6-15-2018

Spatial Thinking in Astronomy Education Research

Merryn Cole
University of Nevada - Las Vegas

Cheryl Cole
Governors State University

Jennifer Wilhelm
University of Kentucky, jennifer.wilhelm@uky.edu

Rebecca Lindell
Tiliadal STEM Education Solutions

Spatial thinking in astronomy education research

Merryn Cole
Department of Teaching and Learning, University of Nevada Las Vegas, 4505 S. Maryland Parkway, Las Vegas, Nevada 89154, USA

Cheryl Cohen
Department of Psychology, Governors State University, 1 University Parkway, University Park, Illinois 60484, USA

Jennifer Wilhelm
Department of Science, Technology, Engineering, and Mathematics Education, University of Kentucky, Lexington, Kentucky 40506, USA

Rebecca Lindell
Tiliadal STEM Education Solutions, 5 N 10th St Suite A-1, Lafayette, Indiana 47901, USA

(Received 15 May 2017; published 15 June 2018)

[This paper is part of the Focused Collection on Astronomy Education Research.] Multiple studies show that spatial thinking skills contribute to students' performance in science, technology, engineering, and mathematics disciplines. The study of astronomy is no different with the understanding of many astronomical phenomena requiring spatial thinking skills. This paper describes traditional and contemporary approaches to characterizing and measuring spatial thinking skills and suggests how they inform research in astronomy education. It summarizes previous literature in astronomy education research and categorizes the research approaches of astronomy education peer-reviewed journal articles and conference proceedings that explicitly consider the role of spatial thinking. Additionally, it recommends directions and curricular approaches for astronomy education research informed by current research in spatial thinking.

DOI: 10.1103/PhysRevPhysEducRes.14.010139

I. INTRODUCTION

Imagine a middle school student attempting to understand why the Moon changes shape over the course of a month. To form a coherent scientific explanation of lunar phases, the student must be able to visualize the movement of Earth around the Sun and the Moon around Earth. The student must also understand the relative sizes of Earth, the Sun, and the Moon and the distances between these bodies. Even when the student holds accurate knowledge about the causation of astronomical phenomena, spatial thinking skills are needed for the student to create accurate mental models of complex phenomena that are too vast to see

> *The process of moving from human wonder at the night sky to a scientific understanding of the structure and evolution of the universe is a remarkable study, made possible to a significant degree by insights and inferences generated by spatial thinkers.*
> *National Research Council, Learning to Think Spatially [1].*

Understanding astronomical phenomena requires the ability to imagine objects from different view perspectives and to track the motion of objects in multidimensional space. Astronomy also requires the ability to recognize patterns, to understand cardinal directions, and to reason about external representations of astronomical phenomena, as represented in diagrams, maps, three-dimensional (3D) animations, virtual reality displays, and classroom demonstrations with physical objects. These abilities are examples of spatial thinking skills, which we define as *the perceptual and cognitive processes that enable humans to create and manipulate mental representations of the spatial properties that exist within and between physical or imagined objects, structures and systems*. In addition to the capacity to form internal representations of spatial entities, we propose that spatial thinking comprises the capacity to comprehend external representations (e.g., maps, diagrams, graphs, etc.) of such entities, and to make inferences or solve problems about the spatial properties or internal and

TABLE I. Historical classifications of spatial skill.

	Spatial visualization	Spatial perception	Mental rotation	Perceptual speed	Closure speed; closure flexibility
Linn and Petersen [4]	Complex, multistep manipulations of spatial information.	The ability to determine spatial relations with respect to one's own body.	The ability to rotate two- or and three-dimensional objects rapidly and accurately.	Not defined	Not defined
Carroll [5]	The processes of apprehending, encoding, and mentally manipulating three-dimensional spatial forms.	Not defined	The ability to visualize the rotation of a three-dimensional object.	The ability to quickly compare figures or symbols, or to perform other very simple tasks involving visual perception.	The ability to identify a stimulus that is obscured by visual noise.

external representations of spatial extent. We distinguish the concept of spatial thinking from that of *spatial ability*, which traditionally has been used to refer to the measurement of spatial skill. We note that cognitive science researchers also use the term *spatial cognition* to refer to the research on spatial thinking skills.

More than half a century ago, a National Science Foundation advisory panel published a report recommending strategies for identifying and nurturing scientific talent [2]. By the first decade of the 21st century, there was a convincing body of longitudinal evidence [2,3] that spatial thinking skills measured in adolescence predicted achievement in science, technology, engineering, and mathematics (STEM) occupations in adulthood. In 2006, the National Research Council (NRC) published *Learning to Think Spatially,* a national research agenda for incorporating explicit instruction in spatial thinking in K–12 curricula [1]. Delineating the principal components of spatial thinking, the NRC urged educators to investigate "…concepts of space, tools of representation, and processes of spatial reasoning" (p. 5). The NRC defined *concepts of space* as the properties, such as scale and size, which define the spatial extent of any scientific discipline; a *representation* as an internal or external embodiment of information about an object or system; and *processes of reasoning* as the mental steps required to solve problems in a particular domain [1].

Motivated by the intrinsically spatial nature of astronomy and the call across STEM disciplines to consider spatial thinking as an important cognitive skill, we review a body of peer-reviewed studies that considered the role of spatial thinking in astronomy education. Our goals in this review are to (i) briefly summarize the research frameworks that traditionally have been used to characterize and measure spatial thinking skill in STEM fields and introduce new approaches, (ii) categorize the research approaches of peer-reviewed journal articles and conference proceedings that explicitly consider the role of spatial thinking in astronomy, and (iii) recommend future directions for astronomy education research that investigates the role of spatial thinking in astronomy and propose curricula that reflect these insights.

Contributions to our understanding of spatial thinking skills have emerged from distinct research traditions in psychology and cognitive science, notably developmental psychology, the psychometric approach, and cognitive psychology. These research frameworks have never been connected by unified theory or common methodologies, resulting in a lack of clarity about the structure of spatial thinking skills. Table I illustrates the similarities and differences in the classification of spatial skills based on approaches from developmental and psychometric research traditions.

Despite these differences, all historical approaches to investigating the structure of spatial ability converge on the idea that spatial ability is not a unitary construct, but rather is composed of subcomponents reflecting different mental processes. Proposals for new frameworks to understand the development, structure, malleability, and assessment of spatial thinking have emerged from cognitive science and disciplinary STEM researchers. In the next section, we briefly describe and compare historical approaches to understanding spatial thinking skill and introduce new approaches.

A. Developmental approach to spatial cognition

Pioneered by Jean Piaget (1896–1980), the developmental approach to psychology investigates the origins and maturation of social and cognitive abilities from infancy through childhood. A central challenge of the developmental approach is parsing the contributions of innateness and environmental influence to the development of competencies.

Piaget was the first scientist to study the development of spatial cognition. He and his colleagues built their theories by observing the behavior of infants and children in natural and experimental settings. Observations of crawling infants led Piaget to emphasize the importance of the motor activity in young child's formation of spatial representations of his or

her immediate environment [6]. Piaget and Inhelder proposed a stagelike progression of spatial awareness in children, beginning with the child's ability to understand topological representations, followed by competencies in understanding projective, and finally Euclidean representations.

The stagelike characterization of cognitive development proposed by Piaget has been disputed by subsequent developmentalists. However, the magnitude of Piaget's contribution to developmental psychology as a whole, and to spatial development in particular, is irrefutable. Piaget identified a number of spatial competencies that develop over childhood, including the ability to use categorical (e.g., near and far) and metric spatial representations to describe spatial extent; facility at shifting between *egocentric* (viewer-dependent) and *allocentric* (viewer-independent) frames of visual reference, and skill at using symbolic spatial representations, including, maps, diagrams, and sketches.

Contemporary developmental psychologists have continued to investigate the emergence and development of spatial skills. There is now a robust body of literature investigating precursors to spatial skills, individual and sex differences in spatial development, and the contributions of motor activity, including gesture, to the development of spatial representations [7]. Research on children's development of spatial representations identified a beneficial reciprocal relationship between early spatial skills and mathematics ability. Gunderson *et al.* [8] demonstrated that 1st and 2nd graders' ability to mentally rotate two-dimensional (2D) figures predicted improvement in their ability to create a meaningful representation of a linear number line by the end of the school year. They also found that the spatial skills of five-year olds, as measured on another mental rotation task, predicted their performance on a measure of symbolic calculation ability.

An approach sometimes taken by developmental psychologists is to investigate sex differences in the emergence and durability of behavioral and cognitive competencies over time. Using meta-analysis to synthesize the results of previously studies, Linn and Petersen used computed effect size differences in mean group (male vs female) performance as reported in 172 spatial studies published from 1974 to 1982, and representing participants from preschool to college age [4]. They categorized the spatial skills represented in these studies along three dimensions: *spatial perception*, which they defined as the ability to determine spatial relations with respect to one's own body; *mental rotation*, the ability to visualize the rotation of three-dimensional objects; and *spatial visualization*, complex, multistep manipulations of spatial information. Linn and Petersen used both psychometric (referring to previous psychometric studies) and cognitive (identifying the mental processes hypothesized to contribute to the skill) rationales to arrive at these categories, "…focus (ing) on the similarities in the processes individuals used for individual (test) items" (p. 1482). In classifying these skills, they noted that while spatial visualization tasks might comprise processes of spatial perception and mental rotation, spatial visualization tasks were distinct from spatial perception and mental rotation because of their inherent complexity and their amenability to analytic and imagistic solution strategies. The meta-analysis of sex differences in spatial skill found the largest sex differences in performance on mental rotation tasks.

Comprehensive reviews of recent developmental literature are found in Newcombe and Huttenlocher [9], Newcombe and Frick [10], and Vasilyeva and Lourenco [11]. Summarizing recent research themes related to the development of spatial abilities, Newcombe and Frick [10] argued for the malleability of spatial skill and the importance of early education of spatial thinking skills in formal and informal settings.

The developmentalists' concern with the malleability of spatial skill is echoed in a recent meta-analysis by Uttal *et al.* [12], which found convincing evidence (an average effect size of 0.47 for training effects vs control) for the malleability of spatial thinking through training and instruction. Uttal *et al.*'s findings support those of an earlier meta-analysis of 60 studies that examined the relationship between spatial ability and participation in spatial activities [12,13]. In this study, Baenninger and Newcombe [13] categorized the training studies in the meta-analysis by content of instruction and duration of training. They identified three types of training content: specific training (training on a specific spatial measure); general training (training on more than one type of measure); and indirect training (training that was not related to a specific spatial measure, but was related to a spatial task). The duration of training studies varied from short (single administrations of training, or training that lasted three weeks or less) to medium (more than one administration over more than three weeks, but less than a semester) and long (training that lasted a semester). Baenninger and Newcombe [13] concluded that training was optimal when it was test specific and administered in at least 3 or 4 sessions over 3 weeks or longer. Uttal and Cohen [14] emphasize the importance of foundational training in spatial thinking at the university level, arguing that STEM novices rely to a far greater degree on the processes of spatial reasoning than do expert scientists, who have created heuristics and abstract mental representations to help them solve problems.

B. Psychometric approach to spatial thinking

The psychometric approach to understanding spatial thinking has focused primarily on discovering and describing the factors of spatial ability. The approach originated among early 20th century scientists who challenged the prevailing view that human intelligence was a unitary construct. By the early 20th century, psychometricians, including Thurstone, had successfully applied exploratory factor analysis to distinguish seven separable intelligence factors: word fluency, verbal comprehension, number

facility, reasoning, associative memory, spatial visualization, and perceptual speed [15]. The further application of the factor analytic approach led to the development of hundreds of standardized spatial ability tests [16], many of which were developed to predict vocational abilities.

In the mid-20th century, a number of exploratory factor analyses of spatial test data were conducted to determine the underlying factor structure of the skills measured by psychometric tests [17]. The results of these analyses varied, in part because the factor analyses were influenced by the types of tests that were used. However, each analysis supported the multi-componential nature of spatial ability [17]. Carroll's [5] reanalysis of 90 sets of psychometric data was the most extensive of this group of exploratory factor analyses. Carroll [5] found consistent support for four visuospatial factors: *spatial visualization*, the processes of apprehending, encoding, and mentally manipulating three-dimensional spatial forms; *perceptual speed*, the ability to quickly compare figures or symbols, or to perform other very simple tasks involving visual perception; and *closure speed* and *closure flexibility*, both of which involve the ability to identify a stimulus that is obscured by visual noise. Carroll found the strongest support for the *spatial visualization* factor, which he defined as "power in solving increasingly difficult problems involving spatial forms."

Carroll's typology has been used as a starting point for conducting studies investigating the role of spatial thinking in some disciplines of science [18]. There is no definitive psychometric test to measure spatial visualization. Typical markers for the test include tests of mental rotation [Vandenberg mental rotation test [19] and the Purdue spatial visualization test: Rotations (PSVT:R) [20]] and form board and surface development tests which require the viewer to imagine the folding and unfolding of a pieces of paper [16]. Other tests that have been used to measure spatial visualization ability include Guay's visualization of views test [16], which asks the participants to imagine a view of an imaginary object from a perspective other than the one given in the test problem, the cube comparison test [21], which requires participants to predict visual patterns on the hidden face of a cube, and the spatial relations subtest of the differential aptitude test [22], which requires participants to visualize a two-dimensional figure that has been folded into a three-dimensional shape.

