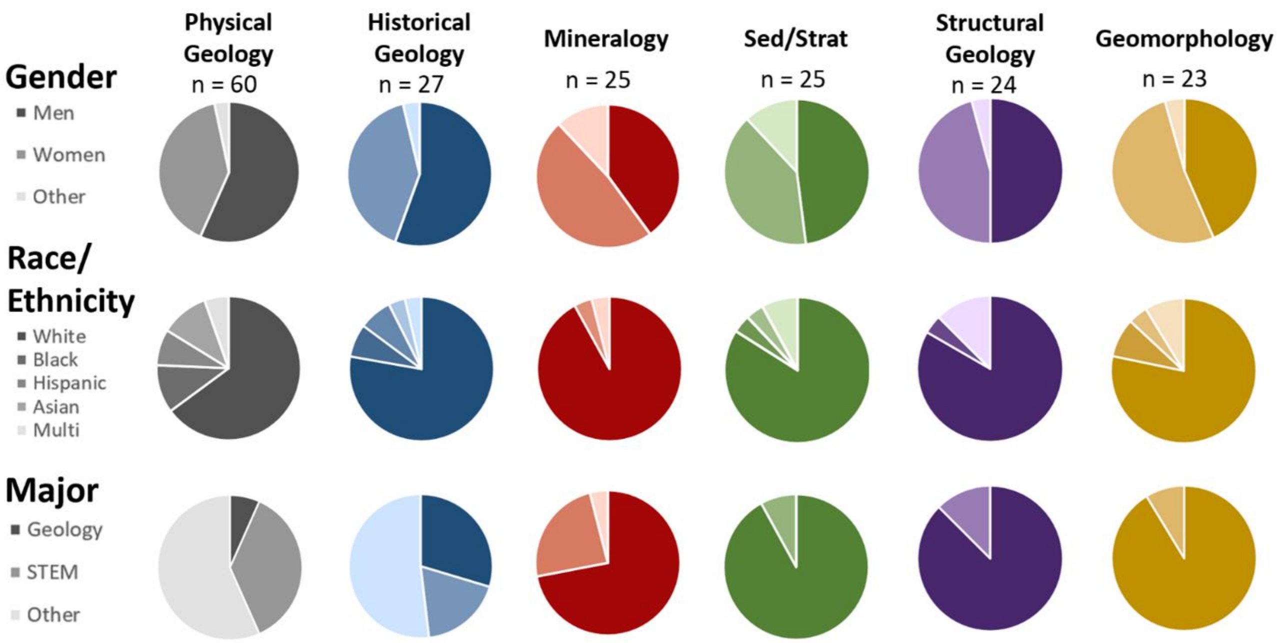
The number of participants and key demographic characteristics for each can be found in Figure 4-2. With the exception of Physical Geology, our sample sizes ranged from 23-27, a few students short of our desired sample size of 30, which is the recommended size for drawing statistical inferences (Hogg et al., 2015). Several patterns emerged from our demographic data, primarily centered around differences between introductory courses (e.g., Physical Geology and Historical Geology) and upper-level courses (e.g., Mineralogy and Structural Geology). Introductory courses were predominantly comprised of students who identified as male and were non-STEM majors (Figure 4-2). The gender ratio was closer to 50:50 (male vs. non-male) in the upper-level courses and most students were declared geology majors with a smaller proportion of other STEM majors (Figure 4-2). Students' race/ethnicity was also more varied in the introductory courses compared to the upper-level courses where over 75% of the students identified as white.

Students' STAT Scores and Measured Gains

Figure 4-3 presents several frequency plots showing the distribution of STAT scores for representative courses and the full sample set. Table 4-4 shows the STAT-Pre and STAT-Post score averages and gains (normalized as percentages for each of the classes in our study). A graphical visualization of STAT-Pre and STAT-Post scores is shown in Figure 4-4, with the relative proportion of each spatial category's contribution to the overall score represented in a stacked column.

Individual students' STAT scores ranged from 10%-20% in Physical Geology to 90%-100% in several upper-level courses with averages from ~40%-70% (Figure 4-4). The frequency plots (Figure 4-3) show a score shift between the STAT-Pre and STAT-Post for all courses (i.e., more students achieve higher scores on the STAT-Post than on the STAT-Pre), suggesting students' spatial thinking improved during the semester. Despite these gains, the average STAT-Post scores were still below 70% for students in upper-level geology courses.

Figure 4-2
Student demographic data for each course in the study

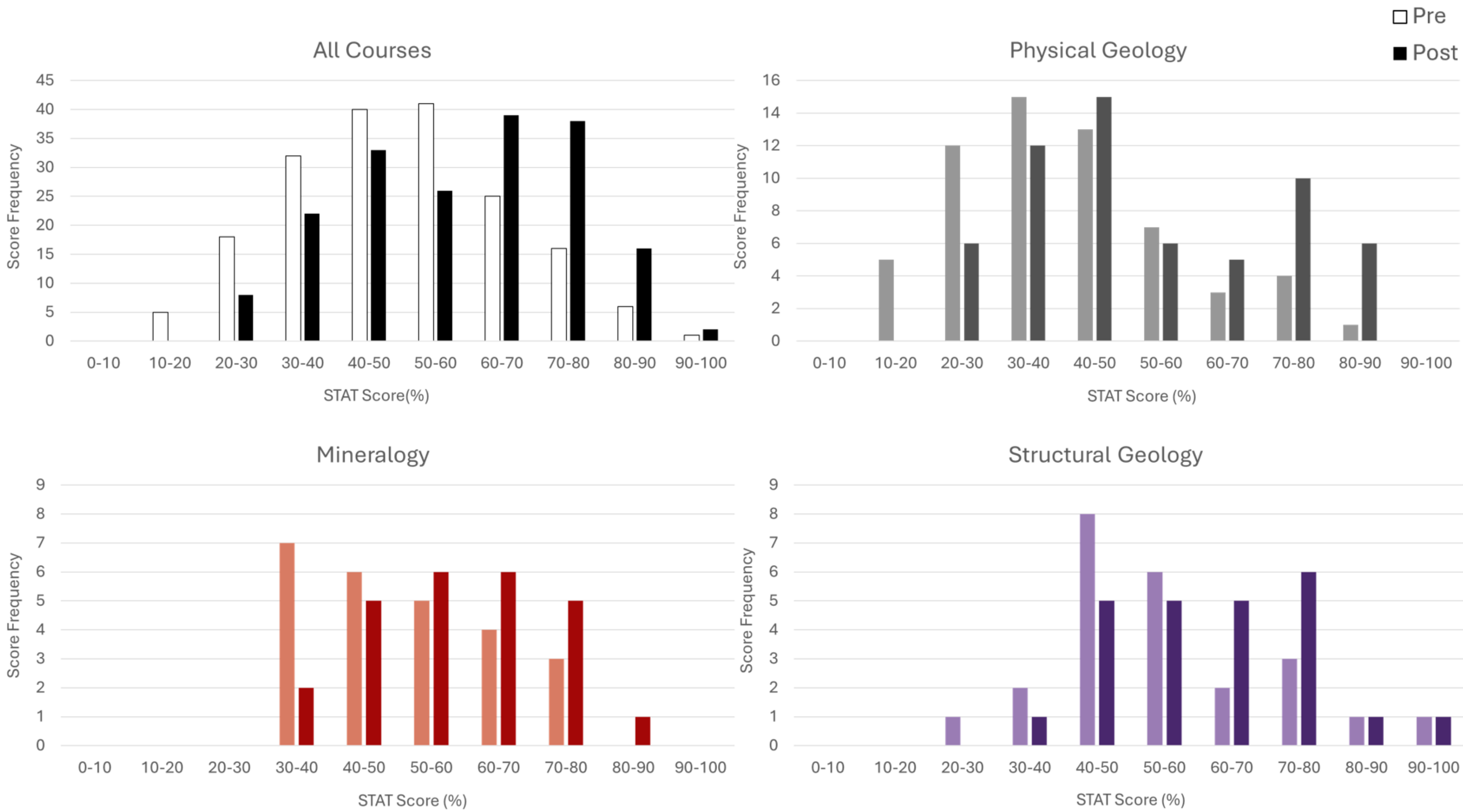


Note. The color gradient provided in the legend applies to all colors (i.e., for gender, the darkest color denotes men). Students' gender was classified as "Other" if they identified as transgender, non-binary, or other gender identity. Similarly, students' race was classified as "Multi" if they selected more than one race/ethnicity option.

Student gains on the aggregate test and at the spatial category and test-subsection level are shown in Table 4-4. While we cannot draw definitive inferences on the statistical significance of the measured gains, we denoted the level of significance (p-value) calculated using paired, two-tail t-tests in Table 4-4. Category scores and individual skill test scores also improved between the pre- and post-tests in most instances. Some notable exceptions include the Zoom Assessment scores in Mineralogy and the Water Level Test scores in Physical Geology and Historical Geology, however these losses were not statistically significant.

Figure 4-3

STAT score frequency distribution plots for entire dataset and select courses



Note. Light-colored bars refer to STAT-Pre scores and dark-colored bars to STAT-Post scores. The x-axis shows STAT score range bins in 10% increments and the y-axis plots the number of students who achieved a score within the specified bins. All four plots show a rightward shift between the STAT-Pre and STAT-Post scores.

Table 4-4
STAT scores and gains for aggregate test, spatial categories, and test-subsections

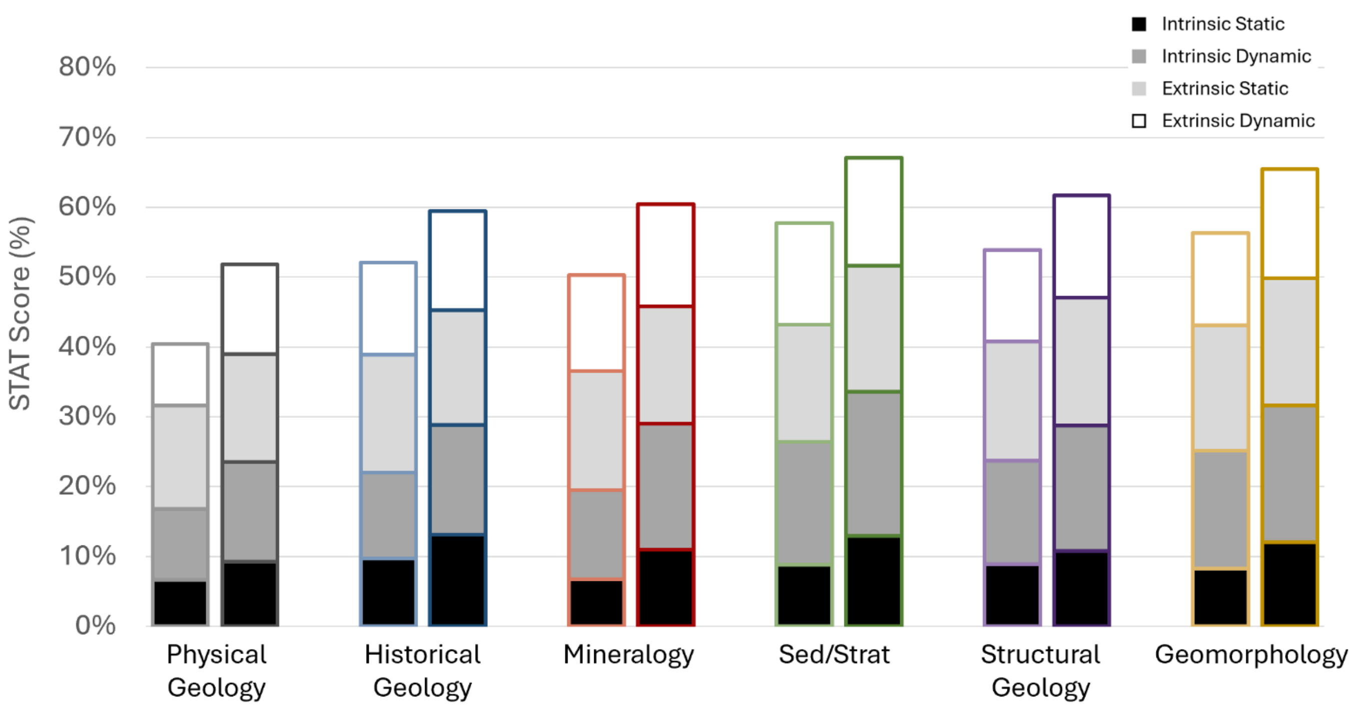
	Physical Geology	Historical Geology	Mineralogy	Sedimentology & Stratigraphy	Structural Geology	Geomorphology
Aggregate	40.44 51.78 11.34***	52.08 59.50 7.42***	50.28 60.46 10.18***	57.78 67.08 9.3***	53.88 61.73 7.85***	56.30 65.44 9.14***
Intrinsic Static/Hidden Figures	33.61 47.78 14.17***	50.00 67.90 17.90**	34.67 56.67 22.00**	45.33 66.67 21.33**	45.83 55.56 9.72*	42.75 62.32 19.57**
Intrinsic Dynamic	34.70 49.19 14.48***	42.55 53.91 11.36**	44.09 62.04 17.96***	60.80 71.29 10.49**	51.11 61.94 10.83**	58.16 67.44 9.28**
Surface Development Test	35.22 57.56 22.33***	49.88 62.96 13.09**	56.27 78.13 21.87***	73.07 87.20 14.13**	63.06 67.78 4.72	71.59 79.13 7.54*
Geologic Block Cross-Sectioning Test	34.44 45.00 10.56***	38.89 49.38 10.49**	38.00 54.00 16.00**	54.67 63.33 8.67*	45.14 59.03 13.89**	51.45 61.59 10.14*
Extrinsic Static	50.09 53.24 3.14	58.02 56.79 -1.23	58.67 58.00 -0.67	57.78 62.00 4.20	58.80 62.96 4.17	61.84 62.80 0.97
Zoom Assessment	40.83 48.33 7.50**	44.44 46.91 2.47	46.67 44.00 -2.67	44.00 49.30 5.33	47.92 52.78 4.86	48.55 50.00 1.45
Water Level Test	68.61 63.06 -5.56	85.19 76.54 -8.64	82.67 86.00 3.33	85.33 87.33 2.00	80.56 83.33 2.78	88.41 88.41 0.00
Extrinsic Dynamic	38.69 56.67 17.98***	58.47 62.96 4.50	60.86 64.86 4.00	64.57 68.57 4.00	58.04 65.18 7.14*	58.39 68.94 10.56**
Spatial Orientation Test	34.38 50.63 16.25***	50.93 58.33 7.40*	50.50 55.00 4.50	54.00 64.00 10.00*	45.83 55.73 9.90*	45.65 60.87 15.22**
Topographic Map Assessment	44.44 64.72 20.28***	68.52 69.14 0.62	74.67 78.00 3.33	78.67 74.67 -4.00	74.31 77.78 3.47	75.36 79.71 4.35

Note. Scores are normalized as percentages. Each cell contains the STAT-Pre score (upper left), STAT-Post score (upper right), and calculated gain or difference between the STAT-Post and STAT-Pre score (bottom). The intrinsic static category only contained one test-subsection (Hidden Figures) and thus represented the same scores. *p < 0.05, **p < 0.01, ***p<0.001

Gains were large and more frequently statistically significant for the intrinsic static and intrinsic dynamic test-subsections. Extrinsic static gains/losses were never statistically significant and extrinsic dynamic gains were significant for Topographic Map Assessment in Physical Geology and for the Spatial Orientation Test for all courses except Mineralogy. The difference in height of intrinsic sections of the stacked columns in Figure 4-4 strikingly illustrates the gains in the intrinsic categories.