The psychometric approach has made important contributions to our understanding of spatial thinking by distinguishing spatial abilities from other cognitive processes and by determining the componential nature of spatial skill. It has also provided assessment tools that permit cognitive psychologists to examine spatial processes in specific STEM disciplines.

C. Cognitive approach to spatial thinking

When students learn about scientific phenomena that are too small or too vast to see with the naked eye, they use their perceptual and cognitive abilities to form internal mental models and to parse external representations, including diagrams, charts, and graphs describing these phenomena. What are the perceptual and cognitive processes that allow students to form mental representations of astronomical phenomena? To what degree do individuals vary in their abilities to use visuospatial skills? The cognitive approach has focused on identifying and describing the mental processes underlying spatial skill. Psychological theories of working memory and of imagery formation contribute to our understanding of the mental processes that contribute to spatial thinking.

Most research on the cognitive processes involved in spatial thinking has focused on the role of visuospatial information in forming spatial representations. There is also growing interest among cognitive and developmental scientists in the contribution of haptic and kinesthetic information, such as through gesture and the manipulation of physical models to the formation of spatial representations [23].

Baddeley's model of *working memory* [24] is the dominant cognitive theory that describes how humans transform new perceptual information into enduring memories. Baddeley's three-part model comprises a *central executive* component and two "slave" systems: a *visuospatial sketchpad* that processes visually based information and a *phonological loop* that processes auditory information. According to this model, the central executive monitors and schedules the operations of visuospatial working memory and phonological loop. All three components of working memory are conceived of as having limited storage and processing capacity albeit with evidence for individual differences in processing ability.

A complementary body of theory describes the processes underlying the formation and transformation of imagery in the visuospatial sketchpad [25,26]. In Kosslyn's model, visuospatial working memory has the potential to combine perceptual input with previously encoded imagery, to combine multiple images into a composite image, and to transform imagery by a number of spatial processes. Kosslyn identified a number of transformational processes that act on imagery including *translate* (move up-down, right-left, or diagonally in two-dimensional space), *rotate, scan,* and *parse* (rearrange the internal parts of an image). In a complex spatial visualization task, the central executive component of working memory would order the series of rotation and parsing processes in visuospatial working memory order to arrive at a solution to a problem.

As applied to spatial visualization tasks, there is evidence that low-spatial individuals lose spatial information while transforming mental images [27–30]. Individual differences in the ability to change view perspective have been documented by Kozhevnikov and Hegarty [31] and Hegarty and Waller [17]. Current models of individual differences in spatial visualization ability specify differences in working memory resources for the storage

and processing of spatial information [17]. As noted by Linn and Petersen [4] there is also evidence for a male advantage on specific spatial thinking tasks, notably mental rotation. Biological, environmental, and social psychological theories have been proposed to explain these differences. Biological theories include hypotheses proposing that hormonal changes at adolescence [32] and genetic contributions to brain laterality [33] disadvantage females on some spatial skills. Countering this view is substantial evidence that environmental influences, in the form of experience in spatial activities from an early age and explicit training can eliminate sex differences on spatial tasks [4,33].

Social psychologists have proposed that stereotype threat may contribute to a male advantage on some spatial tasks. Stereotype threat is experience of anxiety and accompanying degradation to cognitive processes that occurs when a member of a group is reminded of stigmas about their group [34]. In laboratory settings, the male advantage on mental rotation tasks disappeared when female participants were told that they could expect to perform better than men on spatial tasks [35,36] and when they were reminded of their academic status as students in a private elite university [37].

D. Studies of spatial thinking in other STEM disciplines

Science educators have investigated the role of spatial thinking in their fields for decades, and the NRC's call for a systematic approach to spatial education has stimulated more research. In 2006, the National Science Foundation established the Spatial Intelligence and Learning Center (SILC), a consortium of scientists and educators whose collective aim is to investigate the processes of spatial learning and to use this knowledge to develop programs and technologies that will transform education practice in STEM fields [38]. SILC's website [38] is a resource for information on research, publications, tests, instruments, and general information related to the development of spatial intelligence.

STEM researchers in disciplines outside of astronomy have investigated the role of spatial thinking with a variety of empirical approaches, including observational, correlational, and training studies [39]. Correlational studies have found statistically significant correlations between students' performance on psychometric spatial ability tests and their performance on specific tasks in a number of disciplines.

A spatially demanding task that is common in biology, anatomy, and engineering is the ability to visualize the correspondence between 3D structures and their 2D cross sections [40–42]. Cohen and Hegarty [43] found moderate positive correlations between the ability to draw cross sections of novel anatomical-like forms and measures of mental rotation and perspective-taking ability. Rochford [40] found that students who had difficulty in spatial processes such as sectioning, translating, rotating, and visualizing shapes also had difficulty in practical anatomy classes. University students' spatial ability scores predicted their skill in identifying anatomical structures [41]. Ha and Brown [42] found that performance on a measure of cross-sectioning ability among sophomore civil and aeronautical engineering students accounted for 53% of the variance on a mechanics of materials concept inventory. Mechanics of materials is an engineering topic that requires the ability to visualize and analyze the distribution of stress loads on cross sections of inclined planes.

Geoscience education researchers demonstrated that spatial visualization skills, particularly the ability to identify cross sections of 3D structures, is a required skill in introductory geology courses, and that students' spatial thinking skills increase after a structural geology class [44,45]. Liben, Kastens, and Christensen [46] demonstrated that students' performance on a spatial task was related to their performance on geological field and laboratory tasks. Supporting correlational evidence for the importance of spatial thinking skills in geology, Kastens and Ishikawa [47] enumerated specific spatial thinking skills that are important in geosciences: recognizing, describing, and classifying the shape of an object; describing the position and orientation of objects; making and using maps; envisioning processes in three dimensions; and using spatial-thinking strategies to think about geoscience phenomena. In a similar fashion, Liben and Titus [48] described the spatial thinking skills required in a typical field day in geology, describing some of the cognitive processes required for these tasks and suggesting teaching strategies for particularly demanding tasks.

In chemistry, spatial thinking skills contribute to students' ability to distinguish between isomers (molecules with the same composition, but different structural properties) [14]. Significant positive correlations have been found between psychometric spatial ability and chemistry topics, including topics that are not obviously spatial. A number of researchers found that the relationship between spatial ability and understanding chemistry was stronger for questions that required problem-solving skills, rather than those that could be addressed through memorization [20,49,50].

There is also correlational evidence for the contribution of spatial thinking skills to performance in physics. Kozhevnikov and Thornton [51] found small positive correlations between performance on the paper folding test and problems that require the participant to relate force and motion events, and to interpret graphs representing force and acceleration. Additional correlational evidence for the contribution of spatial skills to physics is found in Hegarty and Sims [52] and Kozhevnikov, Hegarty, and Mayer [53].

A variety of researchers have investigated the benefits of spatial training for STEM learning. Some have taken a domain-specific approach to training, training specific

skills that are important to their disciplines. Other studies have used a domain-general approach, by training general spatial processes, such as mental rotation and change in view perspective. Examples of studies using the domain-specific approach include Brinkmann [54], Lord [55], and Small and Morton [56]. Brinkmann incorporated folding cardboard patterns and wooden models of geometric forms in a self-paced instructional program designed to improve the geometry performance of eighth-grade students. Models were used to demonstrate specific spatial concepts in geometry, such as the characteristics of points, lines, angles, planes, and solids. After instruction, the trained group showed significant pre-post-test gains on a measure of geometry performance and transfer to a spatial visualization test. Similarly, Lord [55] used wooden models of geometric solids to train biology students to recognize the cross sections of primitive figures. Students were encouraged to form visual images of three-dimensional solids, and then to imagine the shape that would be formed when the solids were cut at various angles. In an organic chemistry class, Small and Morton [56] found that experience manipulating 3D molecular models and interpreting diagrams significantly improved the performance of an experimental group, compared to a control group that received extra practice on conceptual knowledge in chemistry.

Support for domain-general training is found in extensive work by Sorby [57], who developed semester-long courses designed for engineering students with low spatial visualization skills. Coursework included lectures, sketching, and manipulating multimedia software that modeled rotations, projections, and cross sections of simple geometric objects. Completion of the course was associated with higher grades in engineering and science classes and retention in the undergraduate program.

In a laboratory study, Feng, Spence, and Pratt [58] demonstrated that 10 hours of experience playing an action-based video game significantly reduced the pretraining male advantage on a mental rotation task and a measure of spatial attention, with females realizing larger gains than males. Additional support for a domain-general approach was provided by Wright *et al.* [59] who found that intensive practice on a computerized version of the Vandenberg mental rotation test (MRT) and the paper-folding test over a three-week period transferred to nonpracticed spatial tasks. Similarly, Sanchez [60] demonstrated that university students who played an action-oriented video game for a short duration (25 min) performed significantly better on spatial ability post-tests than members of the control group, who played a word-based video game during the same period. The most noteworthy result from Sanchez [60], however, was the transfer of first-person shooter game training to performance on an essay demonstrating comprehension of the geologic mechanisms causing volcanoes gained from reading a nonillustrated account of the plate tectonics. The control groups' scores on the volcano essay were significantly lower than those of the experimental group. Notably, the video games used in Feng and Spence [58] and Sanchez [60] were "first-person shooter" games, in which the participant is challenged to rapidly shift his or her view perspective of their target. Sanchez [60] argued that a short period of experience manipulating visuospatial information in a goal-directed manner improved the efficiency of visuospatial processes recruited in learning spatially-rich scientific content as well as performance on spatial ability tests.

Historically, studies investigating the role of spatial abilities in STEM fields outside of astronomy have relied on various psychometric measures of spatial visualization, as defined by Carroll, to investigate spatial thinking in particular disciplines [5]. While these tests are predictive of performance in STEM fields, they do not measure the complex, domain-specific thinking required in disciplines. Researchers who are interested in improving spatial thinking have argued for new typologies that capture the nuanced skills required in specific STEM fields [5,61].

E. New typologies for spatial thinking

Addressing the inherent limitations of psychometric tests and the lack of theoretical consensus for any prior typology of spatial thinking skills, Newcombe and Shipley [62] proposed a top-down schema that varies along two dimensions: the intrinsic-extrinsic location of the spatial features of an *object* (e.g., any entity that exists in the natural world, regardless of its spatial scale) and movement (whether objects are static or dynamic). Newcombe and Shipley [62] found support for this typology from cognitive, neural and linguistic evidence that humans make distinctions along these two continua. Support for the intrinsic-extrinsic distinction in spatial thinking was also found in a self-report measures by scientists in diverse fields.

As shown in Table II, crossing these two dimensions yields four broad categories of spatial skills: *intrinsic-static* skills are needed to code (i.e., form a mental representation of) the internal features of static objects, such as the distance between two locations on Earth; *intrinsic-dynamic* skills are used to transform the internal spatial features of objects, as with imaging the mental rotation or cross sectioning of an object; *extrinsic-static* skills are used to code the spatial location of objects relative to other objects or to a reference frame, as with the size and scale of objects in space; and *extrinsic-dynamic* spatial thinking skills are need to transforming the relations between objects as one or more of them, including the viewer moves. An example of an extrinsic-dynamic astronomical representation is a mental model of the planets revolving around the sun as seem from a view perspective other than Earth.

At this writing, support for this typology was found in two studies that tested the hypothesis that specialized spatial skills were required to perform rigid vs nonrigid transformations of objects. In a rigid transformation, the

TABLE II. Newcombe-Shipley topology of spatial skills [62].

	Intrinsic	Extrinsic
Static	Coding the internal features of static objects	Coding the spatial location of objects relative to other objects or to a reference frame
Dynamic	Transforming the internal spatial features of objects, as through mental rotation, cross sectioning, folding, and other plastic deformations	Transforming the relations between objects as one or more of them moves, including the viewer

distances between points (locations) in an object are preserved. In a nonrigid transformation, the distances between points are not preserved. In Resnick and Shipley [63], expert geologists outperformed expert organic chemists on a task that was required in geology, but not in chemistry—identifying the spatial transformations of a brittle object. Testing nonexpert scientists (undergraduates in a psychology department participant pool), Attit, Shipley, and Tikoff [64] also found dissociations between the ability to make rigid vs nonrigid spatial transformations. Evidence from these two studies not only supported the dimensions of the typology, but also suggested that scientists from different disciplines rely on different spatial thinking skills in their work.

The measurement of spatial skills in the psychometric literature focused on how individuals reason about the spatial properties of objects. Cognitive scientists have also recognized the distinction between spatial thinking abilities at different scales, primarily the scale of objects, vistas (such as rooms where the spatial extent can be seen in a single view) and large-scale environments, whose spatial extent can only be experienced through moving through them [65–67]. Cognitive psychologists have identified normal variation in the spatial thinking skills required to interpret interactive animations [43] and virtual reality displays [66]. This evidence suggests that astronomy educators may find variation among astronomy students in the ability to form and transform visuospatial information.

Evidence for the validity of these new spatial typologies argues for what Resnick and Shipley [63] refer to as an *ecological approach* to spatial cognition, one that recognizes that the spatial skills needed in a specific discipline may differ substantially from those needed in other disciplines.

II. LITERATURE REVIEW: ASTRONOMY EDUCATION AND SPATIAL THINKING

A. Selection criteria

The second goal of our paper was to identify the dominant approaches to research that investigated the contribution of spatial thinking in astronomy education. We limited our selection to peer-reviewed journal articles and conference proceedings that explicitly investigated the role of spatial thinking at any level of astronomy education over the past 35 years. Using these key terms "spatial ability," "spatial reasoning," "spatial thinking," "visuospatial," "astronomy," and "astronomy education," we used Google Scholar, PsychInfo, ProQuest, and individual journal indices (e.g., Astronomy Education Review) to search for relevant papers.

B. Categorization of studies

We modified Hegarty's framework to reflect the current literature that addressed research on spatial thinking in astronomy. We identified three primary approaches to investigating the role of spatial thinking in astronomy education: (i) Noninterventional studies, including descriptive studies of students' misconceptions; (ii) interventional studies designed to *remediate* students' spatial misconceptions in astronomy; (iii) learning progressions, which propose sequences through which students may develop accurate explanations of astronomical phenomena.

In the existing literature, researchers used a variety of tools and methods to investigate the role of spatial thinking in astronomy. Some measured the association between domain-general psychometric tests (e.g., PSVT:R, Vandenberg mental rotation test; the paper folding test) and performance on specific tasks in astronomy (content knowledge). Others devised novel, domain-specific tests of spatial skills in astronomy. We have included two sets of tables to help the reader understand the content of the papers within each framework. First, a set of tables was created that shows the ways in which the papers were categorized according to the modified Hegarty framework. This set of tables includes a summary for each study as well as the participants and instruments reported in each paper. In addition, we have also included a set of tables to note how each article within the intervention, nonintervention, and learning progression categories would be classified according to the Newcombe and Shipley [62] extrinsic-intrinsic dynamic-static framework. Categories were assigned based on how the how the assessment and/or instruction of the concepts were described within the paper. This set of tables also includes columns that denote the assessment instrument(s) used as well as their domain specificity or generality. It is important to note that, at first glance, many of the problems undertaken by astronomy education researchers would fall into the extrinsic-dynamic category as we are considering the spatial relations of objects that are certainly moving and the observer is external to the objects and moving as well. However, it is also important to look at both the phenomena itself and

the instructional methods or tools used to teach it, as this may put the research into another Newcombe and Shipley category. Much science content involves phenomena that are too large or small to be seen by the naked eye, so external visualizations such as models, diagrams, and drawings are used to help students learn about the phenomena. The spatial skills needed to understand external visualizations and models are different from those needed to think about the actual phenomena. For instance, when learning about the motion of the Earth, Moon, and Sun system to produce lunar phases, extrinsic-dynamic spatial skills would be needed to consider the relative locations of each moving body in space to produce the phase seen by observers on Earth. If instead an orrery were used to model the lunar phases, intrinsic-static spatial skills would be needed to consider a stationary model or intrinsic-dynamic spatial skills to considering an orrery in motion.

C. Previous reviews of astronomy education literature

In a review of empirical astronomy education literature [68] and a subsequent resource letter [69] Bailey and Slater summarized and classified more than 100 articles, books, and web-based resources that reflected qualitative, quantitative and mixed methods research studies.

They categorized the sources by topics, such as lunar phases, shape of Earth, diurnal movement, cosmology, and astrophysics. They also summarized astronomy assessments and studies that focused on teachers' understandings of astronomical phenomena. Two of the studies reviewed by Bailey and Slater [68] investigated the role of spatial thinking in astronomy [70,71]. Published shortly after the 2001 inauguration of Astronomy Education Review, the literature review stressed the need for astronomy education research that described students' difficulties and more clearly linked theory to classroom practice.

Lelliot and Rollnick [72] reviewed 103 peer-reviewed astronomy education research papers published from 1974 to 2008. Most of the research identified misconceptions of astronomical phenomena held by students. Eighty percent of the studies investigated how students learned about five "big ideas" in astronomy: conceptions of Earth, gravity, the day-night cycle, the seasons, and the Earth-Sun-Moon System. Other topics were the stars, the Solar System, and concepts of size and distance. The most challenging topics were phases of the Moon, the seasons, and gravity, while content related to Earth and the day-night cycle was well understood.

While noting that researchers' theoretical frameworks were often unstated and hard to classify, Lelliot and Rollnick identified four principal theoretical approaches used by researchers: individual or personal constructivism, including Piagetian theories of conceptual development; investigations of mental models or conceptual frameworks held by participants; studies of conceptual change and knowledge acquisition; and cultural, cross-cultural or worldview perspectives on astronomy. Recognizing the intrinsically spatial nature of astronomy, Lelliot and Rollnick recommended that astronomy curricula at all levels of education increase visuospatial learning activities, such as manipulating physical models and interacting with virtual displays. They also recommended that astronomy curricula include instruction in interpreting external representations, including drawings and models, and discussions of distance and scale in the Solar System.

Brazell and Espinosa [73] found in a meta-analysis of 19 studies that planetaria were effective in improving students' understanding of astronomical phenomena, and in some cases in improving students' spatial thinking skills. However, these results should be interpreted with caution as none of the studies that measured spatial skills had been peer reviewed. Sneider, Bar, and Kavanagh [74] identified more than 40 studies, including studies of planetaria, that explored students' conceptual difficulties in understanding explanations for the seasons. Their review identified the spatial content of challenging content related to the seasons, including the physics of light, the rotation of Earth around the Sun, the Sun's path, and the tilt of Earth, but did not identify the spatial assessments that were used in these studies. They proposed a learning progression for the seasons that is discussed in Sec. II. C.

In summary, none of the previous reviews of astronomy education research explicitly investigated the role of spatial thinking in astronomy research. Bailey and Slater [68,69] summarized the content of existing research in a historical framework and noted the need to translate research findings into classroom practice. Two studies investigating spatial thinking were mentioned in this review, but spatial thinking was not a theme.

Lelliot and Rollnick [72] categorize astronomy education research in a conceptual framework of five "big ideas," each of which is brimming with spatially challenging content. Although Lelliot and Rollnick noted the overall importance of developing teachers' and students' visuospatial abilities, they stopped short of defining the spatial challenges inherent in the big ideas. This review will be the first in the field to address the spatial content of astronomy education research.

D. Review of astronomy education literature investigating spatial thinking

1. Noninterventional studies

Noninterventional studies investigate the relationship between students' misconceptions of astronomical phenomenon and spatial reasoning. See Tables III and IV for summaries of each of the research studies and how they fit within the Newcombe-Shipley framework [62]. Many research studies have indicated that students need to develop keen spatial understandings to grasp comprehension of apparent celestial motions, the cause of lunar phases, and an explanation of the seasons [75–78]. A number of studies have demonstrated that children younger

TABLE III. Summary table for noninterventional studies. Note that effect sizes are included if available.

References	Participants	Measures	Findings
Plummer, Bower, and Liben [75]	Children aged 7–9 attending a summer astronomy camp at a planetarium ($N = 15$)	Clinical interviews, perspective taking test	Children were interviewed and asked to complete a written perspective taking test prior to and after a weeklong camp. Interview responses were coded for accuracy of responses, gestures used, and reference frames. Logistic and linear regression were used to examine the relationships between these codes and children's perspective taking ability. Children with better perspective taking skills were more explicit in connecting frames of reference than students with lower perspective taking ability. Children with higher perspective taking ability also used certain kinds of gestures more often for some phenomena. Effect sizes (adjusted R^2) range from 0.00 to 0.37.
Sherrod and Wilhelm [76]	Seventh grade students ($N = 92$)	Sketches, transcribed lesson videos, selected questions from the Lunar Phases Concept Inventory	Students experienced an inquiry-based unit on moon phases. Prior to and after the unit, students completed the Lunar Phases Concept Inventory. After experiencing a lunar modeling lesson, one of the authors facilitated a discussion on the 2D drawings of a first quarter moon phase. Classroom dialogue positively influenced students' learning. Other students critiquing and asking questions about the drawings allowed students time to consider and adjust their models of the Earth-Moon-Sun system.
Wilhelm [77]	Middle school students ($N = 123$)	Lunar Phases Concept Inventory, geometric spatial assessment	Gender difference in middle school students' understanding of lunar phases was investigated. Students completed assessments prior to and after experiencing the REAL curriculum, which included modeling, observing, and journaling. Students made significant pre to post gains in understanding as shown on the Lunar Phases Concept Inventory (Partial $\eta^2 = .703$) overall and for both boys (Partial $\eta^2 = .751$) and girls (Partial $\eta^2 = .667$). Students also made significant gains on the Geometric Spatial Assessment overall (Partial $\eta^2 = .151$) as well as for both boys (Partial $\eta^2 = .127$) and girls (Partial $\eta^2 = .170$).
Wilhelm [78]	Two six year old girls, one eight year old girl	Interviews	Clinical Piagetian interviews were conducted with three young children regarding their understanding of the moon's appearance. One of the questions asked of the children regarded the distance between us (Earth) and the moon. Children said the moon was "far away." Another question asked the children how large the moon was. Answers ranged from about 2 inches to as large as their room.
Venville, Louisell, and Wilhelm [79]	Children (ages 3–8) and their parents from the United States and Australia ($N = 10$)	Interviews	Ten children from Australia and the United States were interviewed about their ideas about the Moon using a detailed, semi-structured interview protocol. The resulting data were analyzed using an interpretive paradigm. The findings represent a set of "knowledge nodes" including (1) children's ontological ideas about the moon, (2) animistic or non-animistic attributes of the moon, (3) permanence of the Moon in the sky, and (4) creatures and artifacts children associate with the moon. The authors note the children's ideas fit neither a consistent framework of knowledge or a knowledge in pieces perspective. Also, the children's ideas often correlated with their environment or previous social or cultural activities.
Subramaniam and Padalkar [80]	Eight people with a background in either architecture or physics	Written questionnaire about explaining lunar phases followed by an interview	Participants first completed questionnaires about the cause of lunar phases and then participated in interviews. The interviewers probed participants' written explanations of lunar phases. The interviews focused on the mechanism that causes lunar phases and the shapes of lunar phases. Only one person provided a correct explanation for the phases of the moon on the questionnaire; the most common incorrect explanation involved eclipses. Participants struggled with drawing phases of the moon, with their shapes frequently inconsistent with their explanations for the mechanism causing lunar phases. The authors also found that providing anchoring situations that were analogous to the Earth/Moon/Sun system helped participants correct misconceptions and provide correct explanations for lunar phases. The authors suggest visualization is an important part of learning science.

(Table continued)

TABLE III. *(Continued)*

References	Participants	Measures	Findings
Kikas [81]	First and second grade students ($N = 176$)	Contour Extraction Test, Moomies and Hands Tests, Memory for Objects, verbal abilities tests, astronomy knowledge questions, interviews	Children were interviewed in both first and second grade. In between, they learned about the astronomy topics in focus. Each of the children were interviewed individually. Students' visuospatial ability affected their factual knowledge positively in the first grade, but negatively affected synthetic knowledge in the second grade. Kikas suggested children with higher visuo-spatial abilities may pay more attention to drawings in popular science books and ask questions to better understand them. Kikas also found that verbal abilities had a significant effect on second graders' scientific knowledge.
Wilhelm, Jackson, Sullivan, and Wilhelm [83]	Sixth grade students ($N = 194$)	Lunar Phases Concept Inventory, Geometric Spatial Assessment, Purdue Spatial Visualization Test Rotations	Students experienced either a business as usual unit or an integrated, mathematics science inquiry unit on lunar phases and related concepts. Assessments were administered prior to and after unit implementation. Experimental (Partial $\eta^2 = 0.37$) and control (Partial $\eta^2 = 0.48$) groups made significant pre to post gains in different areas on the Lunar Phases Concept Inventory. The Experimental groups made gains on periodicity while the control group made gains on geometric spatial visualization. Only girls made gains on the geometric spatial assessment. All groups, except experimental males, made significant gains on the Purdue Spatial Visualization Test Rotations.
Black [86]	College students enrolled in nonmajors science courses ($N = 97$)	Earth science concepts test, Purdue spatial visualization tests rotations, group embedded figures test, differential aptitudes test	The contribution of spatial skills to understanding of earth science concepts was investigated in a nonmajors science courses. Correlations and stepwise regression were used to assess the role of spatial ability in understanding earth science concepts. Spatial ability, as measured on the three tests, accounted for up to one-third of the variation on the earth science concepts test. Significant, moderate, positive correlations were found between the scores on the earth science concepts test and each of the spatial assessments.
Heyer, Slater, and Slater [87]	Students in an undergraduate introductory astronomy course ($N = 154$)	Test of astronomy standards, what do you know, Vandenberg mental rotation test, paper folding test	Students in an undergraduate introductory astronomy course were administered content assessments prior to and after the course. They completed a spatial assessment midcourse, and a subset of students ($N = 14$) completed a think-aloud exercise to validate responses after the postassessments. Moderate to strong correlations were found between pre-to-post gain scores on TOAST and the spatial assessment, suggesting the relationship between spatial thinking and understanding astronomy could explain about 25% of the variation in student achievement. They also found that students left the course still unable to correctly answer one-third or more of the TOAST questions.
Türk [88]	Preservice science teachers ($N = 280$)	Astronomy achievement test (AAT), Astronomy Attitude Scale (AAS), Purdue Spatial Visualization Test Rotations (PSVT:R)	Astronomy content, astronomy attitude, and mental rotation tests were given to preservice teachers in Turkey to investigate whether performance on these assessments were different based on years of study as well as whether there was a correlation between any of the assessments. Significant differences in spatial thinking ability were found between senior and freshmen and sophomores, in favor of seniors. Students in their first three years of university education had a similar understanding of astronomy; however senior students' knowledge of astronomy was significantly higher than those with 1–3 years of university education combined. Similarly senior students' attitudes toward astronomy were significantly higher than the attitudes of students at all other levels of education. Turk found a significant, positive correlation between scores on the AAT and PSVT:R as well as a low positive correlation between the AAS and AAT. (Cohen's f was calculated as .200 for PSVT:R, .175 for AAS, and .172 for AAT).