Figure 4-4

Average STAT-Pre and STAT-Post scores



Note. Pre-test scores (left bar) are outlined in the lighter color and post-test scores (right bar) are outlined in the darker color. The colored sections of each bar represent how students' average spatial category scores contribute to the overall average STAT scores.

Test-Retest Analysis

In the Validity and Reliability section we mentioned the concern of the test-retest effect on students pre- and post-test scores. Unpaired two-tailed t-tests between the post-tests of one course and the pre-tests of the proceeding course in the curriculum (e.g., STAT-Post for Physical

Geology and STAT-Pre for Historical Geology) were used to evaluate this concern. We conducted this evaluation using four post/pre course pairs (Table 4-5). Physical Geology and Historical Geology and Mineralogy and Sedimentation/Stratigraphy represent a fall to spring semester sequence, while the other two pairings represent a spring to fall sequence. These two scenarios present differences in timing with several weeks between testing in the first case and several months in the latter case. Our findings correspond to these differences. In the fall/spring pairings, there is not a significant difference between the STAT-Post from the fall course and STAT-Pre from the spring course. There is however a statistically significant drop in scores for the spring/fall course pairings (Table 4-5). Both findings suggest a low likelihood that the test-retest effect has an impact on students' STAT scores.

Table 4-5
STAT scores and gains – test/retest analysis

Course Sequence	STAT-Post Score	STAT-Pre Score	Difference
Physical Geology to Historical Geology	51.78	52.08	0.30
Historical Geology to Mineralogy	59.50	50.28	-9.21*
Mineralogy to Sedimentology/Stratigraphy	60.46	57.78	-2.69
Sedimentology/Stratigraphy to Structural Geology	67.08	53.88	-13.2**

Note. Scores are normalized as percentages. *p < 0.05, **p < 0.01, ***p<0.001

As an additional method of assessing a potential test-retest effect, several multiple linear regression models were computed using Stata. As described in the Methods section, if a student completed the STAT-Pre or STAT-Post multiple times in a semester, only their first attempt of each the test was recorded. We chose this data management strategy to mitigate any potential test-retest effect in our inferential analyses. However, if we want to evaluate if the test-retest effect exists in our data, students' multiple attempts within a semester should be included. Modifying our dataset resulted in 8 less observations, as students' first attempt was used for all courses they were enrolled in that semester, even if they missed a testing period. For example, in a spring semester, a student may have missed the STAT-Pre in Sedimentology/Stratigraphy but took it in Geomorphology. In the original dataset, the Geomorphology scores were recorded for both courses, but in the modified dataset the student's data was removed.

We predicted students' STAT-Post score by generating four multiple linear regression models (Table 4-6). Models 1 through 3 used the entire modified dataset. Independent variables for all models included test-related factors (STAT-Pre score, number of test attempts), demographic characteristics (gender, race/ethnicity, major), and past/current enrollment in a 1-credit hour spatial thinking training course. The key distinction between Models 1 and 2 lies in the treatment of the major variable. In Model 1, the binary variable indicates whether a student is a geology major or not, while in Model 2, it denotes whether a student belongs to a STEM major or not. Model 3 includes the same variables as Model 1 with the inclusion of a binary variable that identifies whether a student has been enrolled in a spatial training course. In Model 4 (same variables as Model 3), Physical Geology students who took the STAT only one time were removed, as they constituted ~15% of the dataset and were possibly skewing the data to imply that fewer attempts has greater effect on students' STAT-Post scores.

Table 4-6
Multiple linear regression models

	Model 1 *b* (se)	Model 2 *b* (se)	Model 3 *b* (se)	Model 4 *b* (se)
STAT-Pre Score	0.84*** (0.05)	0.84*** (0.05)	0.83*** (0.05)	0.75*** (0.07)
# Attempts	-1.43 (0.74)	-1.06 (0.71)	-1.84** (0.75)	-1.34 (0.85)
Non-Male	-3.11 (1.62)	-3.37* (1.65)	-2.53 (1.62)	-5.17** (1.92)
Non-White	-1.06 (2.01)	-1.02 (2.00)	-1.34 (2.00)	-2.51 (2.10)
Geology Major	2.53 (1.64)		1.80 (1.61)	1.10 (1.99)
STEM Major		1.76 (2.12)		
Spatial Thinking Course			5.36* (2.59)	5.19 (2.66)
Constant	19.99 (2.84)	19.76 (2.89)	20.85 (2.85)	26.02 (4.71)

Note. Multiple linear regression models predicting STAT-Post scores by testing and demographic factors. *b* = coefficients, (se) = robust standard errors. The Non-Male variable represents students who identified as female, transgender, and nonbinary. The Non-White variable represents students who indicated their race/ethnicity as anything other than "white". Models 1 through 3 utilized the full, modified dataset and Model 4 does not include Physical Geology students who only took the STAT once. ***$p < 0.001$; **$p < 0.01$; *$p < 0.05$

All four models accounted for ~67% of the variance in STAT-Post scores and posited that STAT-Pre scores are statistically significant predictor of the post-test scores; however, this is likely not practically significant. Model 3 showed that the number of attempts at the STAT tests predicted a negative effect on students' STAT-Post scores, further supporting our previous finding that the test-retest effect has no impact or a negative on students' scores. Other notable findings include the role of gender (Models 2 and 4) and enrollment in a geology-specific spatial training course. Models 2 and 4 both predicted that non-male students would earn significantly lower STAT-Post scores (~3.4% - 5.2%) than their male peers. While several regression analyses revealed discouraging outcomes for non-male students, there may be a silver lining as Models 3 and 4 predicted that enrollment in a 1-credit spatial training course could significantly improve STAT-Post scores by ~5.2% to 5.4%. Only 12% of the dataset included students who enrolled in this 1-credit hour course, but its effects appear promising.

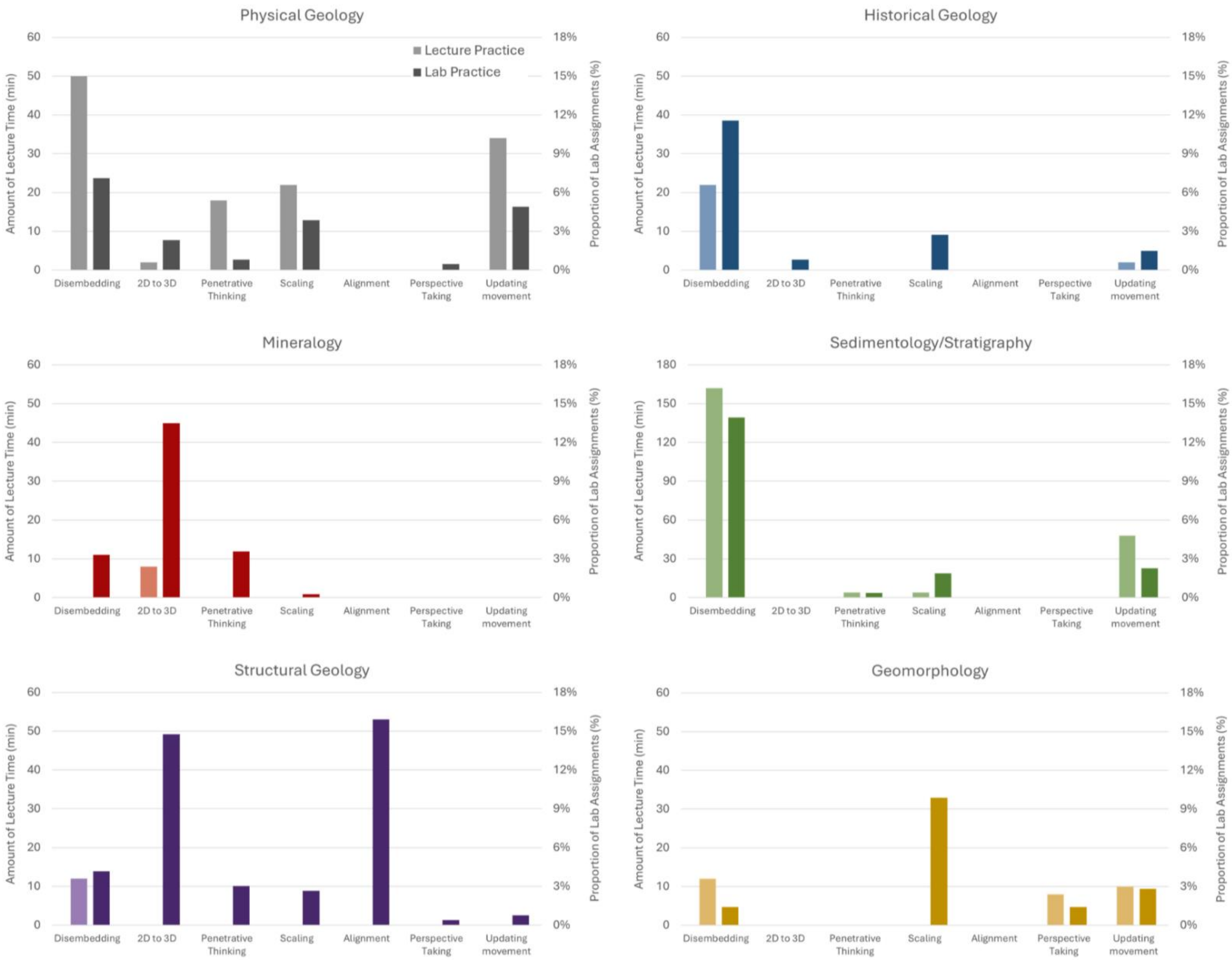
Observations

In Sabatini and McConnell (in preparation), we determined the amount/proportion of time students are exposed to spatial thinking content in the lecture sessions and lab assignments for the courses discussed in this manuscript. We parsed out students' exposure to each of the spatial categories and spatial skills, but we did not further delineate whether that exposure was via passive instruction or active practice. We provide that distinction in Figure 4-5 for the seven skills covered by the STAT.

This visualization reveals that, for this set of skills, if students practiced the skill in lecture, it was also practiced in the lab. The reverse relationship was not true, and more laboratory practice occurred than lecture practice (Figure 4-5). An exception to this finding was the Sedimentology/Stratigraphy course which contained comparable amounts of spatial thinking practice in both settings. For at least five out of the seven skills, there was 15 minutes of practice or less in lecture (except Physical Geology), with similar trends seen in the lab of the courses. In Mineralogy and Structural Geology, practice in only one of the seven skills occurred in the lecture. Other findings revealed that Physical Geology was unique in offering practice across all four spatial categories in both lecture and lab sessions.

Figure 4-5

Amount/proportion of time students practiced spatial thinking skills



Note. The light-colored bars for each course represent lecture (left vertical axis); the dark-colored bars represent labs (right-vertical axis). For the Sedimentology/Stratigraphy, the left-vertical axis encompasses a larger range (up to 180 minutes) than for the other courses (up to 60 minutes). The lab data for Geomorphology represents course homework assignments.

In Sedimentology/Stratigraphy and Historical Geology, disembedding practice was evident in both lecture and lab, although Historical Geology had two hours less practice of in lecture compared to Sedimentology/Stratigraphy, while lab practice remained similar. Mineralogy lacked extrinsic practice entirely and Structural Geology lab showed abundant practice opportunities in 2D to 3D and Alignment.

DISCUSSION

RQ1: How are spatial thinking skills developed by students at multiple points in a traditional undergraduate geology curriculum?