(Table continued)

TABLE III. *(Continued)*

References	Participants	Measures	Findings
Rudman [94]	Undergraduates ($N = 74$)	Cube Comparison Test, Astronomy Geometry Test	Rudman investigated if college students' spatial thinking skills could interfere with their correct causal knowledge of astronomy when they attempted to solve astronomy problems. During the interview, participants gave causal explanations for the basic movements in the Solar System, the day/night cycle, seasons, phases of the Moon and eclipses and were then asked to rate their levels of certainty about the accuracy of their scientific explanations. Following this self-assessment, they were asked to solve problems related to the phenomena they had just explained. Spatial ability shared a positive (though not statistically significant) correlation with astronomy problem solving ability, regardless of the causal models that individuals adopted. Rudman identified four explanatory models of the Seasons: Fixed Tilt (the Earth has a fixed tilt as it rotates around the Sun), Wobbly Tilt (the tilt of the Earth changes as it rotates around the Sun), Elliptical Orbit (the Earth moves in an elliptical orbit around the Sun, causing changes in temperature throughout the year) and Quantum Orbit (the Earth's orbit is closer to the Sun in the summer than in the winter; this model was held by only one of 18 participants.) Performance on the Cube Comparison test predicted the explanatory model held by the students. Students who held the Fixed Tilt model scored higher on the Cube Comparison Test than students who used the Wobbly Tilt model, followed by students who used the Elliptical Model (the level of statistical significance of this prediction is not clearly stated in the article.)
Taylor and Grundstrom [96]	Sixth grade students ($N = 35$)	Textbook images, web images, student drawings	The authors compared diagrams of the Earth-Moon-Sun system in science education materials online, in science textbooks, and produced by students. The size ratios and distances between celestial bodies were compared, finding that none of the sources were accurate. They suggest that warnings of the inaccuracy of the scale be included in drawings and that astronomy instructors use physical models in addition to drawings. In all three sources of imagery, the median ratio of Earth-Moon size was significantly lower than it actually is. Sixth graders visualized the Moon as twice as large as it actually is, and textbook images depicted the Moon as 1.5 times its actual size. Web depictions of Earth-Moon scale were closer to the true ratio, but still significantly lower. The depiction of the Earth-Moon distance was also inaccurate. Students depicted the Moon as 16 times closer to the Earth than it is. The median distance depicted in textbook images was 23 times closer than it is. The median distance between the Earth and the Moon depicted in web pages was 14 times closer than actuality.

TABLE IV. Newcombe-Shipley [64] categories for noninterventional studies. Note that while the domain specificity or generality and assessment instrument columns are intended to align, any alignment with the Newcombe-Shipley framework column is incidental.

References	Concept(s) investigated	Assessment instrument(s)	Domain-specificity or generality of instrument(s)	Shipley-Newcombe framework [64]
Plummer, Bower, and Liben [75]	Apparent celestial motion	Interviews	Domain specific	Extrinsic dynamic
		Perspective taking test	Domain general	
	Seasons			
Sherrod and Wilhelm [76]	Lunar phases	Analysis of verbal discourse	Domain specific	Extrinsic dynamic
				Intrinsic static
		2D drawings of Earth-Moon-Sun system	Domain specific	
Wilhelm [77]	Lunar phases	Lunar Phases Concept Inventory	Domain specific	Extrinsic dynamic
		Geometric spatial assessment	Domain general	
Wilhelm [78]	Location, size, and appearance of the Moon	Interviews	Domain specific	Extrinsic dynamic
Venville, Louisell, and Wilhelm [79]	Childrens' notions about the Moon	Interviews	Domain specific	Intrinsic static
				Intrinsic dynamic
				Extrinsic dynamic
Subramaniam and Padalkar [80]	Lunar phases	Interviews	Domain specific	Extrinsic dynamic
				Intrinsic dynamic
Kikas [81]	Earth properties or motions; solar system; seasons	Contour extraction test	Domain general	Intrinsic static
		Test of visual memory for objects	Domain general	Intrinsic dynamic
		Two mental rotation tests	Domain general	
		Interviews	Domain specific	
Wilhelm, Jackson, Sullivan, and Wilhelm [83]	Lunar phases	Purdue spatial visualization test: Rotations	Domain general	Extrinsic dynamic
				Intrinsic static
		GSA	Domain general	Intrinsic dynamic
		LPCI	Domain specific	
Black [86]	Wide range of Earth science concepts	Earth science concepts test	Domain specific	Extrinsic dynamic
		PSVT:R	Domain general	Extrinsic static
		Group embedded figures	Domain general	Intrinsic dynamic
		Differential aptitudes	Domain general	Intrinsic static
Heyer, Slater, and Slater [87]	Range of Astronomy topics typically encountered in ASTRO 101 courses	Test of Astronomy Standards (TOAST)	Domain specific	Extrinsic dynamic
		What do you know (WDYK)	Domain specific	Extrinsic static
		Vandenberg mental rotation test	Domain general	Intrinsic static
		Paper folding test-Vz-2	Domain general	
Türk [88]	Range of basic astronomy concepts encountered by pre-service teachers	Astronomy achievement test	Domain specific	Extrinsic dynamic
		PSVT:R	Domain general	Intrinsic static
		Astronomy attitude scale	Not relevant to spatial issues	
Rudman [94]	Basic motion in solar system; Day-night cycle; seasons; phases of the Moon; eclipses	Astronomy-based geometry test	Domain specific	Extrinsic dynamic
		Cube comparison	Domain general	Intrinsic dynamic
		Astronomy questionnaire	Domain specific	
		Interviews	Domain specific	
Taylor and Grundstrom [96]	Earth-Moon-Sun system scale and distance	Student-drawn diagrams	Domain specific	Intrinsic static
				Extrinsic static

than nine have limited ability to change their view perspective and little to no understanding of the causes of astronomical phenomena [75,78,79].

There is also evidence that young children's understanding of astronomical understanding (e.g., knowledge of the Moon and stars) can be influenced by their social and cultural milieu (stories, family experiences, etc.). In addition, many children had little to no experiences with focused celestial observations. Wilhelm [78] determined that not only were young children influenced by stories and movies to understand astronomical phenomena, but also had a natural inclination to animate celestial objects such as the Sun and Moon. For example, a child in this study explained that the Moon was sometimes half (first quarter Moon phase) because it (the Moon) "wants to be half" and sometimes when it was happy it looked fuller (waxing gibbous) [78], (p. 263).

Research conducted with older children (aged 12–14) regarding spatial visualization and the cause of lunar phases understanding illustrates that children's spatial thinking skills improve as they gain more experience understanding perspective, directional space, and the necessary geometries for particular lunar phases [77]. Children in this age group also begin to advance their scientific reasoning through work with 2D and 3D modeling [76,80]. Sherrod and Wilhelm [76] conducted a study with 92 middle school students where classroom dialogue was examined regarding a Moon finale lesson using 2D and 3D models to investigate the cause of lunar phases. Students reconstructed their understanding of lunar concepts related to geometric Earth-Moon-Sun configurations after productive classroom discourse that allowed them to consider and challenge their misconceptions. Subramaniaum and Padalkar [80] explored how eight students reasoned with models to explain Moon phases. In this study, students had correct mental models of the Earth-Moon-Sun system, but were unable to explain each phase scientifically. Subramaniaum and Padalkar claimed, "In order to successfully explain lunar phases one needs to shift perspectives as one reasons, from a space based to an Earth based viewpoint" and that conceptual elements such as illumination boundary (terminator line) "belong to the domain of geometry of the sphere" [80], (p. 19).

Plummer, Bower, and Liben [75] developed a novel instrument for investigating elementary school children's ability to shift their imagined view perspective from one reference frame to another. Children who were skilled at changing their view perspective provided more coherent explanations of the relationships between Earth-bound and space-bound frames of reference when explaining the apparent motion of the Sun and the stars, and for seasonal changes in the constellations. The findings suggest that children with lower perspective taking skills may need support in learning to explicitly connect reference frames.

Kikas [81] examined the role of spatial ability and verbal ability in young children's (176 first and second grade students) understanding of knowledge of Earth. Kikas measured students' spatial thinking skills with four tests: the Contour Extraction test [82] in which participants are asked to identify a specified shape in a complex image, a test of visual memory for objects, and two mental rotation tests. Verbal reasoning and students' memory for words and sentences were also assessed. Students' visuospatial ability affected their factual knowledge positively in the first grade, but negatively affected synthetic knowledge in the second grade. Kikas suggested children with higher visuospatial abilities may pay more attention to drawings in popular science books and ask questions to better understand them. Kikas also found that verbal abilities had a significant effect on second graders' scientific knowledge.

In a study with children (aged 12–13 years), Wilhelm *et al.* [83] examined the associations between students' spatial thinking ability and lunar-related content knowledge. Student's spatial thinking abilities were measured with the PSVT:R [20] and the geometric spatial assessment (GSA) [84], a 16-item multiple-choice test that assesses four spatial domains (periodic patterns, geometric spatial visualization, cardinal directions, and spatial projections). In addition, students completed the Lunar Phases Concept Inventory (LPCI) [85], a 20-item multiple-choice survey that assesses students' understanding of lunar-phase concepts such as Moon motion, orbital periodicity, and cause of phases. Questions on the LPCI can be mapped to the four spatial domains of the GSA (See Table V). Significant

TABLE V. Concept domains: Lunar Phases Concept Inventory (LPCI) science domains and corresponding Geometric Spatial Assessment (GSA) domains [83].

LPCI scientific domains			Corresponding GSA spatial domains
A—Period of the Moon's orbit around Earth	B—Period of the Moon's cycle of phases		Periodic Patterns (PP)
C—Direction of the Moon's orbit around Earth	E—Phase due to Sun-Earth-Moon positions	G—Cause of lunar phases	Geometric Spatial Visualization (GSV)
D—Moon motion from earthly perspective	F—Phase location in sky-time of observation		Cardinal Directions (CD)
H—Effect of lunar phase with change in earthly location			Spatial Projection (SP)

positive correlations were found between the PSVT:R and the GSA, and between the LPCI and the GSA.

A number of studies have found significant correlations between mental rotation ability and comprehension of astronomy concepts among college students [86–88]. Black [86] investigated the contribution of nonscience majors' spatial skills to their understanding of earth science concepts, including astronomy. Students' understanding of earth science concepts were assessed with a multiple-choice assessment developed by Black. Students' spatial skills were assessed with the PSVT:R [20], the group embedded figures test, which assesses the ability to disembed a shape from background noise [89], and the Differential Aptitudes Test: Spatial Relations [22]. Spatial ability, as measured on the three tests, accounted for up to one-third of the variation on the earth science concepts test.

Heyer, Slater, and Slater [87] examined the relationship between nonscience majors' spatial thinking and understanding of astronomy. They used the Test of Astronomy Standards (TOAST) [90] and What Do You Know (WDYK) [91–93] to measure astronomy understanding. Spatial thinking skills were measured in a two-part spatial reasoning assessment using questions drawn from the Vandenberg mental rotation test [19] and the paper folding test-Vz-2 [21]. Moderate to strong correlations were found between pre-to-post gain scores on TOAST and the spatial assessment, suggesting the relationship between spatial thinking and understanding astronomy could explain about 25% of the variation in student achievement. They also found that students left the course still unable to correctly answer one-third or more of the TOAST questions.

Rudman [94] investigated if college students' spatial thinking skills could interfere with their correct causal knowledge of astronomy when they attempted to solve astronomy problems. In this exploratory study, he analyzed students' performance on two spatial measurements, a short-answer questionnaire of astronomy knowledge and astronomical problem solving, and structured one-on-one interviews. Spatial thinking skills were measured with the cube comparison test [21] and a 21-item astronomy geometry (AG) test that was designed for this study and tested students' ability to solve problems related to rotation, revolution, occlusion, tilt, light, and a combination of those phenomena. During the interview, participants gave causal explanations for the basic movements in the Solar System, the day-night cycle, seasons, phases of the Moon, and eclipses and were then asked to rate their levels of certainty about the accuracy of their scientific explanations. Following this self-assessment, they were asked to solve problems related to the phenomena they had just explained.

Rudman found that spatial ability shared a positive (though not statistically significant) correlation with astronomy problem solving ability, regardless of the causal models that individuals adopted. From coded interview content, Rudman identified four explanatory models of the seasons: fixed tilt (Earth has a fixed tilt as it rotates around the Sun), wobbly tilt (the tilt of Earth changes as it rotates around the Sun), elliptical orbit (Earth moves in an elliptical orbit around the Sun, causing changes in temperature throughout the year) and quantum orbit (Earth's orbit is closer to the Sun in the summer than in the winter; this model was held by only one of 18 participants). Performance on the cube comparison test predicted the explanatory model held by the students. Students who held the fixed tilt model scored higher on the cube comparison test than students who used the wobbly tilt model, followed by students who used the elliptical model (the level of statistical significance of this prediction is not clearly stated in the article).

The spatial skills, astronomy knowledge, attitudes toward science, and mental models of pre-service science teachers are of interest to astronomy education researchers. Türk [88] investigated spatial thinking skills, understanding of and attitudes toward astronomy of 280 (male = 121; female = 159) preservice science teachers who had from 1–4 years of university education. There was an equal distribution of participants, and a near-equal distribution of sex across years of instruction. Astronomy knowledge was measured using the astronomy achievement test (AAT), attitudes toward astronomy were measured using the astronomy attitude scale (AAS), and the PSVT:R [20] was used to assess spatial thinking. Significant differences in spatial thinking ability were also found between seniors and freshmen and sophomores, in favor of seniors. Students in their first three years of university education had a similar understanding of astronomy; however, senior students' knowledge of astronomy was significantly higher than those with 1–3 years of university education combined. Similarly, senior students' attitudes toward astronomy were significantly higher than the attitudes of students at all other levels of education. Türk found a significant, positive correlation between scores on the AAT and PSVT:R as well as a low positive correlation between the AAS and AAT.

Heywood, Parker, and Rowlands [95] observed how preservice science teachers developed mental models of the Earth-Sun-Moon system. Participants were 26 female students who participated in five 3-h sessions of instruction during their third year of a four-year science education degree program. At specified times during the course, students were asked to draw and annotate diagrams illustrating the day-night cycle and the shape and direction of the Sun's path in relation to the horizon. A smaller number of participants were interviewed for 40 min after the instruction ended. From these two sources of data, the researchers identified four types of models held by participants: the real perception model (RPM), which used participants' personal perceptions as the only source of explanation; and imagined position model (IPM) model, which reflected participants' ability to switch between Earth and Sun perspectives of the Earth-Sun system; the Sun-Earth model (SEM), which

integrates chronological motion into a three-dimensional view of Earth and the Sun from a remote position in space; and the light traveling model (LTM) that incorporates reasoning about the direction with which light travels from the Sun to reach Earth. The authors noted that the models were not mutually exclusive but represented the most frequently employed and identifiable models expressed by the participants. The researchers gave a detailed account of the evolution of the mental models of three students, noting the points at which students were stymied by or successful at reconciling the RPM model with the SEM model. The researchers also mentioned that individual students' use of gesture during individual interviews supported their efforts to integrate multiple view perspectives and to integrate models of chronological motion.

Diagrams of astronomical phenomena are an essential element of astronomy curricula. Taylor and Grundstrom [96] investigated the accuracy of two spatial parameters (scale and distance) in diagrams of the Earth-Moon system collected from three sources: 6th grade students ($n = 35$); educational and governmental websites ($n = 44$); and eight middle school textbooks produced by four publishers ($n = 30$). In all three sources of imagery, the median ratio of Earth-Moon size was significantly lower than it actually is. Sixth graders visualized the Moon as twice as large as it actually is, and textbook images depicted the Moon as 1.5 times its actual size. Web depictions of the Earth-Moon scale were closer to the true ratio, but still significantly lower. The depiction of the Earth-Moon distance was also inaccurate. Students depicted the Moon as 16 times closer to Earth than it is. The median distance depicted in textbook images was 23 times closer than it is. The median distance between Earth and the Moon depicted in web pages was 14 times closer than actuality. The authors recommended that textbook publishers attempt to more accurately portray scale and distance in diagrams, and to add warnings when an image was not drawn to scale. The authors also recommended that astronomy instructors use physical models to illustrate the correct scale and size of the Earth-Moon system and suggested further investigations on the use of textbook and web diagrams that allow for multiple levels of magnification.

The noninterventional studies used a mix of domain-general and domain-specific assessments and well as domain-specific interviews to assess the content understanding and spatial thinking skills of students from early childhood through college. In most cases, the spatial thinking skills were measured using domain-general, psychometric tests while the content knowledge was measured using domain-specific content assessments (interviews or written tests). The exceptions were studies by Wilhelm *et al.* [77,83], where spatial domains were mapped onto a domain-specific content test.

2. Interventional studies

Interventional studies investigate the effectiveness of classroom instruction on understanding of astronomical phenomenon and spatial reasoning. See Tables VI and VII for summaries of each of the research studies and how they fit within the Newcombe-Shipley framework [62]. In many of these studies, the researcher assessed the effectiveness of instruction on spatial thinking skills, or the mastery of astronomical concepts. Multiple astronomy educators' research focuses on how the comparison of 2D and 3D visualizations affects students' spatial thinking skills. These studies focused on different population groups: middle school, high school, and college Astronomy 101 students within a course, as well as those who attended museums.

In other science and mathematical domains, researchers have taken a domain-general approach to training spatial skills prior to content instruction and then assessing whether this domain-general training approach facilitated learning of the subject matter. Within the astronomy education literature, where the reliance is often instead on either domain-specific training of spatial skills or correlational studies.

Plummer, Kocareli, and Slagle [97] interviewed 8–9 year-old students before and after they participated in classroom-based and/or planetarium-based instruction on daily celestial motion. During the interviews, students were asked to use a flashlight and a small planetarium dome to explain apparent celestial motion. The students then explained their demonstrations using models of Earth, the Moon, and the Sun. Plummer *et al.* [97] found students improved their perspective-taking ability as a result of instruction, "going beyond just taking someone else's perspective, by moving their own perspective out into the solar system" (p. 1100). Students in this study improved their understanding of daily celestial motion when they experienced instruction "that supported their ability to visualize Earth-based observations and develop explanations by engaging in multiple modalities: observe visual simulations, engage in guided gesturing, and participate in kinesthetic and psychomotor modelling" (p. 1101). They found instruction needs to address both Earth-based and space-based perspectives, as focusing on only one perspective resulted in students providing less developed explanations of daily celestial motion.

Plummer, Wasko, and Slagle [98] interviewed third grade students ($n = 24$) about the daily apparent motions of the Sun, Moon, and stars. They found that half of the students held naive mental models and the other half of the students could explain the apparent motion of the Sun but struggled with the apparent motion of the Moon and stars. Plummer, Wasko, and Slagle [98] also described an instructional approach using computer simulations and hands-on modeling to support students moving between Earth-centered and Sun-centered frames of reference. Pre- and postinterviews were used with the group of students

TABLE VI. Summary table for interventional studies. Note that effect sizes are included if available.

References	Participants	Measures	Findings
Plummer, Kocareli, and Slagle [97]	Third grade students ($N = 99$)	Interviews	Students were interviewed prior to and after experiencing one of four instructional conditions. The conditions emphasize (1) space-based perspective, (2) Earth-based perspective with a planetarium, (3) developing explanations for the Earth-based perspective, (4) and a combination of constructing explanations and the planetarium. Instruction that included both space-based and Earth-based perspectives was better for helping students develop sophisticated, scientifically accurate explanations for daily celestial motion, but not all students were able to develop adequate explanations after instruction. Instruction should include both Earth-based and space-based perspectives for students to learn about daily celestial motion, but some students are unable to fully develop scientifically accurate explanations even with good instruction.
Plummer, Wasko, and Slagle [98]	Third grade students ($N = 24$) and third grade students from a gifted program ($N = 18$)	Interviews	Third grade students were interviewed prior to astronomy instruction. Findings show that about half of the students were working from naïve mental models, while the other half could use scientifically accurate explanations of the Sun's apparent motion but were less likely able to explain the apparent motion of the Moon and stars. An intervention with gifted third grade students designed to support students' learning to shift frames of reference between an Earth-based and space-based perspective showed promise for helping students learn to provide scientifically accurate explanations for celestial motion; more of the students could provide sophisticated explanations after instructions than prior to instruction. In midelementary school, most students think the observed motions of celestial objects are the actual motions of the objects rather than the apparent motions from an Earth-based perspective. Some children recognize these motions are apparent motions but cannot explain the causes. A few students can provide scientifically accurate explanations for the Sun's apparent motion but struggle in doing the same for the Moon or stars. The authors suggest the mismatch between how students describe apparent daily celestial motion and their explanations may be due to differences in spatial ability. Students need to be supported in learning to shift between frames of reference to develop understanding of daily celestial motion.
Wilhelm, Toland, and Cole [99]	Sixth grade students ($N = 468$) from 8 teachers' classrooms	Lunar Phases Concept Inventory, Purdue spatial visualization test: Rotations	Students completed assessments on lunar phases and spatial thinking prior to and after experiencing an astronomy unit related to moon phases. Multilevel modeling was used to identify how curricular experiences (i.e., experimental versus control groups), gender, and/or race or ethnicity affected spatial-scientific learning. Results showed differences between gender and race or ethnicity groups for some spatial domains and between treatment groups (i.e., experimental and control). Gender and race or ethnicity were significant predictors of LPCI post-test scores; boys tended to have higher scores than girls, and students of color tended to have lower scores than white students. Within the periodic patterns domain, the differences between experimental and control groups depended on gender of the student. For the geometric spatial visualization domain, only gender (boys tended to do better than girls) and pretest score were predictors of post-test score. Only the pretest scores were significant predictors for the spatial projection domain or the Purdue spatial visualization test: Rotations. Effect sizes ranged from 0 to .68.

(Table continued)

TABLE VI. *(Continued)*

References	Participants	Measures	Findings
Jackson, Wilhelm, Lamar, and Cole [100]	Sixth grade students ($N = 2021$)	Lunar Phases Concept Inventory, geometric spatial assessment, Purdue spatial visualization test: Rotations	Students' geometric spatial development within an Earth-space unit was examined by gender and race groups and by control and experimental groups. Prior to and after instruction, students completed the Lunar Phases Concept Inventory, geometric spatial assessment, and Purdue spatial visualization test: Rotations to assess understanding of lunar phases and spatial thinking. All students within the experimental group made significant gains in understanding geometric spatial visualization while within the control group the gains in understanding were only made by white students and boys. The results indicate that support is needed to all groups to develop their geometric spatial visualization and in turn scientific understanding. Effect sized ranged from 0.20 to 0.50.
Cole, Wilhelm, and Yang [101]	Sixth grade students ($N = 333$)	Lunar Phases Concept Inventory, teacher survey, daily moon observation journals	Students kept daily moon observation journals for about a month during a unit on lunar phases. The authors used regression analysis to examine the relationship between moon journaling and what students learned about lunar phases. The results show that students who kept moon journals did better on post assessments, both in terms of content and spatial thinking, than students who put less effort into moon journaling. Students who put more effort into daily moon observation journals (as evidenced by total journal rubric score or by number of entries) scored better on the post test than students who put less effort into moon journaling. These results held true for the overall LPCI post-test scores as well as for two spatial domains: Periodic patterns and geometric spatial visualization.
Padalkar and Ramadas [102]	Seventh grade students in India ($N = 64$)	Student drawings	Students experienced an author-designed pedagogical sequence intended to help students develop accurate mental models of the Earth and its motions. Models, gestures, and diagrams were used during the teaching sequence and students were asked to draw as part of their own explanations. Students initially were hesitant to draw, but through the intervention students became more comfortable creating and sharing drawings as part of their explanations. Students could provide adequate drawings and explanations for many parts of their Earth model but struggled with using parallel sun's rays.
Price and Lee [103]	Middle school students ($N = 19$)	Questions adapted from letter rotation task, block rotation task, paper folding task. Interviews.	Students performed spatial cognition tasks in either 2D or 3D. The tasks were focused on letter rotation, block rotation, or paper folding. Students first completed the 2D spatial tasks on paper, then they performed the 3D spatial tasks using a GeoWall. Response accuracy and completion time were measured for each kind of task. Students' performance on spatial tasks was compared when tasks were presented as 2D or 3D images. Response accuracy did not differ between the two conditions, but time to complete the tasks did. 3D tasks that required more manipulation took longer than similar tasks presented in 2D. The authors suggest students need to be more time to become familiar with 3D representations.
Cid and Lopez [104]	Students in introductory college astronomy course ($N = 170$)	Lunar Phases Concept Inventory	Students participated in a laboratory on phases of the moon in an introductory college astronomy course. Students either experienced 3D stereo visualizations using a GeoWall system while the other half experienced the same labs in 2D using the same system. Pre- and post-testing was conducted using the Lunar Phases Concept Inventory. Both students who experienced the 2D ($r^2 = 0.129$) and 3D ($r^2 = 0.069$) versions of the laboratory showed significant pre-post gains on the Lunar Phases Concept Inventory; there was no significant difference between the two groups of students.