The results described above corroborate findings from past work (Ormand et al., 2014; 2017; Titus & Horsman, 2009; Gold et al., 2018) in that students enter and matriculate through an undergraduate geology program with a wide range of skills. Like our colleagues, we found that students achieve higher test scores in upper-level courses (e.g., Structural Geology) compared to students enrolled in introductory geology courses (e.g., Physical Geology; Figure 4-4). Our study stands out from previous research by administering spatial thinking tests across six core geology courses. This comprehensive approach allowed us to delve deeper into students' skill acquisition and its breakdown across a typical undergraduate geology curriculum. For example, we found that students STAT scores (Figure 4-3; Table 4-4) leveled off once students completed their second semester sophomore courses. STAT scores and gains were similar for Sedimentation/Stratigraphy, Structural Geology, and Geomorphology. Findings from Ormand et al. (2014) and Titus and Horsman (2009) suggested this possibility. While test scores were higher for more advanced courses, larger gains were observed in lower-level courses. This stagnation could be due to a ceiling effect or that students rely on other ways of knowing (Hambrick et al., 2011).

We found that students' STAT scores significantly decreased between the end of the spring semester and beginning of the fall semester (Table 4-5). We suggest two potential explanations for this finding. First, summer learning loss is a common phenomenon for students at all educational levels (Cooper, 2003) with undergraduate students losing approximately half a course letter grade of learning during summer break (van de Sande & Reiser, 2018). A second factor contributing to this finding is based on our assumption that student cohorts followed a quasi-longitudinal trajectory. Although this scenario would have been ideal, practical constraints arose as students entered and exited the study at different points in the curriculum. These discrepancies in timing could introduce variability in the students' learning experiences and subsequently impact their performance on the STAT assessments. For instance, students who entered the study later may have missed certain instructional sessions or spatial thinking training activities, resulting in lower average scores compared to those who were present for the full duration of the study.

Aggregate gains were statistically significant for all courses; however, improvements primarily came from the intrinsic static and intrinsic dynamic categories (Table 4-2). Disembedding was the only intrinsic static skill assessed by the STAT using questions from the Hidden Figures Test. With the exception of one course (Ormand et al (2014), previous studies have typically not registered a significant improvement in disembedding skills measured by the Hidden Figures Test (Ormand et al, 2014; Gold et al., 2018). Disembedding is required to successfully perform other spatial skills, as a person must first focus in on the pertinent spatial information before they can mentally transform those characteristics or relate them to other objects. It is reasonable then that improvements in students' disembedding skills would contribute to gains in other skills also improve. On the other hand, our findings regarding intrinsic dynamic skills align with previous studies that reported substantial improvements, however score ranges varied amongst studies. Physical Geology and Structural Geology students in our study showed similar pre- and post-test scores for the Surface Development Test (Table 4-4) as those in Titus and Horsman (2009), however, Sedimentology/Stratigraphy students in the Kreager et al. (2022) study attained lower scores than those in our investigation (Table 4-4). Previous GBCT findings from Bagher et al. (2022) indicated slightly higher levels of penetrative thinking in Physical Geology compared to our study (Table 4-4). Conversely, Ormand et al. (2017) reported slightly lower GBCT scores in Mineralogy (Table 4-4. In Structural Geology, our study showed similar GBCT scores (Table 4-4) compared to previous research by Ormand et al., (2014; 2017) and Hannula (2019), which reported ranges of 49-74%.

By comparison, gains in the extrinsic domains occurred in fewer courses and were smaller compared to gains for intrinsic skills (Table 4-4). The Zoom Assessment in Physical Geology was the only test that showed statistically significant gains for the extrinsic static category. In contrast, in the extrinsic dynamic category, there were consistent improvements in Spatial Orientation Test scores for all courses except Mineralogy (Table 4-4). We are confident in the findings for the Spatial Orientation Test due to its robust trustworthiness established in cognitive psychology (Kozhevnikov & Hegarty, 2001) and use in geosciences (Polifka et al., 2022). This is the first time that these skills have been reported to improve in geology courses. Water Level Test gains/losses were never significant, a result that is consistent with previous work (Ormand et al, 2017, Hannula, 2019). This may be due to a ceiling effect as students' pre-test scores on this measure were the highest of all the seven sub-sections, leaving little room for improvement (Table 4-4). Additionally, some researchers believe the Water Level Test does not accurately measure students' perception of horizontal (Vasta & Liben, 1996). The second extrinsic dynamic category test, the Topographic Map Assessment, showed large gains for Physical Geology students but no other significant improvements. The Physical Geology gains can be attributed to the last two laboratory assignments in that course which focused on interpreting topographic maps. Subsequent to the Physical Geology course, the TMA pre-test scores were among the highest recorded for all tests, perhaps limiting the potential for additional improvements.

Table 4-7
Joint display table

	STAT Scores		Course Observations	Mixed Methods Inferences
	Test Subsection	Gain (%)	Skill (listed in order of exposure prevalence)	How quantitative and qualitative results converge
Physical Geology	HF	14.17	*Disembedding*	Converge • HF & TMA gains and disembedding & updating movement practice, respectively • ZA gains and scaling exposure in lecture Diverge • 2D/3D, penetrative thinking, and perspective taking not emphasized skills, but there were gains for SD, GBCT, & SOT
	SD	23.33	2D/3D	
	GBCT	10.56	Penetrative Thinking	
	ZA	7.50	Scaling	
	WLT	-5.56	Alignment	
	SOT	16.25	Perspective Taking	
	TMA	20.28	**Updating movement**	
Historical Geology	HF	17.90	**Disembedding**	Converge • HF gains & disembedding practice Diverge • 2D/3D, penetrative thinking, and perspective taking not emphasized skills, but there were gains for SD, GBCT, & SOT • Students exposed to updating movement content, but no TMA gains
	SD	13.09	2D/3D	
	GBCT	10.49	Penetrative Thinking	
	ZA	2.47	Scaling	
	WLT	-8.64	Alignment	
	SOT	7.40	Perspective Taking	
	TMA	0.62	Updating movement	
Mineralogy	HF	22.00	Disembedding	Converge • SD gain and substantial 2D/3D practice in lecture & lab Diverge • Disembedding and penetrative thinking not emphasized skills, but gains observed for HF and GBCT
	SD	21.87	**2D/3D**	
	GBCT	16.00	Penetrative Thinking	
	ZA	-2.67	Scaling	
	WLT	3.33	Alignment	
	SOT	4.50	Perspective Taking	
	TMA	3.33	Updating movement	

Table 4-7 (continued).

Table 4-7 continued	STAT Scores		Course Observations	Mixed Methods Inferences
	Test Subsection	Gain (%)	Skill (listed in order of exposure prevalence)	How quantitative and qualitative results converge
Sedimentology/ Stratigraphy	HF	21.33	**Disembedding**	Converge • HF gains & significant amount of disembedding practice in lecture & lab Diverge • 2D/3D, penetrative thinking, and perspective taking not emphasized skills, but there were gains for SD, GBCT, & SOT
	SD	14.13	2D/3D	
	GBCT	8.67	Penetrative Thinking	
	ZA	5.33	Scaling	
	WLT	2.00	Alignment	
	SOT	10.00	Perspective Taking	
	TMA	-4.00	*Updating movement*	
Structural Geology	HF	9.72	Disembedding	Converge • Emphasized skills and gains in all spatial categories (except no gains for extrinsic static skills) Diverge • Gains not associated with individual, emphasized skills
	SD	4.72	**2D/3D**	
	GBCT	13.89	Penetrative Thinking	
	ZA	4.86	Scaling	
	WLT	2.78	**Alignment**	
	SOT	9.90	Perspective Taking	
	TMA	3.47	Updating movement	
Geomorphology	HF	19.57	Disembedding	Converge • Gains and emphasized skills in intrinsic dynamic and extrinsic dynamic categories Diverge • Gains not associated with individual, emphasized skills
	SD	7.54	2D/3D	
	GBCT	10.14	Penetrative thinking	
	ZA	1.45	**Scaling**	
	WLT	0.00	Alignment	
	SOT	15.22	Perspective Taking	
	TMA	4.35	Updating movement	

Note. Key findings from quantitative and qualitative analyses. Only significant gains between STAT-Pre and STAT-Post scores were included. Skills were listed in the *Course Observations* column if at least one of two criteria were met. **Bolded skill** met both the exposure and practice criteria. *Italicized skills* met only the practice criteria. Underlined skills met only the exposure criteria. Greyed out tests and/or skills indicate the gains were not significant and/or skill did not meet either *Course Observation* criteria, respectively.. Test subsection abbreviations: Hidden Figures test (HF), Surface Development test (SD), Geologic Block Cross-Sectioning Test (GBCT), Zoom Assessment (ZA), Water Level Test (WLT), Spatial Orientation Test (SOT), Topographic Map Assessment (TMA).

RQ2: To what extent do undergraduate students' scores on spatial thinking skills tests converge with their instructors' integration spatial thinking skills in their geology courses?

The primary goal of this project was to investigate the role course content and instructional mode played in spatial skill development as students progressed through an undergraduate geology curriculum. Table 4-7 displays the key findings from the quantitative and qualitative strands of this study. Joint displays like this are commonly used in mixed methods investigations as they enable researchers to gain deeper insights into complex phenomena.

Quantitative data was included in the joint display (Table 4-7) if the differences between the pre- and post-test sub-section scores were positive and statistically significant. The STAT assessed seven of the thirteen skills that were referenced in Newcombe and Shipley's (2015) typology. We identified spatial skills in the *Course Observations* column of the joint display (Table 4-7) if students encountered a significant amount of exposure and/or practice in the lecture and/or lab assignments in the course. Sabatini and McConnell (in preparation) defined the exposure criteria as either >120 minutes of lecture that referenced the skill or if the skill was present in 10% of lab activities. The criteria for student practice values were defined as either engaging in practice for 30 minutes or more during lecture sessions or accounting for more than 5% of practice time in lab assignments. We anticipated that courses which met both exposure and practice criteria would likely see greater gains in student test scores on the related skills.

While convergence between the quantitative and qualitative data could often be linked to specific course content and activities, there were many examples where improvements in test scores had no obvious connection with course content. Convergence manifested itself in two ways in our dataset: where there were significant gains in skills that were emphasized and where there were not significant gains for skills that were not emphasized. In Mineralogy labs, 2D/3D skills are highlighted, focusing on stereographic projections and sketching and averaged ~22% gains (Tables 4-4, 4-7). Disembedding gains in Physical Geology, Historical Geology, and Sedimentology/Stratigraphy improved can be attributed to significant practice in lecture and labs. For example, Sedimentology/Stratigraphy students were exposed to nearly three hours of disembedding content in lecture and students completed three laboratory assignments dedicated to various forms of stratigraphic correlation. complemented by several additional labs focused on the identification of sedimentary rocks, lending to students' encounters with disembedding. Likewise, there were also similar types of lab assignments in Physical Geology and Historical Geology. The scaling and updating movement skills also exhibited this type of convergence in Physical Geology. We attributed these gains to lectures that included considerable time devoted to the size of objects and lab periods that were devoted to topographic map interpretation (Tables 4-4, 4-7).

We considered convergence to also include situations in which there were no gains for skills which were not overly emphasized in the geology course. The Mineralogy course stands out as a notable example of this phenomenon, with extrinsic categories being virtually absent from the course content, leading to no observable gains in extrinsic skills for students. Further, extrinsic static skills (e.g., scaling, alignment) show up rarely in the classes we observed, perhaps contributing to the lack of Zoom Assessment and Water Level Test gains in nearly all courses.

Divergence in the data becomes evident when significant gains in spatial skills are expected based on course content and activities, but these improvements are not observed in students' test scores or vice versa. The former case was most apparent in Structural Geology, where students spent ample lab time engaged in 2D/3D thinking and alignment, via plotting/interpreting stereographic projections and measuring strike/dip of bedding planes, respectively. It is possible that these differences could be attributed to a ceiling effect for the Surface Development test and WLT for upper-level students. This pattern was also observed for the updating movement skill in Historical Geology, Sedimentology/Stratigraphy, and Geomorphology. Updating movement examples in these courses more closely resembled mental animation of dynamic processes (e.g., mountain building, dune migration, sediment transport), which was not captured with the TMA items used in the STAT,

Overlap and correlation between skills in the same spatial category presented a possible cause for divergence where scores improved despite an apparent absence of the related skill. For example, mental transformation was an emphasized skill in Geomorphology (Sabatini & McConnell, in preparation) but was not measured by the STAT. However, Geomorphology students showed significant gains on the Surface Development Test which simulates the mental transformation of a 2D figure into a 3D shape. Ormand et al. (2014) found a moderate correlation between intrinsic dynamic skills, therefore explaining a possible interaction. Methodological constraints pose an additional reason for divergence (i.e., course observations only included lecture sessions and lab assignments). For example, Sabatini and McConnell (in preparation) concluded that students are rarely exposed to perspective taking in lecture and lab settings. Penetrative thinking and perspective taking in geology are often associated with field work and geologic mapping (Hannula, 2019; Kastens & Ishikawa, 2006) , which typically occur on field trips not during lecture sessions. This could have affected our observational inferences and caused us to underestimate its occurrence. Alternatively, students could be developing skills in other geology courses (e.g., hydrogeology, geophysics) while enrolled in the courses we observed.

Limitations

The limited sample size in our study impacts the robustness of our inferences. The study began in the first “regular” semester after the COVID-19 pandemic when attendance was still affected by student illness or campus attendance policies. This resulted in some students missing either the pre-test or post-test who were therefore excluded from the study. Physical Geology was unaffected by the limited sample size issue because it is a large-enrollment course primarily taken by non-majors to fulfill general education lab science requirements.

In an ideal scenario, the study would have tracked multiple student cohorts as they progressed through the program's curriculum sequence. However, practical constraints, such as limited project duration and varied student adherence to the recommended course sequence (e.g., transfer students, course timing limitations) resulted in some students taking courses out of sequence. This issue notably impacted the STAT results for Structural Geology, as students from the Fall 2021 cohort, who contributed significantly to the average performance, were not part of the samples for lower-level courses.