(Table continued)

TABLE VI. *(Continued)*

References	Participants	Measures	Findings
Schneps, Ruel, Sonnert, Dussault, Griffin, and Sandler [105]	High school students ($N = 152$)	Astronomy and Space Science Concept Inventory	Students were exposed to 3D simulations of the solar system in either a true to scale form or an orrery-like simulation where scale relationships were exaggerated to focus on surface features of the planets. Students could manipulate the simulations in each condition on iPads. Pre- and post-testing was done using questions from the Astronomy and Space Science Concept Inventory. Students improved in understanding within both conditions, but the true to scale simulation helped more when scale was important in understanding the concept. Students either used the true to scale (TTS) display followed by the Orrery display or saw the displays in the opposite order. The TTS produced higher gain scores than the Orrery display. When the item types (scale vs general items), the TTS view resulted in higher grain scores for the scale items than the Orrery view, but there was no difference between views for the general items. Effect sizes ranged from 0.05 to 0.38 SD.
Meyer, Mon, and Hibbard [106]	Students in undergraduate astronomy or physical science course.	Lunar Phases Concept Inventory	Students were asked to make moon observations for about one and one-quarter of a cycle of lunar phases. The students used sextants made in class to describe the location of the moon and also shade in the unlit portion of the moon on a drawing. Prior to and after this project, students completed the Lunar Phases Concept Inventory, which the authors used to quantitatively describe changes in students' understanding of lunar phases. Lunar Phases Concept Inventory score increased from 41% prior to instruction to 56% post-instruction (Cohen's $d = 0.88$).

who experienced the instruction ($n = 17$). The interview data supported the use of the instructional approach. However, the researchers noted that testing the instructional approach with gifted students limits the generalizability.

Studies concerning the development of spatial-scientific understanding of lunar-related content have demonstrated middle level students making significant gains after participating in focused curricular interventions [83,99–101]. Each of these interventions used two astronomy-specific assessments to measure students' understanding of spatial, the GSA [84], and the LPCI [85]. During the interventions teachers implemented a project-based Earth-Space curriculum with their students after receiving intense professional development and training. Students were also required to keep Moon observation journals for at least five weeks, noting lunar patterns and motions, and the geometric orientation of the Earth-Moon-Sun system in specific phases.

Cole, Wilhelm, and Yang [101] found that keeping Moon observation journals contributed to middle school students' understanding of lunar phases. For every one point increase in the overall Moon journal scores, students improved their post-test score overall and also on the periodic patterns (PP) and geometric spatial visualization (GSV) spatial domains of the LPCI by 1%. Additional Moon observation journal entries also led to an increase in the overall score and the score on the GSV and PP spatial domains on the post-test.

Wilhelm [77] examined gender differences in middle school students' spatial-scientific understandings before and after a focused curricular intervention. Girls tended to not do as well as boys on the preassessments (LPCI and GSA); however, they performed just as well or better on postassessment LPCI spatial domain items, namely, GSV (visualizing from above, below, or within a system's plane) and CD (documenting an object's vector direction relative to a set location). Other intervention studies not only looked at gender differences in spatial-scientific understandings pre- and postcurricular interventions, but also at racial or ethnic differences [99,100]. Wilhelm *et al.* [99] compared students' spatial understandings by gender and race within and between control (business as usual) and treatment (project-based Earth-space unit). Findings showed all students (boys, girls, white, students of color) in the treatment group tended to have similar clustered significant spatial-scientific understandings postintervention while students in the business as usual group showed only boys with significant gains. Jackson *et al.* [100] found similar results in a study examining differences in learning between an experimental group of middle level students that received a curricular intervention and a control group. Findings showed experimental groups of boys, girls, students of color, and white students showed significant gains in their GSV understanding while control students' significant gains were limited to boys and white students.

TABLE VII. Newcombe-Shipley [64] categories for interventional studies.

References	Concept investigated	Assessment instrument(s)	Domain-specificity or generality of instrument(s)	Shipley-Newcombe framework [64]
Plummer, Kocareli, and Slagle [97]	Daily celestial motion	Interviews	Domain specific	Extrinsic static Extrinsic dynamic Intrinsic static
Plummer, Wasko, and Slagle [98]	Daily celestial motion	Interviews	Domain specific	Intrinsic dynamic Extrinsic dynamic Extrinsic static
Wilhelm, Toland, and Cole [99]	Lunar phases	LPCI PSVT:R	Domain specific Domain general	Extrinsic dynamic Extrinsic static Intrinsic dynamic
Jackson, Wilhelm, Lamar, and Cole [100]	Lunar phases	LPCI GSA PSVT:R	Domain specific Domain specific Domain general	Extrinsic dynamic Extrinsic static Intrinsic dynamic
Cole, Wilhelm, and Yang [101]	Lunar phases	Moon observation journals LPCI	Domain specific Domain specific	Extrinsic dynamic Intrinsic dynamic
Padalkar and Ramadas [102]	Earth-Moon-Sun system	Drawings	Domain specific	Intrinsic static Intrinsic dynamic Extrinsic dynamic
Price and Lee [103]	2D and 3D representations	Letter rotation Block rotation Paper folding Interviews	Domain general Domain general Domain general Domain specific	Intrinsic static
Cid and Lopez [104]	Lunar concepts learned in 2D or 3D stereo visualization	LPCI	Domain specific	Extrinsic dynamic Extrinsic static
Schneps, Ruel, Sonnert, Dussault, Griffin, and Sandler [105]	Solar system; general astronomy concept	Astronomy and Space Science Concept Inventory	Domain specific	Intrinsic static Extrinsic static Extrinsic dynamic
Meyer, Mon, and Hibbard [106]	Lunar phases	LPCI Student moon observations	Domain specific Domain specific	Extrinsic dynamic Intrinsic static Intrinsic dynamic

Diagrams are an essential form of external representation in astronomy. However, in order to understand diagrams, students must be able to understand the correspondence between 2D representations and 3D, often dynamic, phenomena. Padalkar and Ramadas [102] investigated how to integrate diagrams with other spatial tools in a year-long grade 8 intervention on the Earth-Moon-Sun system. Based on earlier assessments of the astronomy knowledge of grade 4 and 7 students, Padalkar and Ramadas identified three categories of diagrams often used by astronomy learners: diagrams that represent a model or part of a system, frequently drawn from an allocentric perspective; those representing a phenomenon, or patterns in a phenomenon over time, most often drawn from an egocentric perspective; and diagrams that attempt to explain or predict an astronomical event. Explanatory and predictive diagrams often combine egocentric and allocentric view perspectives. The intervention encouraged students to use physical gestures to map the correspondence between diagrams and concrete models of the Earth-Moon-Sun system and to convey spatial properties such as length, orientation, direction, or the dynamic trajectories of rays of light or celestial bodies. After 45 days of contact over a year, students' diagrams changed from picturelike representations of phenomena to more schematicized diagrams that utilized consistent, appropriate view perspectives.

Price and Lee [103] investigated the role of 2D versus 3D utilizing a Geowall located within an urban astronomy

museum, to assess 18 middle school students' spatial thinking skills. Each student completed three different spatial thinking tasks first in two dimensions using paper and then three dimensions using the Geowall. The spatial cognition tasks were based on published spatial thinking assessments. Price and Lee utilized questions for each of the following skills: letter rotation task, block rotation task, and the paper-folding task. These three tasks required an increasing degree of 3D manipulations. After completion of each task in two and three dimensions, the students participated in a short interview of how they perceived their experience. Though the accuracy of the responses did not differ, the time it took to complete the 3D manipulations was much greater. The researchers concluded that additional time is needed with the use of 3D representations to allow individuals to become familiar with stereoscopic visualizations.

Cid and Lopez [104] utilized a pre-post-test design to investigate the effectiveness 3D Geowall visualizations of the highly spatial concept of lunar phases as compared to 2D representations of the same phenomenon. The researchers implemented the study within a typical introductory college astronomy course for nontechnical majors (ASTRO 101), which consisted of 270 students. The sample was split into two groups: in one group students were taught lunar phases utilizing 2D representations, while the other group used the Geowall to experience 3D visuals of the Earth-Moon-Sun system. Rather than measure spatial thinking skills, the researchers constructed a pre-post design utilizing the LPCI [85] to assess student understanding of lunar phases. There was no significant difference between the two groups on the assessment.

Schnepps *et al.* [105] also investigated the relative contributions of 2D vs 3D visualizations of the solar system at overcoming common incorrect alternate understandings of astronomical phenomena among 152 high school students. The researchers hypothesized that a photorealistic, simulated 3D solar environment could convey the scale of the Earth-Sun-Moon system more effectively than 2D representations, such as textbook diagrams. Unlike the 3D platforms in previous studies that used stereoscopic visualization [103,104], the 3D representations in this study were implemented with an interactive pinch-to-zoom tablet interface. Students who experienced the 3D visualization demonstrated higher learning gains than those experiencing the 2D visualizations, as measured on a subset of 16 questions from the Astronomy and Space Science Concept Inventory. The authors hypothesized that the 3D environment permitted students to develop more accurate representations of scale and views of planetary bodies from different perspectives.

Meyer, Mon, and Hibbard [106] utilized the LPCI [85] in a pre-post test design to measure ASTRO 101 students' conceptual understanding of lunar phases. Results included a significant difference in pre- and postinstruction on the LPCI, but that is to be expected even with traditional presentation. Unfortunately, without comparison to other classes and/or published research, these results may not be generalizable. The researchers did not assess students' spatial thinking skills, but only focused on their conceptual understanding.

Similar to the noninterventional studies, many of the interventional studies could be placed into multiple categories within the Newcombe-Shipley [62] framework. These studies tended to focus on domain-general measures of spatial skill, reserving the domain-specific measures for assessing content knowledge. The exceptions to this were the Wilhelm *et al.* studies [99–101] where spatial abilities were mapped onto the questions of the domain-specific LPCI. While studies varied in the ways the content was taught or assessed, placing them in multiple categories besides the extrinsic-dynamic category of the phenomenon, there was not a consistent way in which the studies address the phenomena.

3. Learning progressions

Learning progressions (LPs) are empirically grounded, testable hypotheses that propose how students develop complex and complete understanding of a core scientific idea [107]. See Tables VIII and IX for summaries of each of the learning progression studies and how they fit within the Newcombe-Shipley framework [62]. LPs generally specify learning goals, measures of progress, and achievement markers that represent the integration of progressively more sophisticated levels of thinking about a specific scientific idea. The development of LPs was motivated by a call from U.S. education policy makers to create clear instructional standards, curricula, and assessments for science education [108].

Plummer [109] argued that sound spatial thinking is essential for developing a scientific explanation for celestial motion. According to Plummer [109] one of the challenges in creating a scientific explanation of celestial motion is reconciling the Sun's apparent motion through the sky with perspectives of Earth from space. Over a series of studies, Plummer and colleagues [109–111] proposed a series of LPs for how elementary school students develop a scientific understanding of celestial motion. Implicit in creating a scientific explanation of celestial motion is reconciling the Sun's apparent motion through the sky with perspectives of Earth from space. Plummer and Maynard [111] developed the LP by integrating a series of *construct maps*, representations of models of cognition that specify the lower and upper level anchors of spatial thinking skill required for understanding a given concept. Their learning progression has six levels of understanding, beginning with a naïve view of astronomy and a culminating in a scientific explanation of celestial motion.

TABLE VIII. Summary table for learning progressions.