Limited availability of validated and reliable spatial skill assessment instruments poses a challenge, particularly in the context of undergraduate geology education. Among the available tools, some may not be suitable for accurately measuring skill levels in this specific field. For instance, the Zoom Assessment, which prompts students to determine the magnification of small items like pennies or grasshopper legs, may not effectively capture the range of scaling tasks inherent in geology, which span a wide range of scales from atomic to astronomical. Similarly, the items in the Topographic Map Assessment may not fully encompass the entirety of updating movement, particularly in terms of mental animation. The Water Level Test, while commonly used, may not provide an accurate measure of students' perception of horizontality, as scores tend to start and remain consistently high. Additionally, there is a lack of well-established instruments to assess skills such as categorization, sequential thinking, locating, and relations among objects, all of which were commonly observed in the courses included in this study.

In addition to our concerns about instrumentation, we must further discuss the potential impact of the test-retest effect on our study. While our investigation aimed to elucidate this effect, our findings revealed losses. In contrast, other studies employing control group experiments reported gains. For instance, Ormand et al. (2014) observed gains of 5.5% for the Hidden Figures test and 2.4% for the GBCT, although these were not statistically significant. Similarly, Atit et al. (2015) found non-significant gains of 5.9% for the GBCT. Notably, our study recorded significantly higher gains for both instruments across all analyzed courses (see Table 4-4). Moreover, a meta-analysis of spatial training studies underscores the potential impact of the test-retest effect in within-subject studies, such as ours, suggesting an effect size of 0.75. Despite our study revealing losses in both the repeated measures scenario and the regression analysis, it is plausible that these losses could be attributed to test fatigue (Ackerman & Kander, 2009). Consequently, our observed gains may be underestimated due to this repeated test effect.

The complexity inherent in geological tasks and skills impacts both datasets. In the qualitative analysis, our protocol necessitated selecting a single skill to represent the activities in class, typically focusing on the most salient or obvious skill. For instance, when students are tasked with constructing a geologic cross-section, we would typically categorize this as penetrative thinking. However, this activity involves students disembedding relevant information from a geologic map, understanding the orientations of various geologic units and structures in relation to horizontal (alignment), and discerning spatial relationships between them (relations among objects). In the quantitative realm, task complexity is reflected in the instrument items. While the GBCT primarily measures penetrative thinking, it also encompasses elements of disembedding. Thus, isolating individual skills for analysis proves challenging.

CONCLUSIONS

Several key findings emerged from our analysis of student performance and course content in our study investigating spatial thinking skill development across undergraduate geology courses. Despite facing challenges such as a limited sample size and practical constraints due to COVID-19 disruptions, our research shed light on the complexities of skill acquisition in geology education.

Throughout a semester, students' spatial thinking skills showed improvement across various courses, as evidenced by gains in their STAT scores. However, these gains were not uniform across all courses, with upper-level classes generally yielding higher scores compared to introductory ones. At the beginning of students' first geology course (Physical Geology), students' average STAT scores were ~40% and improved by ~12%. Students' scores further improved by ~7% in students' first-year spring course, Historical Geology. Between Historical Geology and Mineralogy, students' STAT scores drop significantly, but rebounded by the end of Mineralogy and increased another 7% by the end of students' second year courses. After this point in the curriculum, STAT scores appear to stabilize. Gains were observed in certain categories, notably intrinsic static and intrinsic dynamic skills, while they were less frequent in others, such as extrinsic static and extrinsic dynamic skills. Specific skills like those assessed by the GBCT (penetrative thinking) and the Surface Development test (visualizing 2D/3D) have been previously documented to show improvement (Titus & Horsman, 2009; Ormand et al., 2014; 2017). Conversely, tests like Hidden Figures (disembedding) and the SOT (perspective taking) exhibited gains that have not been reported by previous studies in the geosciences.

Our analysis revealed the nuanced nature of skill acquisition in geology education has the potential to be influenced by course content and instructional modes. For instance, courses emphasizing tasks such as topographic map interpretation and stratigraphic correlation showed significant gains in corresponding spatial skills. However, challenges persisted in accurately assessing certain skills, such as scaling and updating movement, underscoring the need for validated assessment instruments tailored to the discipline. Despite these obstacles, our mixed methods work reinforces the notion that sustained, targeted spatial training can yield improvements in the spatial skills of geology students. This holds true not only for standard undergraduate courses but also for a 1-credit course specifically designed to enhance geoscience-specific spatial thinking through active practice (see Regression results).

Despite these limitations, our study contributes valuable insights into spatial skill development in undergraduate geology courses. Moving forward, addressing these challenges will be crucial for advancing our understanding of how students acquire and apply spatial thinking skills in the context of geoscience education.

CHAPTER 5 - CONCLUSION

The preceding chapters outlined detailed account of our research process, spanning from identifying needs of the GER community to the development and implementation of a geoscience-specific spatial thinking observation protocol (STOP) and administration of an aggregate spatial thinking test (STAT) in six common undergraduate geology courses. The primary objective was to explore the development of skills among undergraduate students within the geology curriculum, while also investigating the potential impact of instructional methods and course content on skill acquisition. Based on the discussions and conclusions from Chapters , 2, 3, and 4, it is evident that the study of spatial thinking skills in undergraduate geology courses is multifaceted and complex.

Chapter 2 highlighted the diversity in how spatial skills are defined and assessed within geoscience education research. The review emphasized the need for a standardized framework, such as Newcombe and Shipley's spatial thinking typology, to facilitate comprehensive assessments of spatial skills. Furthermore, the chapter underscored the importance of future research focusing on skill development across all categories of the typology, especially those that have received limited attention.

In Chapter 3, the analysis of curricular patterns and data provided insights into the incorporation of spatial thinking instruction across various geology courses. Both lecture and lab components of the courses heavily featured discussions or applications of spatial skills, comprising a significant portion of the curriculum. Intrinsic spatial skills were found to be more prevalent in courses such as Mineralogy and Sedimentology/Stratigraphy, whereas extrinsic skills were commonly observed in courses like Physical Geology, Geomorphology, and

Structural Geology. However, some spatial thinking skills, such as penetrative thinking and alignment, were infrequently taught in the undergraduate geology courses analyzed.

Chapter 4 delved into a quasi-longitudinal assessment of students' spatial thinking skills across multiple geology courses. The results showed improvements in certain skill categories over time, with upper-level courses generally yielding higher scores compared to introductory ones. The nuanced nature of skill acquisition was elucidated, emphasizing the influence of course content and instructional modes on skill development.

Synthesizing these findings, it is apparent that spatial thinking plays a significant role in undergraduate geology education. The study underscores the need for a standardized framework for assessing spatial skills and calls for future research to explore skill development across all categories comprehensively. Additionally, it emphasizes the importance of targeted spatial training and the need for validated assessment instruments tailored to the discipline. Overall, this dissertation contributes valuable insights into spatial skill development in undergraduate geology courses and sets the stage for further exploration in this field. Moving forward, addressing the identified challenges will be crucial for advancing our understanding of how students acquire and apply spatial thinking skills in geoscience education.

REFERENCES

Ackerman, P.L. & Kanfer, R. (2009). Test Length and Cognitive Fatigue: An Empirical Examination of Effects on Performance and Test-Taker Reactions. *Journal of Experimental Psychology: Applied, 15(2),* 163-181. https://doi.org/10.1037/a0015719

Atit, K., Gagnier, K., & Shipley, T.F. (2015). Student Gestures Aid Penetrative Thinking. *Journal of Geoscience Education, 63(1),* 66-72. https://doi.org/10.5408/14-008.1

Atit, K., Weisberg, S.M., Newcombe, N.S., & Shipley, T.F. (2016). Learning to interpret topographic maps: Understanding layered spatial information. *Cognition Research: Principles and Implications, 1(2),* 1-18. https://doi.org/10.1186/s41235-016-0002-y

Bagher, M.M., Sajjadi, P., Carr, J., La Femina, P., & Klippel, A. (2020). Fostering Penetrative Thinking in Geosciences Through Immersive Experiences: A Case Study Visualizing Earthquake Locations in 3D. *Proceedings of 6th International Conference of the Immersive Learning Research Network, iLRN 2020.* 132-139. doi: 10.23919/ iLRN47897.2020.9155123.

Bralower, T.J., Feiss, P.G. & Manduca, C.A. (2008). Preparing a New Generation of Citizens and Scientists to Face Earth's Future. *Liberal Education*, *94*(2), 20-23. Retrieved from https://www.aacu.org/publications-research/periodicals/preparing-new-generation-citizens-and-scientists-face-earths

Carlisle, D., Tyson, J., & Nieswandt, M. (2015). Fostering spatial skill acquisition by general chemistry students. *Chemistry Education Research and Practice*, *16*, 478-517. https://doi.org/10.1039/C4RP00228H

Chatterjee, A. (2008). The Neural Organization of Spatial Thought and Language. *Seminars in Speech and Language, 29(3),* 226-238. DOI:10.1055/s-0028-1082886

Cheek, K.A., LaDue, N.D., & Shipley, T.F. (2017). Learning About Spatial and Temporal Scale: Current Research, Psychological Processes, and Classroom Implications. *Journal of Geoscience Education, 65(4),* 455-472. https://doi.org/10.5408/16-213.1

Cooper, H. (2003). Summer learning loss: The problem and some solutions. Retrieved from https://eric.ed.gov/?id=ED475391

Cooper, M.M., Caballero, M.D., Ebert-May, D., Fata-Hartley, C.L., Jardeleza, S.E., Krajcik, J.S., Laverty, J.T., Matz, R.L., Posey, L.A., and Underwood, S.M., (2015). Challenge faculty to transform STEM learning. *Science, 350*(6258), 281-282. DOI: 10.1126/science.aab0933

Crawford, A.B., and Burnham, P.S. (1946). *Forecasting college achievement: A survey of aptitude tests for higher educations*. New Haven, Yale University Press

Creswell, J. W. (Ed.). (2003). Research design: Qualitative, quantitative, and mixed methods approaches. (2nd ed.). Thousand Oaks: Sage.

Creswell, J. & Plano Clark, V. (2018). *Designing and conducting mixed methods research*. (3rd ed). Thousand Oaks, CA: Sage.

Egger, A.E., Viskupic, K., & Iverson, E.R. (2019). Results of the National Geoscience Faculty Survey (2004-2016). *National Association of Geoscience Teachers*, 82. Retrieved from https://serc.carleton.edu/NAGTWorkshops/CE_geo_survey/index.html

Ekstrom, R. B., Dermen, D., & Harman, H. H. (1976). *Manual for kit of factor-referenced cognitive tests*. Retrieved from https://www.ets.org/Media/Research/pdf/Manual_for_Kit_of_Factor-Referenced_Cognitive_Tests.pdf

Eliot, J. (1987). The Psychometric Approach. In *Models of Psychological Space: Psychometric, Developmental, and Experimental Approaches* (pp. 37-82). Springer. https://doi.org/10.1007/978-1-4612-4788-3

Freeman, S., Eddy, S.L., McDonough, M., Smith, M.K., Okoroafor, N., Jordt, H., & Wenderoth, M.P. (2014). Active learning increases student performance in science, engineering, and mathematics. *PNAS, 111(3),* 8410-8415. https://doi.org/10.1073/pnas.131903011

Giorgis, S. (2015). Google Earth mapping exercises for structural geology students—A promising intervention for improving penetrative visualization ability. *Journal of Geoscience Education, 63*(2), 140-146. https://doi.org/10.5408/13-108.1

Giorgis, S., Mahlen, N., & Anne, K. (2017). Instructor-Led Approach to Integrating an Augmented Reality Sandbox into a Large Enrollment Introductory Geoscience Course for Nonmajors Produces No Gains. *Journal of Geoscience Education, 65(3)*, 283-291. https://doi.org/10.5408/17-255.1

Gold, A.U., Pendergast, P.M., Ormand, C.J., Budd, D.A., & Mueller, K.J. (2018). Improving spatial thinking skills among undergraduate geology students through short online training exercises. *International Journal of Science Education, 40(18),* 2205-2225. https://doi.org/10.1080/09500693.2018.1525621

Guay, R. B. (1976). *Purdue Spatial Visualization Test*. Purdue Research Foundation.

Hambrick, D. Z., Libarkin, J. C, Petcovic, H. L., Baker, K.M., Elkins, J., Callahan, C.N., Turner, S. P., Rench, T.A. & LaDue, N. D. (2012). A test of the circumvention- of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology*, *141(3),* 397-403. https://doi.org/10.1037/a0025927

Hannula, K.A. (2019). Do geology field courses improve penetrative thinking? *Journal of Geoscience Education*, *67(2)*, 143-160. https://doi.org/10.1080/10899995.2018.1548004

Hegarty, M., Newcombe, N.S., Goodchild, M.F., Janelle, D.G., Shipley, T.F., & Sinton, D. (2012). Spatial Thinking Across the College Curriculum. *Specialist Meeting Report*. Retrieved from https://www.academia.edu/66792734/Spatial_Thinking_across_the_College_Curriculum_Final_Report

Hogg, R.V., Tanis, E. and Zimmerman, D. (2015). *Probability and Statistical Inference. 9th Edition,* Pearson.

Hoyek, N., Collet, C., Rastello, O., Fargier, P., Thiriet, P., & Guillot, A. (2009). Enhancement of mental rotation abilities and its effect on anatomy learning. *Teaching and Learning in Medicine, 21(3)*, 201–206. DOI: 10.1080/10401330903014178

Jackson, D., Kaveh, H., Victoria, J., Walker, A., & Burstyn, N. (2019). Integrating an augmented reality sandbox challenge activity into large-enrollment introductory geoscience lab for nonmajors produces no learning gains. *Journal of Geoscience Education, 67(3),* 237-248. https://doi.org/10.1080/10899995.2019.1583786

Jacovina, M., Ormand, C., Shipley, T. F., & Weisberg, S. (2014). Topographic Map Assessment. Retrieved from https://www.silc.northwestern.edu/topographic-map-assessment-tma/

Jo, I., Hong, J.E., & Verma, K. (2016). Facilitating spatial thinking in world geography using Web-based GIS. *Journal of Geography in Higher Education, 40(3),* 442-459.

Johnson, B. and Turner, L.A. (2003) Data Collection Strategies in Mixed Methods Research. *In Tashakkori, A.M. and Teddlie, C.B., Eds., Handbook of Mixed Methods in Social and Behavioral Research*, SAGE Publications, Thousand Oaks, 297-319.

Jones, M.G., Gardner, G., Taylor, A.R., Wiebe, E., & Forrester, J. (2010). Conceptualizing Magnification and Scale: The Roles of Spatial Visualization and Logical Thinking. *Research in Science Education (41),* 357-368. https://doi.org/10.1007/s11165-010-9169-2

Kali, Y. & Orion, N. (1996). Spatial Abilities of High-School Students in the Perception of Geologic Structures. *Journal of Research in Science Teaching, 33(4),* 369-391. https://doi.org/10.1002/(SICI)1098-2736(199604)33:4<369::AID-TEA2>3.0.CO;2-Q

Kastens, K.A. & Ishikawa, T. (2006). Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. *In Earth and Mind: How Geologists Think and Learn about the Earth*. https://doi.org/10.1130/2006.2413(05)

Kastens, K., Manduca, C.A., Cervato, C., Frodeman, R., Goodwin, C., Libena, L.S., . . . Titus, S. (2009). How Geoscientists Think and Learn. *Eos Trans. AGI, 90(31),* 265. Retrieved from https://serc.carleton.edu/serc/EOS-90-31-2009.html

Klyce, A. & Ryker, K. (2022). What does a degree in geology actually mean? A systematic evaluation of courses required to earn a bachelor of science in geology in the United States. *Journal of Geoscience Education, 71(1),* 3-19. https://doi.org/10.1080/10899995.2022.2076201

Kozhevnikov, M. & Hegarty, M. (2001). A dissociation between object-manipulation and perspective taking spatial abilities. *Memory & Cognition, 29,* 745-756. https://doi.org/10.3758/BF03200477

Kreager, B.Z., Ladue, N.D., Shipley, T.F., Powell, R.D.,& Hampton, B.A. (2022). Spatial skill predicts success on sequence stratigraphic interpretation. *Geosphere, 18(2),* 750-761. https://doi.org/10.1130/GES02428.1 4

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta- analysis. *Child Development, 56(6),* 1479–1498.

Liben, L.S., Kastens, K.A., & Christensen, A.E. (2011). Spatial Foundations of Science Education: Illustrative Case of Instruction on Introductory Geological Concepts. *Cognition and Instruction, 29(1),* 45-87. https://doi.org/10.1080/07370008.2010.533596

Liben, L.S. & Titus, S.J. (2012). The importance of spatial thinking for geoscience education: Insights from the crossroads of geoscience and cognitive science. In K.A. Kastens and C.A. Manducs (Eds.), *Earth and Mind II: A Synthesis of Research on Thinking and Learning in the Geosciences.* https://doi.org/10.1130/2012.2486(10)

Manduca, C.A. & Kastens, K.A. (2012). Mapping the domain of spatial thinking in the geosciences. In K.A. Kastens and C.A. Manducs (Eds.), *Earth and Mind II: A Synthesis of Research on Thinking and Learning in the Geosciences.* https://doi.org/10.1130/2012.2486(09)

Mathison, S. (1988).Why triangulate? *Educational Researcher, 17(2)*, 13–17. https://doi.org/10.3102/0013189X017002013

McLaughlin, J.A. & Bailey, J.M. (2022). Students need more practice with spatial thinking in geoscience education: a systematic review of the literature. *Studies in Science Education,* 1-10. https://doi.org/10.1080/03057267.2022.2029305

Miles, M. B., & Huberman, A. M. (1994). Qualitative data analysis: An expanded source book (2nd ed.). Thousand Oaks, CA: Sage.

National Research Council. (2006). Learning to Think Spatially. Washington, D.C.: National Academies Press.

Newcombe N.S. & Shipley, T.F. (2015). Thinking About Spatial Thinking: New Typology, New Assessments. In: Gero J. (Ed.) *Studying Visual and Spatial Reasoning for Design Creativity*. Springer. https://doi.org/10.1007/978-94-017-9297-4_10

Nyarko, S. (2021). *In an Era of Soft Skills: Investigating Teamwork Skills in Geosciences* (3801) [Doctoral dissertation, Western Michigan University]. ScholarWorks.

Onweugbuzie, A.J., & Teddlie, C. (2003). A framework for analyzing data in mixed methods research. In A. Tashakkori & C. Teddlie (Eds.), *Handbook of mixed methods in social & behavioral research* (pp.351-383). Thousand Oaks, CA: Sage.

Ormand, C.J., Manduca, C., Shipley, T.F., Tikoff, B., Harwood, C.L., Atit, K., & Boone, A.P. (2014). Evaluating Geoscience Students' Spatial Thinking Skills in a Multi-Institutional Classroom Study. *Journal of Geoscience Education, 62(1),* 146-154. https://doi.org/10.5408/ 13-027.1

Ormand, C.J., Shipley, T.F., Tikoff, B., Dutrow, B., Goodwin, L.B., Hickson, T., . . . Resnick, I. (2017). The Spatial Thinking Workbook: A Research-Validated Skills Curriculum for

Geology Majors. *Journal of Geoscience Education, 65(4),* 423-434. https://doi.org/10.5408/16-210.1

Pallrand, G.J. & Seeber, F. (1984). Spatial Ability and Achievement in Introductory Physics. *Journal of Research in Science Teaching, 21(5),* 507-516. https://doi.org/10.1002/tea.3660210508

Piaget, J., & Inhelder, B. (1967). Systems of reference and horizontal–vertical coordinates. In *The child's conception of space*. New York: W.W. Norton and Co.

Polifka, J.D., Cervato, C. & Holme, T.A. (2022). Measuring the role of spatial ability and multiple external representations in introductory geology students' knowledge of plate tectonics. *Journal of Geoscience Education,* 1-16. https://doi.org/10.1080/10899995.2022.2135351

Resnick, I. & Shipley, T.F. (2013). Breaking new ground in the mind: An initial study of mental brittle transformation and mental rigid rotation in science experts. *Cognitive Processing, 14(2),* 143-152. https://doi.org/10.1007/s10339-013-0548-2

Resnick, I., Davatzes, A., Newcombe, N.S., & Shipley, T.F. (2017). Using analogy to learn about phenomena at scales outside human perception. *Cognitive Research: Principles and Implications, 2(21),* 1-17. DOI: 10.1186/s41235-017-0054-7

Ryker, K., Jaeger, A. J., Brande, S., Guereque, M., Libarkin, J., & Shipley, T. F. (2018). Research on cognitive domain in geoscience learning: Temporal and spatial reasoning. In St. John, K (Ed.) *Community framework for geoscience education research*. National Association of Geoscience Teachers. https://doi.org/10.25885/ger_framework/7

Sabatini, S.M. & McConnell, D.A. (in preparation). Characterizing Spatial Thinking Instruction in Undergraduate Geology Courses Using an Observation Protocol.

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124(2),* 262–274. https://doi.org/10.1037/0033-2909.124.2.262

Shea, D., Lubinski, D., & Benbow, C. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology, 93(3),* 604-614. https://doi.org/10.1037/0022-0663.93.3.604

Shepard, R.N., & Cooper, L.A. (1982). Mental images and their transformations. Cambridge: MIT Press.

Smith, M.K., Jones, F.H.M., Gilbert, S.L., & Wieman, C.E. (2013) The Classroom Observation Protocol for Undergraduate STEM (COPUS): A New Instrument to Characterize University STEM Classroom Practices. *CBE—Life Sciences Education, 12,* 618-627. https://doi.org/10.1187/cbe.13-08-0154

Sorby, S.A. (2009). Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education, 31(3),* 459-480. https://doi.org/10.1080/09500690802595839

St. John, K., & McNeal, K. (2017). The Strength of evidence pyramid: One approach for characterizing the strength of evidence of geoscience education research (GER) community claims. *Journal of Geoscience Education*, 65(4), 363–372. https://doi.org/10.5408/17-264.1

Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology, 69(5)*, 797–811. https://doi.org/10.1037/0022-3514.69.5.797

Stieff, N., Lira, M.E., & Scopelitis, S.A. (2016). Gesture Supports Spatial Thinking in STEM. *Cognition and Instruction, 34(2),* 80-99. https://doi.org/10.1080/07370008.2016.1145122

Tarampi, M.R., Heydari, N., & Hegarty, M. (2016). A Tale of Two Types of Perspective Taking: Sex Differences in Spatial Ability. *Psychological Science, 27(11),* 1507-1516. https://doi.org/10.1177/0956797616667459

Teasdale, R., Viskupic, K., Bartley, J.K., McConnell, D., Manduca, C., Bruckner. M., Farthing, D., & Iverson, E. (2017). A multidimensional assessment of reformed teaching practice in geoscience classrooms. *Geosphere, 13(2),* 608-627. https://doi.org/10.1130/GES01479.1

Titus, S. & Horsman, E. (2009). Characterizing and Improving Spatial Visualization Skills. *Journal of Geoscience Education, 57(4),* 242-254. https://doi.org/10.5408/1.3559671

Uttal, D. H. & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why, and how?. In *Psychology of learning and motivation, 57,* 147-181. https://doi.org/10.1016/B978-0-12-394293-7.00004-2

Uttal, D.H., Meadow, N.G., Tipton, E., Hand, L.L., Alden, A.R., Warren, C., & Newcombe, N.S. (2013). The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. *Psychological Bulletin, 139(2),* 352-402. https://doi.org/10.1037/a0028446

Vandenberg, S.G., & Kuse, A.R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599–604. https://www.silc.northwestern.edu/vandenberg-kuse-mental-rotation-test-redrawn-version/

van de Sande, C. & Reiser, M. (2018). The effect of summer break on engineering student success in calculus. *International Journal of Research in Education and Science (IJRES), 4(2)*, 349-357. DOI:10.21890/ijres.409264

Vasta, R., & Liben, L. S. (1996). The Water-Level Task: An Intriguing Puzzle. *Current Directions in Psychological Science, 5(6)*, 171-177. https://doi.org/10.1111/1467-8721.ep11512379

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

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101(4),* 817–835. https://doi.org/10.1037/a0016127

Wieman, C. (2007). Why Not Try a Scientific Approach to Science Education. *Change: The Magazine of Higher Learning, 39(5),* 9-15. DOI: 10.3200/CHNG.39.5.9-15

APPENDICES

APPENDIX A

Spatial Thinking Aggregate Test

The following pages include questions that are designed to examine different aspects of spatial thinking.
The test is divided into seven separate parts. Several parts begin with sample questions to demonstrate what is expected.
Part 1: Hidden Figures – 6 questions/4 minutes
Part 2: Surface Development – 3 questions/3 minutes
Part 3: Geologic Block Cross-Sectioning – 6 questions/3 minutes
Part 4: Scaling: Zoom Assessment – 6 questions/3 minutes
Part 5: Water Level Test – 3 questions/1 minute
Part 6: Spatial Orientation Test – 4 questions/2 minutes
Part 7: Topographic Map Assessment – 3 questions/3 minutes

Each part has a limited time to complete the questions.
Stop the task when the assigned time comes to an end.
Do not return to the questions later.
Do not begin the next part until notified by the researcher or instructor.

Name (print) ________________________________

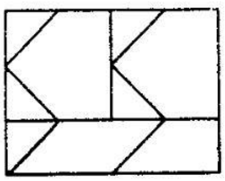
Spatial Thinking Aggregate Pre-Test

Part 1 – Hidden Figures Test

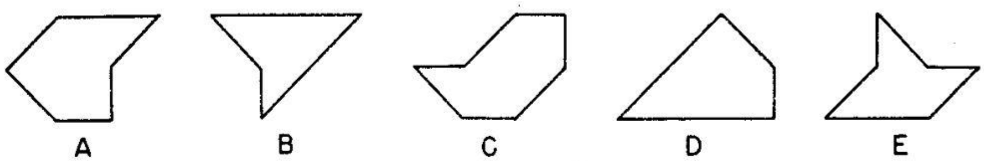
This is a test of your ability to tell which one of five simple figures can be found in a more complex pattern. At the bottom of each question in this test are five simple figures lettered A, B, C, D, and E. Above each row of figures is a pattern. Circle the letter of the figure which you find in the pattern.

NOTE: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five letter figures.

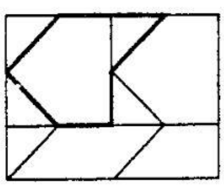
Now try an example. Find one of the five figures below in this pattern.



Circle one:

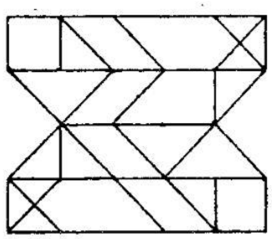


The figure below shows how figure A is included in the problem.

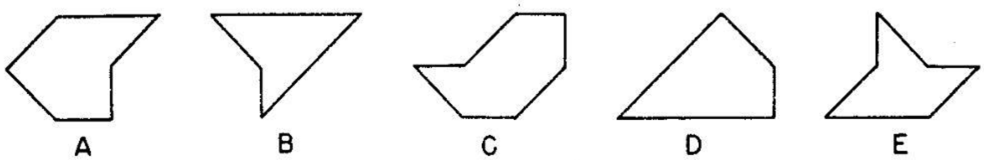


You will have 4 minutes for the next 6 questions. When you have finished, STOP. Please do not go on to Part 2 until you are asked to do so.