References	Participants	Measures	Findings
Sneider, Bar, and Kavanagh [74]	Review	N/A	The authors reviewed 41 studies on learning about seasons. A learning progression for the cause of seasons was suggested, starting with the day/night cycle, moving to an Earth-based perspective, and ending with a space-based perspective. One of the four main reasons why learning the cause of seasons is so difficult is that combining Earth-based and space-based perspectives requires spatial reasoning skills.
Plummer [109]	Third grade students ($N = 99$) in a suburban school district in the northeastern United States.	Interviews	The role of spatial knowledge and reasoning in learning progressions on daily celestial motion and lunar phases are explored. The learning progression was applied to a study of children learning about such ideas. Plummer found that students' progression through the learning progression was shaped by their spatial ability. Students' spatial visualization ability affects their learning of daily celestial motion and lunar phases.
Plummer and Krajcik [110]	Students in first ($N = 20$), third ($N = 20$), and 8th ($N = 20$) grades in the Midwest. Sixty-three first and second grade students in the Midwest.	Interviews	The authors used prior research to develop a set of learning progressions for topics related to celestial motion. The authors also used the learning progressions to analyze learning due to an instructional intervention in a planetarium. Four learning progressions related to daily celestial motion are presented and compared to students' learning prior to and after a short planetarium program. Targeted instruction improved students' understanding compared to business as usual instruction; learning progression levels were used to categorize learning.
Plummer and Maynard [111]	8th Grade students ($N = 38$)	13 question assessment based on teacher-generated questions and questions from *Reason for Seasons*	Rasch analysis was used to revisit construct maps addressing the reason for the seasons. For students to move through the levels to progressively more sophisticated explanations, students need to be able to move between space-based and Earth-based perspectives. Authors suggest that instruction on celestial motion needs to intentionally address the spatially complex connection between Earth-based observations and space-based perspectives. Construct map shows progression of ideas about celestial motion consistent with the NGSS. Findings also show that making the connection between Earth-based and space-based perspectives is a major challenge for students' learning to explain celestial motion.
Testa, Galano, Leccia, and Puddu [112]	Italian students ($N = 300$) at the beginning (age 14) and end (age 18) of secondary school.	Two-tier instrument designed by the authors	Item response analysis and curve integral method were used to develop a learning progression about celestial motion. Findings support spatial reasoning as a key factor in learning about celestial motion. The learning progression supports the idea that spatial reasoning is key to understanding celestial motion. The authors also found that causal reasoning is also essential. The authors suggest first teaching the geometry of the Earth-Moon-Sun system, followed by the motion of the system, and finally the spatial relationship between Earth-based and space-based perspectives.

Testa, Galano, Leccia, and Puddu [112] proposed a Learning Progression that integrates three stages of visuospatial and conceptual mastery required to develop an accurate understanding of celestial motion. Their learning progression argues that mastery of visuospatial concepts of the Earth-Moon-Sun system is required before students can understand celestial motion. The second stage of their learning progression involves understanding the physical consequences of Earth's movement around the Sun, including changing solar radiation. The third stage of their LP involves moving through different frames of reference. Their learning progression has six levels of understanding, beginning with a naïve view of astronomy and a culminating in a scientific explanation of celestial motion.

Sneider, Bar, and Kavanagh [74] proposed a learning progression of the seasons that spans three age ranges. At

TABLE IX. Newcombe-Shipley [64] categories for learning progressions.

References	Concept investigated	Assessment instrument(s)	Domain-specificity or generality of assessment instrument(s)	Shipley-Newcombe framework [64]
Sneider, Bar, and Kavanagh [74]	Seasons	N/A (Review Article)	N/A	Extrinsic dynamic
Plummer [109]	Daily celestial motion; Lunar phases	Interviews	Domain specific	Extrinsic dynamic Extrinsic static
Plummer and Krajcik [110]	Celestial motion	Interviews	Domain specific	Extrinsic dynamic
Plummer and Maynard [111]	Seasons	13 question assessment based on teacher-generated questions and questions from *Reason for Seasons*	Domain specific	Extrinsic dynamic
Testa, Galano, Leccia, and Puddu [112]	Celestial motion	Two-tier instrument designed by the authors	Domain specific	Extrinsic dynamic

grades 4–5, instruction focuses on the day-night cycle. At this level students must reconcile their perceptual observations of day and night with spatial concepts such as the shapes and relative sizes of Earth and the Sun, the tilt of Earth, and the rotation of Earth around the Sun. For grades 6–8, the curriculum includes instruction in the physics of light (light travels in straight lines; sunlight that strikes Earth at a steep angle conveys more warmth than sunlight that strikes Earth at an oblique angle) and the daily and seasonal fluctuations in climate due to the Sun's changing path in the sky and the latitude of the location). At grades 9–12 the curriculum integrates their understanding of seasons from the perspective of Earth with a vision of Earth from space. At this level students are challenged to synthesize their knowledge of Earth's climate zones, their understanding of Earth's orbit around the Sun and the tilt of Earth's axis to understand the reason for changes in the length of daylight across the seasons.

The literature on learning progressions in astronomy has focused primarily on extrinsic-dynamic content. This is possibly because the focus of the paper was on the astronomical phenomenon rather than specific instruction, making it difficult to categorize the studies into other relevant Newcombe-Shipley [62] categories.

III. CONCLUSIONS AND RECOMMENDATIONS

A. Discussion

In this review of literature, we have categorized and described studies that investigated the intimate relationship between spatial thinking skill and understanding of astronomical phenomena (i.e., celestial motions, cause of phases, cause of seasons). We summarized the developmental, psychometric and cognitive approaches for describing and measuring spatial thinking skills, surveyed research on the role of spatial skills in other STEM disciplines, and introduced new typologies for studying spatial skill. We classified and reviewed three types of studies: noninterventional, interventional, and learning progressions. The noninterventional studies are consistent with a history in astronomy education research of understanding the nature of students' misconceptions and content understanding. However, the studies reviewed here build on the literature by also investigating the contribution of spatial thinking. In some noninterventional studies, researchers found significant correlations between astronomy content assessments and tests of spatial skills.

Our paper showed that young children struggle to explain the cause of Moon phases, sometimes relying on perceptual experiences and cultural and social influences (stories, media, etc.) for explanations. As students move through middle school, their role of spatial thinking in the discipline becomes more apparent, as those who struggle to change view perspectives while visualizing the Earth-Moon-Sun system seem to have more difficulty in explaining astronomical phenomena as well [75–78,81].

Interventional studies used planetaria [97], focused curricula [83], and compared 2D and 3D [103–105] models to enhance students' understanding of astronomical concepts. Each of these studies showed pre-to-post intervention improvements in students' understanding of astronomical concepts as well as commented on an aspect of spatial thinking. For instance, the Wilhelm *et al.* studies [83,84] showed that students improved in spatial thinking as well as in understanding the cause of phases of the Moon. However, there is a relative lack of experimental studies found in the literature that focus on spatial thinking in astronomy education. While there are a few studies [83,99,100] where an explicit control or business as usual group is used as a comparison for the treatment group, this setup is rarely found within the spatial literature in astronomy education. Several other studies compared

multiple conditions, such as Cid and Lopez [104] where 2D and 3D versions of the same lab were compared, but none of these were explicitly identified as a control group. Another limitation is that effect sizes are not included in all papers, and when they are included they are often reported in different ways, making comparisons across papers more difficult. Without experimental (or quasiexperimental) designs, it is more difficult to rule out other cognitive processes besides spatial thinking skills (e.g., executive function, visual working memory, general intelligence) that may be contributing to students' learning gains.

Learning progressions are a methodology that is gaining momentum in astronomy education. Given the importance of spatial competency in astronomy, learning progressions are an excellent method for guiding curriculum that takes into account the inherent spatial concepts in the content. Learning progressions can also be used to guide the instructor in the need to evaluate the spatial skills of individual students and provide remediation if necessary to individuals as they progress from naïve understandings of astronomy to scientifically accurate understandings.

The literature demonstrated that students' spatial-scientific understandings could be developed through purposeful curricular interventions, technologies, and experiences. Such interventions resonate with the National Research Council's [1] charge to enhance learners' abilities to visualize relationships between static and moving objects while taking into account distance, direction, and perspective.

B. Recommendations

Regardless of the type of study (e.g., noninterventional, interventional, or learning progression), we recommend that researchers specify and describe which tests they use to assess content knowledge and spatial skills. Differences in definitions exist not only between cognitive psychologists and education researchers, but also between science education and discipline-based education researchers and even within each of these groups. Given the variability of terminology in categorizing spatial skills, we recommend that researchers identify the astronomical context of the spatial skill they attempted to measure and the tools they used to assess this construct. For example, a researcher might state that they used test X to measure the ability to change view perspective required when explaining the geometric orientation of the Earth-Moon-Sun system at specific lunar phases. Researchers should cite any content or spatial test by name, whether they are using the test as originally designed or in a modified version. If the study used a modified version of a test, we recommend that the researcher indicate how it was modified and why the modified version was used.

We recommend further research investigating the relative utility of domain-independent and domain-specific spatial tests in astronomy education research. There is evidence that some domain-general spatial tests, such as the PSVT:R [20], show significant positive correlations with assessments of astronomy content, such as the LPCI [85]. However, such a correlation alone does not explain where and how the mental rotation skills assessed in the PSVT:R are utilized in solving astronomy problems. Other data are needed to explain the correlation. There is also evidence that the domain-general spatial thinking problems on the GSA [84] show significant positive correlations with astronomy knowledge as represented in the LPCI [85], when the spatial items are mapped to content knowledge hypothesized to use similar skills.

If we find that it is important to develop additional domain-specific assessments of spatial ability in astronomy, we need to develop these tests. If we find that domain-general measures of spatial thinking are adequate, we then need to determine which tests have the greatest predictive validity for different content areas of astronomy. Other STEM disciplines take note of the different spatial requirements of learning from models, drawings, and gestures; astronomy educators should do so as well. While we have used the Newcombe-Shipley framework [62] to further categorize the studies we reviewed, we were only able to categorize studies based on our interpretation of the instruction and/or assessments described within each paper. Many of the papers were assigned multiple category labels. We recommend that researchers adopt a similar common framework and include information on (i) the object or system considered, (ii) whether intrinsic or extrinsic properties were addressed or emphasized, and (iii) the ways in which 1 and 2 apply to instruction and/or assessment in their study.

Another question is whether the spatial thinking skills acquired in astronomy can be transferred to another scientific domain. We also recommend that researchers investigating differences in spatial skills also test for other measures of cognitive resources, such as attentional control. It is important to both identify which spatial skills are important for understanding astronomy and identifying whether it is the spatial skill alone or in conjunction with other cognitive resources. Additional robust experimental (or quasiexperimental) studies are needed where interventions can be compared with control or business as usual groups where assessments and effect sizes for each condition are clearly reported.

We have included studies in this review that address spatial thinking in astronomy across a wide range of grade levels, from early childhood through college. Given this range, we also suggest researchers attend to the appropriateness of the specific test to the age of students, checking and reporting on reliability and validity of the test for their study's population.

These studies also highlight the need for instructors to be aware of students who may have less developed spatial skills [79] and to plan instruction that addresses both content and the development and/or use of spatial thinking in the

discipline [56] so that all students have the opportunity to be successful in learning astronomy. Whether this awareness of students of differing spatial ability is screened for prior to instruction or becomes apparent throughout instruction, instructors need to be aware of the role spatial thinking plays in understanding astronomy and other STEM content as well as best practices for addressing the challenges of teaching inherently spatial content. Teachers also need to learn how to create spatially rich lessons that allow students to develop spatial thinking skills in addition to the content. The intervention literature we reviewed showed that simply covering content is not enough, but it is important to also cover content in a way that builds spatial skills as well. Teachers also need to consider the role of both domain-general and domain-specific spatial skills that are relevant to understanding astronomy content, and helping students learn those spatial skills that are most relevant. Astronomy education researchers should also investigate the role of teachers' spatial thinking ability may play in students' understanding of astronomy content. In order for teachers to create spatially rich learning environments, they need to be aware of their own spatial thinking skills as well as how to foster the needed spatial skills in their students.

[1] *National Research Council, Learning to Think Spatially: GIS as a Support System in the K-12 Curriculum* (National Academies Press, Washington, DC, 2006).

[2] J. Wai, D. Lubinski, and C. P. Benbow, Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance, J. Educ. Psychol. **101**, 817 (2009).

[3] R. M. Webb, D. Lubinski, and C. P. Benbow, Spatial ability: A neglected dimension in talent searches for intellectually precocious youth, J. Educ. Psychol. **99**, 397 (2007).

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

[5] J. Carroll, *Human Cognitive Abilities: A Survey of Factor-Analytic Studies* (Cambridge University Press, New York, NY, 1993).

[6] J. Piaget and B. Inhelder, *The Child's Conception of Space* (W. W. Norton & Company, New York, 1967).

[7] S. B. Ehrlich, S. C. Levine, and S. Goldin-Meadow, The importance of gesture in children's spatial reasoning, Dev. Psychol. **42**, 1259 (2006).

[8] E. A. Gunderson, G. Ramirez, S. L. Beilock, and S. C. Levine, The relation between spatial skill and early number knowledge: The role of the linear number line, Dev. Psychol. **48**, 1229 (2012).

[9] N. S. Newcombe and J. Huttenlocher, Development of spatial cognition, *Handbook of Child Psychology* (John Wiley & Sons, Inc, New York, 2006), 6th ed., Vol. 2, Doi: 10.1002/9780470147658.chpsy0217.

[10] N. S. Newcombe and A. Frick, Early education for spatial intelligence: Why, what, and how, Mind, Brain Educ. **4**, 102 (2010).