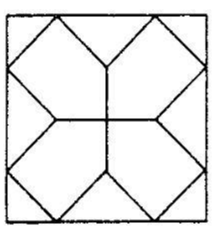
1. Which of the five simple figures is found in the more complex pattern? Note: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.



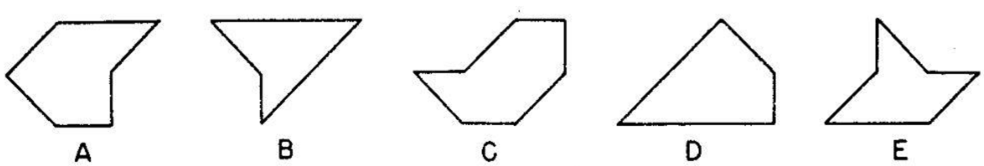
Circle one:



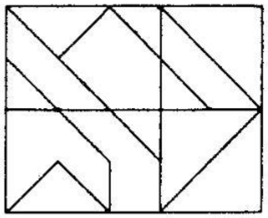
2. Which of the five simple figures is found in the more complex pattern? Note: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.



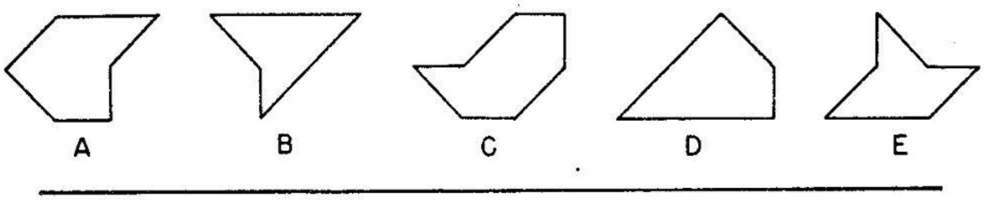
Circle one:



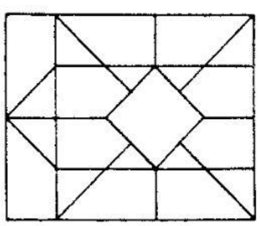
3. Which of the five simple figures is found in the more complex pattern? Note: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.



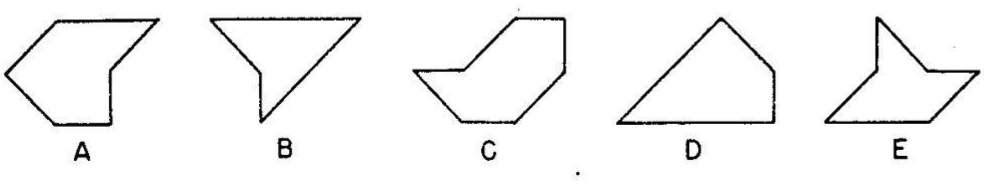
Circle one:



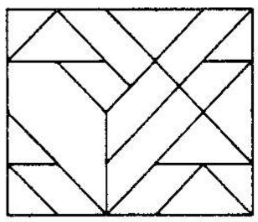
4. Which of the five simple figures is found in the more complex pattern? Note: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.



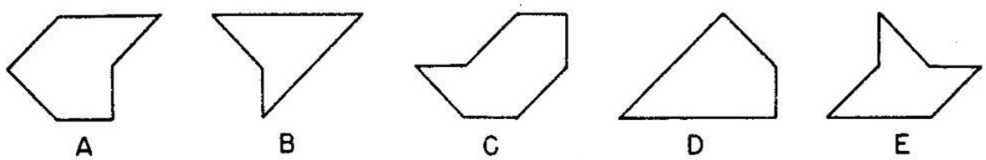
Circle one:



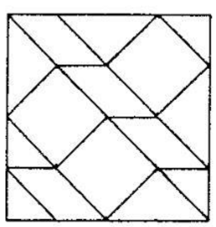
5. Which of the five simple figures is found in the more complex pattern? Note: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.



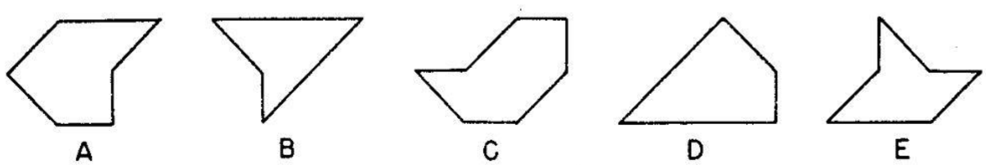
Circle one:



6. Which of the five simple figures is found in the more complex pattern? Note: There is only one of these figures in each pattern, and this figure will always be right side up and exactly the same size as one of the five lettered figures.

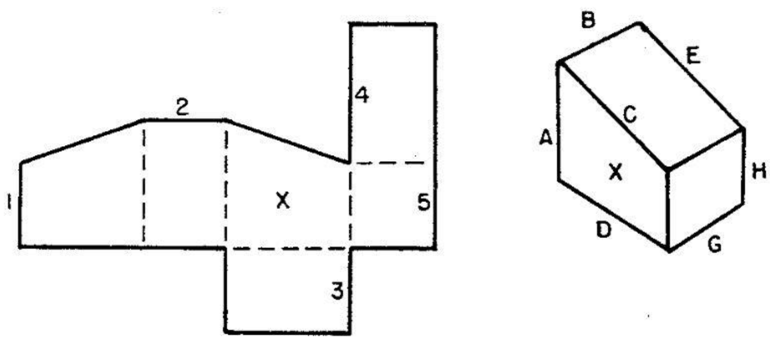


Circle one:



Part 2 – Surface Development Test
In this test you are trying to imagine or visualize how a piece of paper can be folded to form some kind of object. Look at the two drawings below. The drawing on the left is a piece of paper which can be folded on the dotted lines to form the object drawn at the right. You are to imagine the folding and are to figure out which of the letter edges on the object are the same as the numbered edges on the piece of paper at the left. Write letters of the answers in the numbered spaces at the far right.

Now try this practice problem below. Numbers 1 and 4 are already correctly marked for you.



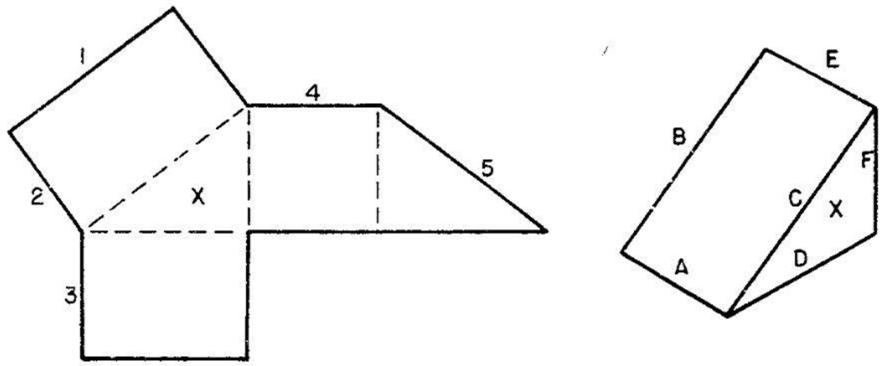
1: H
2:
3:
4: C
5:

NOTE: The side of the flat piece marked with the X will always be the same as the side of the object marked with the X. Therefore, the paper must always be folded so that the X will be on the outside of the object.

In the above problem, of the side with edge 1 is folded around to form the back of the object, the edge 1 will be the same as side H. If the side with edge 5 is folded back, then the side with edge 4 may be folded down so that the edge 4 is the same as edge C. The other answers are as follows: 2 is B; 3 is G: and 5 is H. Notice that two of the answers can be the same.

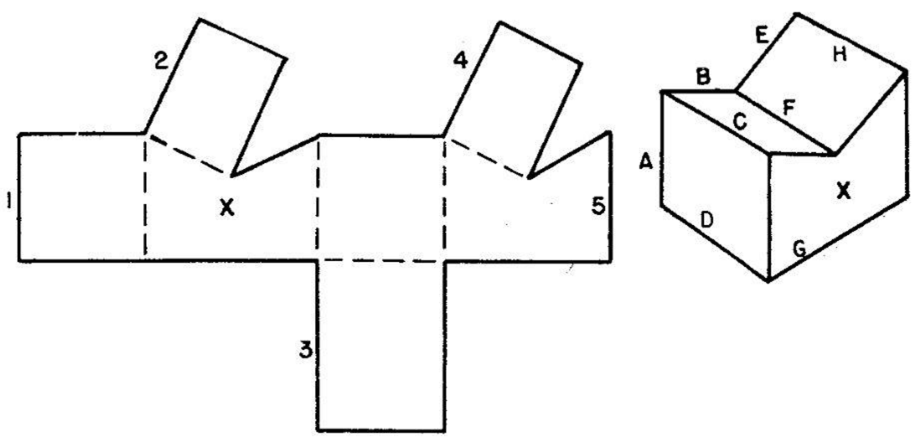
You will have <u>3 minutes</u> for the next <u>3 questions</u>. When you have finished, STOP. Please do not go on to Part 3 until you are asked to do so.

7. The drawing on the left is a piece of paper which can be folded along the dotted lines to form the object shown on the right. Match the lettered edges of the 3D shape to the numbered edges of the 2D paper. Write your answers in the provided box.



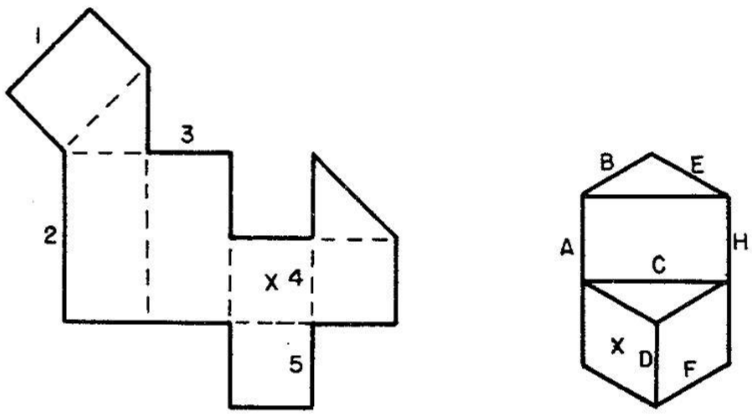
1:
2:
3:
4:
5:

8. The drawing on the left is a piece of paper which can be folded along the dotted lines to form the object shown on the right. Match the lettered edges of the 3D shape to the numbered edges of the 2D paper. Write your answers in the provided box.



1:
2:
3:
4:
5:

9. The drawing on the left is a piece of paper which can be folded along the dotted lines to form the object shown on the right. Match the lettered edges of the 3D shape to the numbered edges of the 2D paper. Write your answers in the provided box.



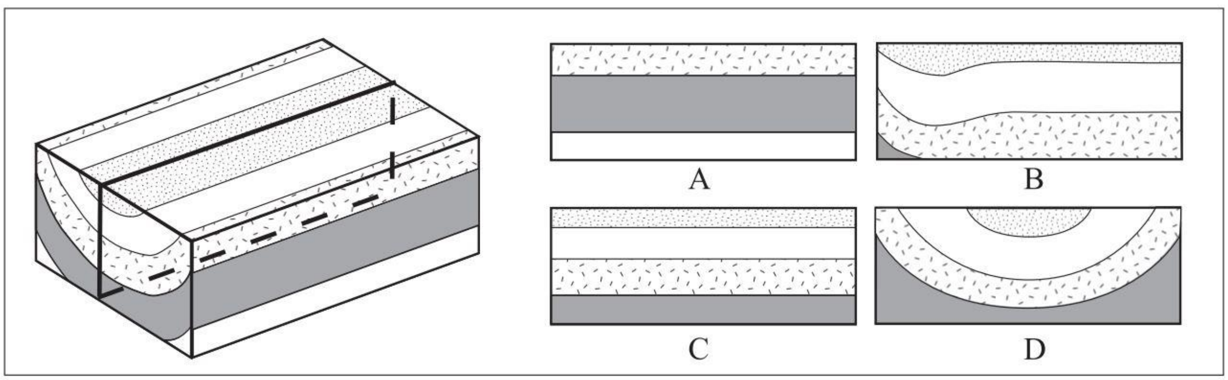
1:
2:
3:
4:
5:

Part 3 – Geologic Block Cross-Sectioning Test
This test is designed to assess your ability to mentally slice through a three-dimensional geologic structure expressed in a block diagram.

For each item below:

1. Study the geologic structure that is displayed in the 3-D block diagram.
2. Determine what the cross-section of that geologic structure would look like on the surface of the vertical plane intersecting the block.
3. Choose the multiple-choice answer that illustrates the structure along that plane. Where more than one answer appears to be possible, choose the MOST LIKELY answer.

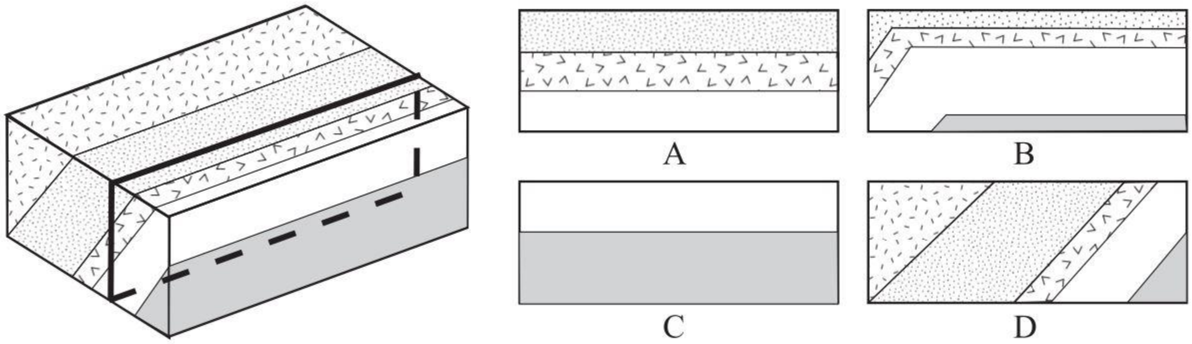
Here is an example:



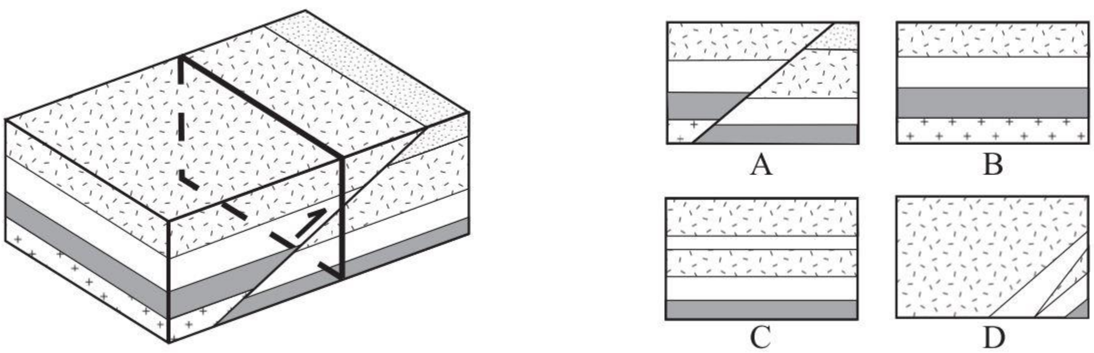
The answer to this example problem is **C**. It shows the layers in the correct positions, with the correct thicknesses, and in the correct orientations.

You will have 3 minutes for the next 6 questions. When you have finished, STOP. Please do not go on to Part 4 until you are asked to do so.

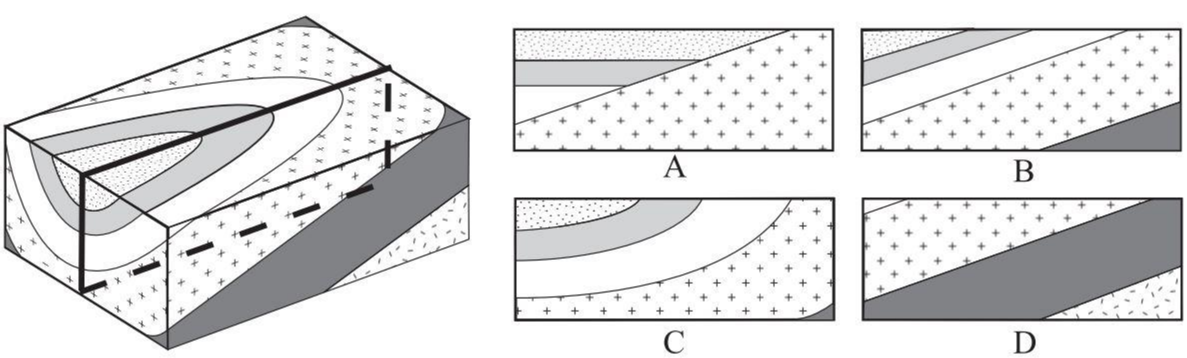
10. Choose which cross-section illustrates what the geologic structure would look like on the surface of the vertical place intersecting the block.



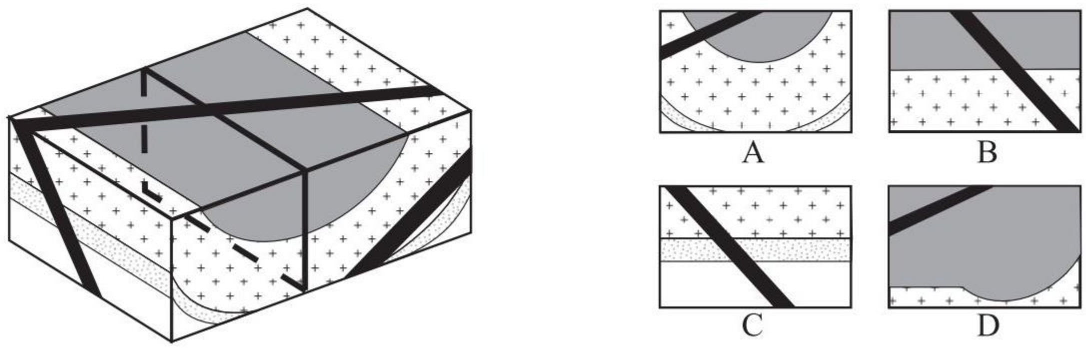
11. Choose which cross-section illustrates what the geologic structure would look like on the surface of the vertical place intersecting the block.



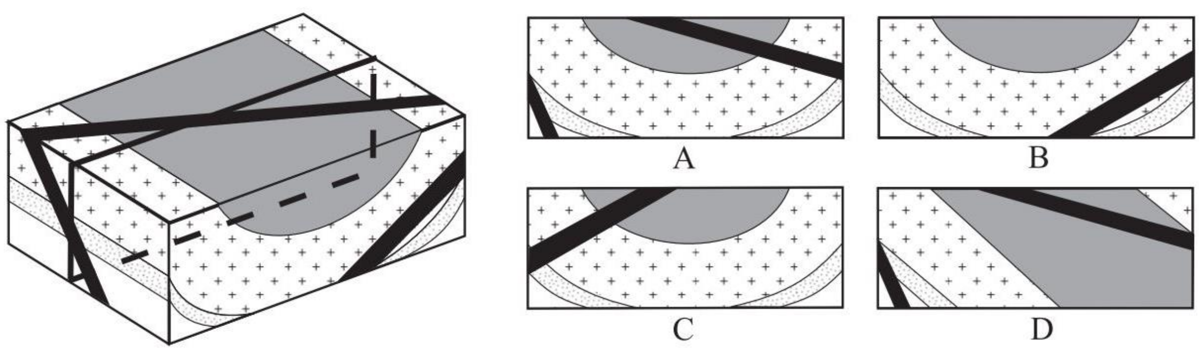
12. Choose which cross-section illustrates what the geologic structure would look like on the surface of the vertical place intersecting the block.



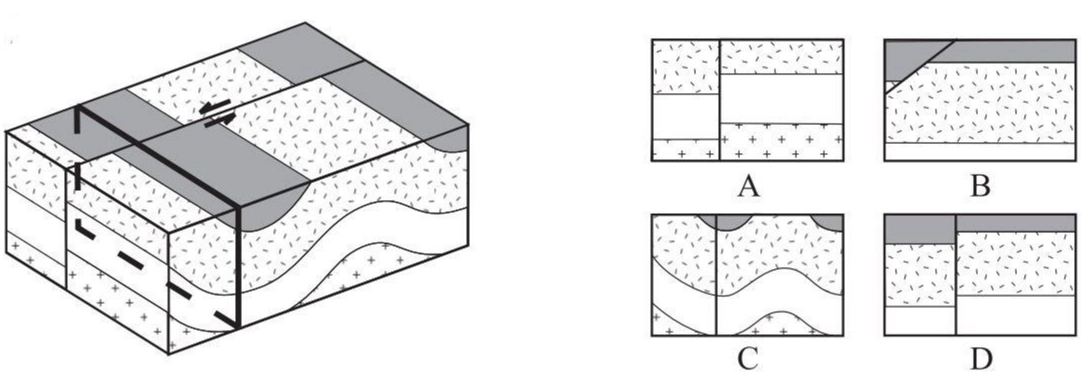
13. Choose which cross-section illustrates what the geologic structure would look like on the surface of the vertical place intersecting the block.



14. Choose which cross-section illustrates what the geologic structure would look like on the surface of the vertical place intersecting the block.



15. Choose which cross-section illustrates what the geologic structure would look like on the surface of the vertical place intersecting the block.



Part 4 – Scaling: Zoom Assessment
The Zoom Assessment involves students' use of magnification and scale and was designed to assess students' knowledge of magnification and includes tasks that called for them to:

1. mentally increase or decrease image size by selected amounts (10×, 100× etc…),
2. calculate sizes of enlarged objects when given an actual size and the number of times the object is enlarged or magnified,
3. identify how objects move when magnified and manipulated under a microscope,
4. enlarge an object and match the enlarged object size to a known object, and
5. observe miniaturized objects (such as a doll house chair) and estimate the number of times the object is smaller than the actual (human-sized) object.

Read the following prompts and answer the questions accordingly.

You will have 3 minutes for the next 6 questions. When you have finished, STOP. Please do not go on to Part 5 until you are asked to do so.

16. If a ladybug is 8mm long and is magnified 20 times (20x) how big would the image of the bug be?

    a. 80 mm
    b. 800 mm
    c. 160 mm
    d. 16 mm

17. Examine the enlarged letter 'a' below. How many times larger is the enlarged letter 'a' compared to this letter 'a'?

    a. 10 times
    b. 25 times
    c. 40 times
    d. 100 times

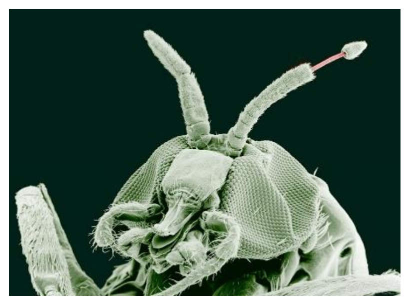
18. If you were to magnify a penny by 50 times (50x), the image of the penny would be approximately as large as which of the following?

a. A typical shoe
b. A baseball bat
c. A car
d. A typical house

19. If a human hair were enlarged to size shown in the image, how many times would it have been enlarged?

a. 5 times (5x)
b. 100 times (100x)
c. 1,000 times (1000x)
d. 10,000 times (10000x)

20. The image shows a fly head that has been magnified 100 times. How big is the fly before it was magnified?



a. 5 mm
b. 20 mm
c. 0.5 m
d. 1 m

21. This image of a penny has been enlarged from the size of the actual penny. How many times would you estimate that the word Liberty has been enlarged?

a. 2 times
b. 5 times
c. 10 times
d. 50 times

Part 5 – Water Level Task

In this test, you are to imagine a bottle half full of water and what the water level line would look like if the bottle was not in a vertical position.

For each item below:

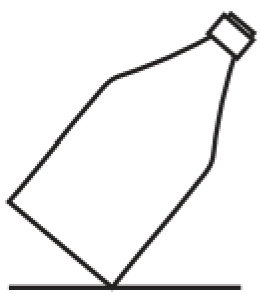
1. Study the bottle and imagine it half full of water.

2. Determine what the water level would look like in the bottle.

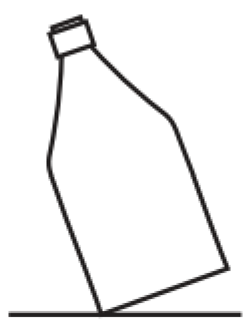
3. Draw a line to represent the water level surface in the bottle.

You will have <u>1 minute</u> for the next <u>3 questions</u>. When you have finished, STOP. Please do not go on to Part 6 until you are asked to do so.

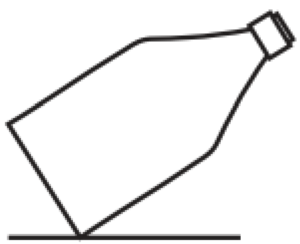
21. Draw a line to show how the water line would look if the bottle were half full of water.



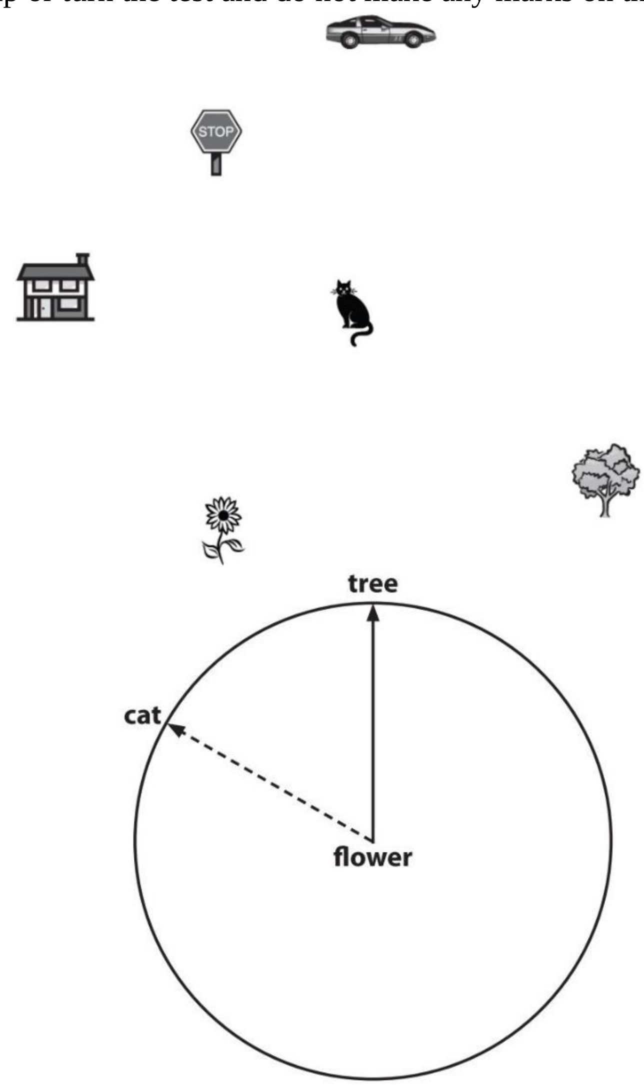
22. Draw a line to show how the water line would look if the bottle were half full of water.



23. Draw a line to show how the water line would look if the bottle were half full of water.

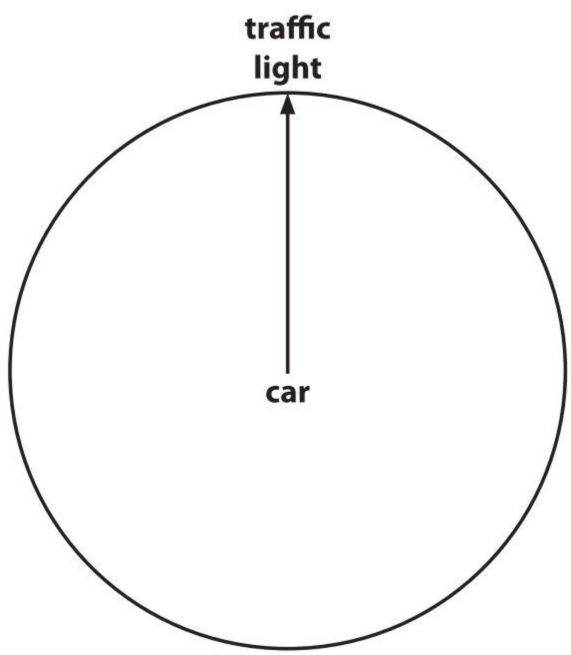


Part 6 – Perspective Taking/Spatial Orientation Test
This is a test of your ability to imagine different perspectives in space. On each of the following pages you will see a picture of an array of objects and an "arrow circle" with a question about the direction between some of the objects.
Look at the sample item below. In this item you are asked to imagine that you are standing at the flower, which is named in the center of the circle, and facing the tree, which is named at the top of the circle. Your task is to draw an arrow pointing to the cat. In the sample item this arrow has been drawn for you. In the test items, your task is to draw this arrow. Can you see that if you were at the flower facing the tree, the cat would be in this direction?
Please do not pick up or turn the test and do not make any marks on the maps.

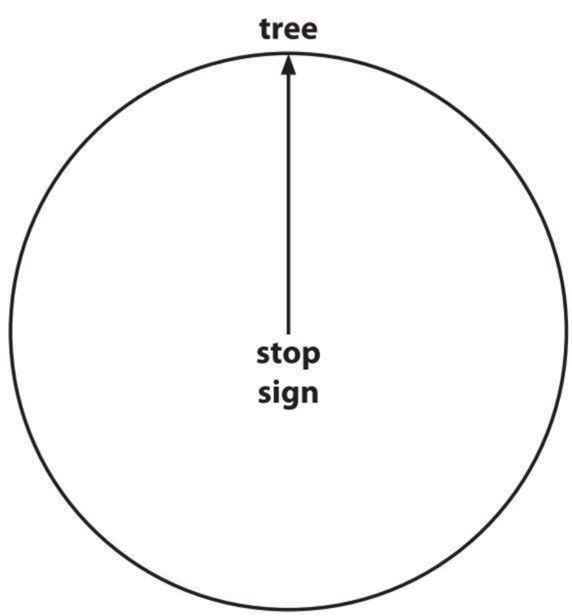


You will have <u>2 minutes</u> for the next <u>4 questions</u>. When you have finished, STOP. Please do not go on to Part 7 until you are asked to do so.

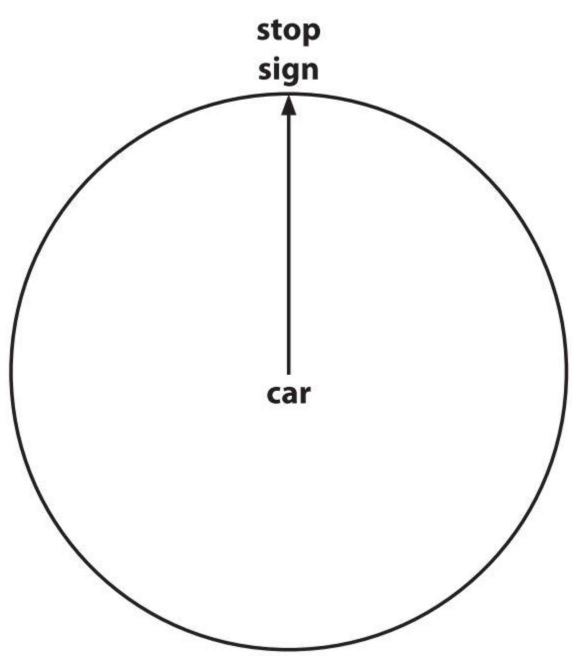
24. Imagine you are standing at the **car** and facing the **traffic light**. Draw an arrow pointing to the **stop sign.** Do not pick up or rotate the test.



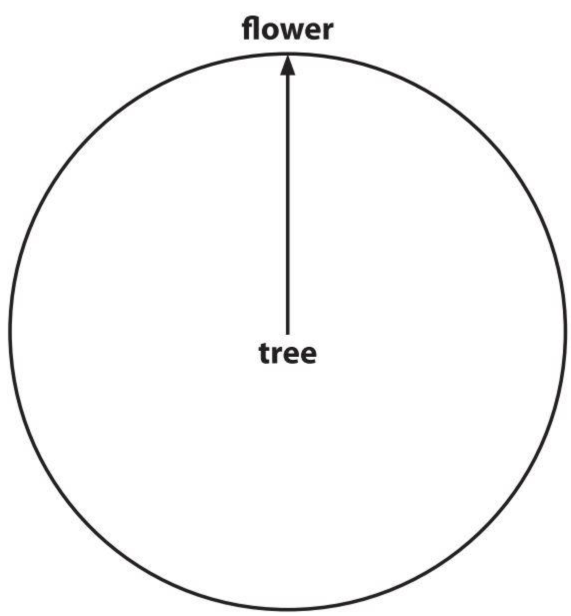
25. Imagine you are standing at the **stop sign** and facing the **tree**. Draw an arrow pointing to the **traffic light.** Do not pick up or rotate the test.



26. Imagine you are standing at the **car** and facing the **stop sign**. Draw an arrow pointing to the **tree.** Do not pick up or rotate the test.



27. Imagine you are standing at the **tree** and facing the **flower**. Draw an arrow pointing to the **house.** Do not pick up or rotate the test.



Part 7 – Topographic Map Assessment

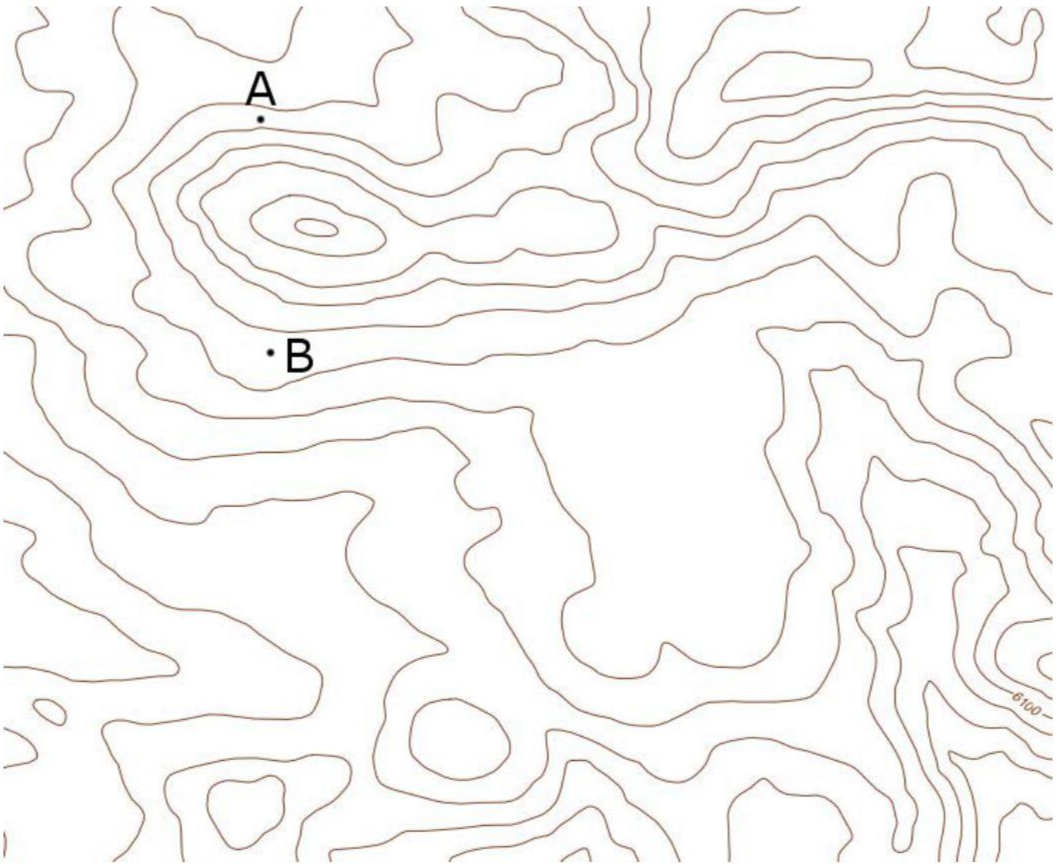
The Topographic Map Assessment involves the use and understanding of topographic maps. Individuals must be able to understand the rules/conventions of topographic maps and be able to visualize terrains from contour maps to solve problems correctly.

The following set of problems involve updating an object's (yourself or a stream) position over time and space.

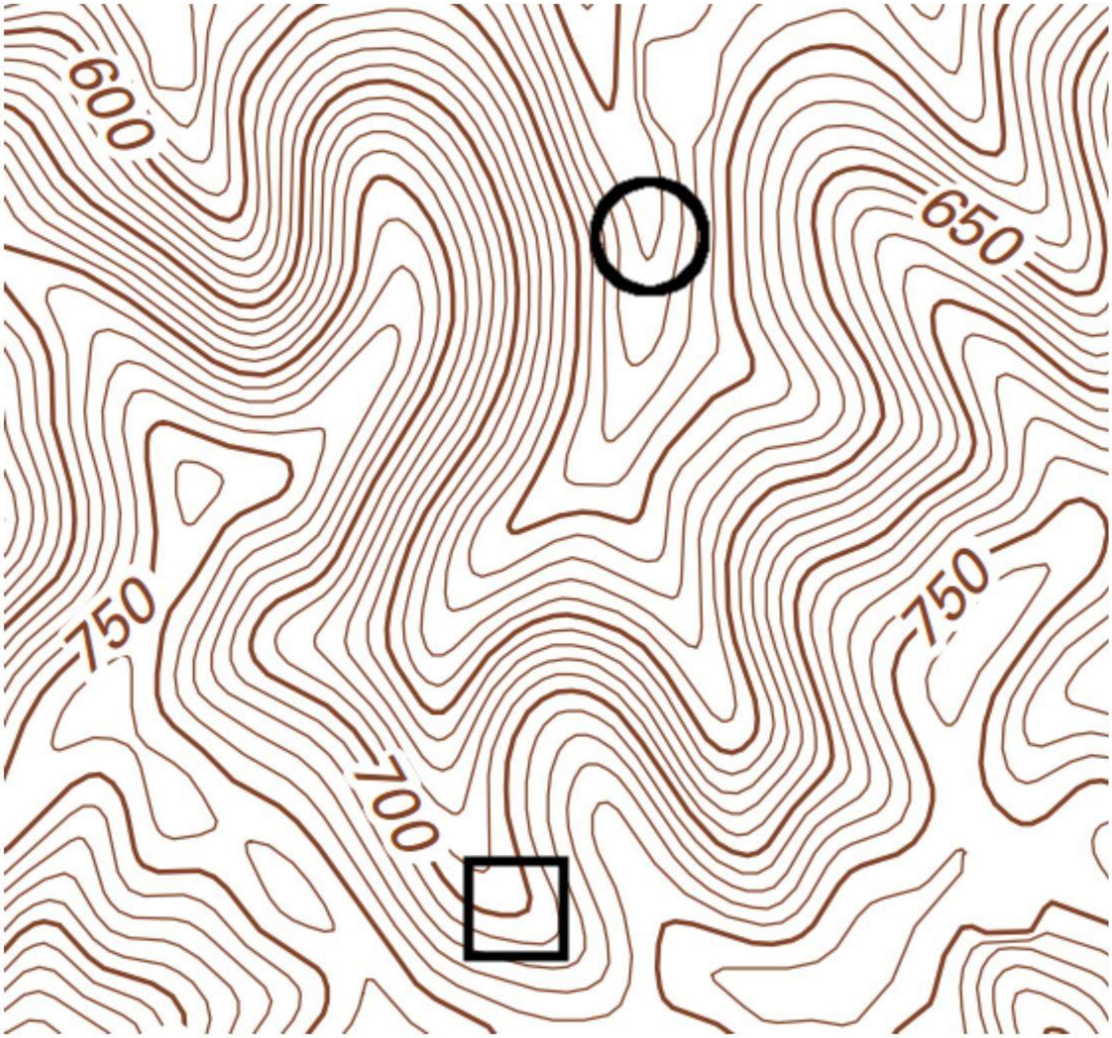
Read the following prompt and follow all directions within, such as sketching and providing explanations when prompted.

You will have <u>3 minutes</u> for the next <u>3 questions</u>. When you have finished, STOP.

28. The contour interval for this map is 100 ft. Imagine you had to walk to get from point A to point B and wanted to do so as easily as possible. Sketch the route you would build and explain why you chose that particular path.



29. Imagine there is a stream that connects the circle and the square. Please draw the path you believe the stream would follow. In addition, clearly mark the direction you believe the water would flow and explain why.



30. Imagine that you are standing at the square, but you want to get to a place (on the map) where you would be able to see a small lake at the circle. Assume there is no vegetation. Please draw a line from the square to another place on the map that indicates a spot where you can see the circle. Explain below, why you chose the spot as well the route to get there.