[11] M. Vasilyeva and S. F. Lourenco, Development of spatial cognition, Wiley Interdisciplinary Rev.: Cogn. Sci. **3**, 349 (2012).

[12] D. H. Uttal, N. G. Meadow, E. Tipton, L. L. Hand, A. R. Alden, C. Warren, and N. S. Newcombe, The malleability of spatial skills: A meta-analysis of training studies, Psychol. Bull. **139**, 352 (2013).

[13] M. Baenninger and N. Newcombe, The role of experience in spatial test performance: A meta-analysis, Sex Roles **20**, 327 (1989).

[14] D. H. Uttal and C. A. Cohen, Spatial thinking and STEM education: When, why, and how?, Psychol. of learning and motivation **57**, 147 (2012).

[15] L. L. Thurstone, *Primary Mental Abilities* (The University of Chicago Press, Chicago, IL, 1938).

[16] J. Eliot and I. M. Smith, *An International Directory of Spatial Tests* (Nfer-Nelson, Windsor, Berks, 1983).

[17] M. Hegarty and D. Waller, Individual differences in spatial abilities, in *The Cambridge Handbook of Visuo-spatial Thinking*, edited by P. Shah and A. Miyake (Cambridge University Press, New York, 2005), p. 121.

[18] L. G. Humphreys, D. Lubinski, and G. Yao, Utility of predicting group membership and the role of spatial visualization in becoming an engineer, physical scientist, or artist, J. Appl. Psych. **78**, 250 (1993).

[19] S. Vandenberg and A. Kuse, Mental rotations: Group test of three-dimensional spatial visualization, Perceptual Motor Skills **47**, 599 (1978).

[20] G. M. Bodner and R. B. Guay, The Purdue visualization of rotations test, Chemical Educator **2**, 1 (1997).

[21] R. B. Ekstrom, J. W. French, H. H. Harman, and D. Dermen, *Manual from Kit of Factor- References Cognitive Tests* (Educational Testing Service, Princeton, NJ, 1976).

[22] G. K. Bennett, H. G. Seashore, and A. G. Wesman, *Differential Aptitude Tests for Personnel and Career Assessment* (Pearson, 1991).

[23] M. W. Alibali, Gesture in spatial cognition: Expressing, communicating, and thinking about spatial information, Spatial Cognit. Comput. **5**, 307 (2005).

[24] A. Baddeley, Working memory, Science **255**, 556 (1992).

[25] S. Kosslyn, *Image and Mind* (Harvard University Press, Cambridge, MA, 1980).

[26] S. Kosslyn, J. Brunn, K. Cave, and R. Wallach, Individual differences in mental imagery ability: A computational analysis, Cognition **18**, 195 (1984).

[27] P. A. Carpenter, M. A. Just, T. A. Keller, W. Eddy, and K. Thulborn, Graded functional activation in the visuospatial

system with the amount of task demand, J. Cogn. Neurosci. **11**, 9 (1999).
[28] M. A. Just and P. A. Carpenter, Cognitive coordinate systems: Accounts of mental rotation individual differences in spatial ability, Psychol. Rev. **92**, 137 (1985).
[29] D. F. Lohman, Spatial abilities as traits, processes, and knowledge, in *Advances in the Psychology of Human Intelligence*, edited by R. J. Sternberg (Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1988), Vol. 4, pp. 181–248.
[30] R. J. Mumaw and J. W. Pellegrino, Individual differences in complex spatial processing, J. Educ. Psychol. **76**, 920 (1984).
[31] M. Kozhevnikov and M. Hegarty, A dissociation between object manipulation spatial ability and spatial orientation ability, Memory Cogn. **29**, 745 (2001).
[32] R. Diamond, S. Carey, and K. J. Back, Genetic influences on the development of spatial skills during early adolescence, Cognition **13**, 167 (1983).
[33] M. B. Casey, R. L. Nuttall, and E. Pezaris, Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills, Dev. Psychol. **35**, 1237 (1999).
[34] C. M. Steele and J. Aronson, Stereotype threat and the intellectual test performance of African-Americans, J. Personality Social Psychol. **69**, 797 (1995).
[35] A. Moè and F. Pazzaglia, Following the instructions!: Effects of gender beliefs in mental rotation, Learning and individual differences **16**, 369 (2006).
[36] A. Moè, Are males always better than females in mental rotation? Exploring a gender belief explanation, Learning and individual differences **19**, 21 (2009).
[37] M. S. McGlone and J. Aronson, Stereotype threat, identity salience, and spatial reasoning, J. Appl. Dev. Psychol. **27**, 486 (2006).
[38] Spatial Intelligence and Learning Center, http://spatiallearning.org.
[39] M. Hegarty, Spatial thinking in undergraduate science education, Spatial Cognit. Comput. **14**, 142 (2014).
[40] K. Rochford, Spatial learning disabilities and underachievement among university anatomy students, Med. Educ. **19**, 13 (1985).
[41] A. X. Garg, G. Norman, and L. Sperotable, How medical students learn spatial anatomy, Lancet **357**, 363 (2001).
[42] O. Ha and S. Brown, Spatial reasoning difference between civil and mechanical engineering students in learning Mechanics of Materials course: A case of cross-sectional inference, *Proceedings of the 2017 ASEE Annual Conference & Exposition*, Columbus, OH (2017).
[43] C. A Cohen and M. Hegarty, Individual differences in use of external visualisations to perform an internal visualisation task, Appl. Cogn. Psychol. **21**, 701 (2007).
[44] Y. Kali and N. Orion, Spatial abilities of high-school students in the perception of geologic structures, J. Res. Sci. Teach. **33**, 369 (1996).
[45] N. Orion, D. Ben-Chaim, and Y. Kali, Relationship between Earth-science education and spatial visualization, J. Geosci. Educ. **45**, 129 (1997).
[46] L. Liben, K. A. Kastens, and A. E. Christensen, Spatial foundations of science education: The illustrative case of instruction on introductory geological concepts, Cognit. Instr. **29**, 45 (2011).
[47] K. A Kastens and T. Ishikawa, Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields, *Geological Society of America, Special Papers* Vol. 413 (Geological Society of America, Boulder, Colorado, 2006), p. 53.
[48] L. S. Liben and S. J. Titus, The importance of spatial thinking for geoscience education: Insights from the crossroads of geoscience and cognitive science, in *Geological Society of America, Special Papers* Vol. 486 (Geological Society of America, Boulder, Colorado, 2012), pp. 51–70.
[49] R. Grabow, Ph.D. thesis, California State University, Fullerton, CA, 2003.
[50] C. S. Carter, M. A. Larussa, and G. M. Bodner, A study of two measures of spatial ability as predictors of success in different levels of general chemistry, J. Res. Sci. Teach. **24**, 645 (1987).
[51] M. Kozhevnikov and R. Thornton, Real-time data display, spatial visualization ability, and learning force and motion concepts, J. Sci. Educ. Technol. **15**, 111 (2006).
[52] M. Hegarty and V. K. Sims, Individual differences in mental animation during mechanical reasoning, Memory Cogn. **22**, 411 (1994).
[53] M. Kozhevnikov, M. Hegarty, and R. E. Mayer, Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers, Cognit. Instr. **20**, 47 (2002).
[54] E. H. Brinkmann, Programmed instruction as a technique for improving spatial visualization, J. Appl. Psych. **50**, 179 (1966).
[55] T. Lord, Enhancing the visuo-spatial aptitude of students, J. Res. Sci. Teach. **22**, 395 (1985).
[56] M. Y. Small and M. E. Morton, Research in college science teaching: Spatial visualization training improves performance in organic chemistry, J. Coll. Sci. Teach. **13**, 41 (1983).
[57] H. Gerson, S. Sorby, A. Wysocki, and B. Baartmans, The development and assessment of multimedia software for improving 3-D spatial visualization skills, Comput. Appl. Eng. Educ. **9**, 105 (2001).
[58] J. Feng, I. Spence, and J. Pratt, Playing an action video game reduces gender differences in spatial cognition, Psychol. Sci. **18**, 850 (2007).
[59] R. Wright, W. L. Thompson, G. Ganis, N. S. Newcombe, and S. M. Kosslyn, Training generalized spatial skills, Psychon. Bull. Rev. **15**, 763 (2008).
[60] C. A. Sanchez, Enhancing visuospatial performance through video game training to increase learning in visuospatial science domains, Psychon. Bull. Rev. **19**, 58 (2012).
[61] M. Hegarty, Components of spatial intelligence, Psychol. Learning Motivation **52**, 265 (2010).
[62] N. S. Newcombe and T. F. Shipley, Thinking about spatial thinking: New typology, new assessments, *Studying Visual and Spatial Reasoning for Design Creativity* (Springer, Dordrecht, 2015), pp. 179–192.
[63] I. Resnick and T. F. Shipley, Breaking new ground in the mind: an initial study of mental brittle transformation and

mental rigid rotation in science experts, Cognit. Processing **14**, 143 (2013).
[64] K. Atit, T. F. Shipley, and B. Tikoff, Twisting space: Are rigid and non-rigid mental transformations separate spatial skills?, Cognit. Processing **14**, 163 (2013).
[65] D. R. Montello, Scale and multiple psychologies of space, *European Conference on Spatial Information Theory* (Springer, Berlin, Heidelberg, 1993), pp. 312–321.
[66] M. Hegarty, D. R. Montello, A. E Richardson, T. Ishikawa, and K. Lovelace, Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning, Intelligence **34**, 151 (2006).
[67] M. Hegarty, A. E. Richardson, D. R. Montello, K. Lovelace, and I. Subbiah, Development of a self-report measure of environmental spatial ability, Intelligence **30**, 425 (2002).
[68] J. M Bailey and T. F. Slater, A review of astronomy education research, Astron. Educ. Rev. **2**, 20 (2003).
[69] J. M Bailey and T. F. Slater, Resource letter AER-1: Astronomy education review, AJP **73**, 677 (2005).
[70] P. L. Callison and E. L. Wright, The effect of teaching strategies using models on preservice elementary teachers' conceptions about Earth-Sun-Moon relationships, *Proceedings of the Annual Meeting of the National Association for Research in Science Teaching*, April, Atlanta, GA (1993), as cited in Ref. [46].
[71] C. I. Sneider, S. Pulos, and S. 1983, Children's cosmographies: Understanding the Earth's shape and gravity, Sci. Educ. **67**, 205 (1983) as cited in Ref. [46].
[72] A. Lelliott and M. Rollnick, M. Big ideas: A review of astronomy education research 1974–2008, Int. J. Sci. Educ. **32**, 1771 (2010).
[73] B. D. Brazell and E. Spinoza, Meta-analysis of planetarium efficacy research, Astron. Educ. Rev. **8**, 1 (2009).
[74] C. Sneider, V. Bar, and C. Kavanagh, Learning about seasons: A guide for teachers and curriculum developers, Astron. Educ. Rev. **10**, 010103-1 (2011).
[75] J. D. Plummer, C. A. Bower, and L. Liben, The role of perspective taking in how children connect reference frames when explaining astronomical phenomena, Int. J. Sci. Educ. **38**, 345 (2016).
[76] S. E. Sherrod and J. A. Wilhelm, A study of how classroom dialogue facilitates the development of geometric spatial concepts related to understanding the cause of Moon phases, Int. J. Sci. Educ. **31**, 873 (2009).
[77] J. Wilhelm, Gender differences in lunar-related scientific and mathematical understandings, Int. J. Sci. Educ. **31**, 2105 (2009).
[78] J. Wilhelm, A case study of three childrens' original interpretations of the Moon's changing appearance, School Sci. Math. **109**, 258 (2009).
[79] G. J. Venville, R. D. Louisell, and J. A. Wilhelm, Young children's knowledge about the Moon: A complex dynamic system, Res. Sci. Educ. **42**, 729 (2012).
[80] K. Subramaniam and S. Padalkar, Visualization and reasoning in explaining the phases of the Moon, Int. J. Sci. Educ. **31**, 395 (2009).
[81] E. Kikas, The effect of verbal and visuospatial abilities on the development of knowledge of the Earth, Res. Sci. Educ. **36**, 269 (2006).
[82] A. Toomela, Drawing as a verbally mediated activity: A study of relationships between verbal, motor, and visuospatial skills and drawing in children, Int. J. Behav. Dev. **26**, 234 (2002), as cited in Ref. [63].
[83] J. Wilhelm, C. Jackson, A. Sullivan, and R. Wilhelm, Examining differences between preteen groups' spatial-scientific understandings: A quasi-experimental study, J. Educ. Res. **106**, 337 (2013).
[84] J. Wilhelm, B. Ganesh, S. Sherrod, and J. Ji, Spatial assessment, assessment developed at Texas Tech University, Lubbock, TX (2007).
[85] R. S. Lindell and J. P. Olsen, Developing the lunar phases concept inventory, in *Proceedings of the Physics Education Research Conference 2002* (AIP, New York, 2002).
[86] A. A. Black, Spatial ability and Earth science conceptual understanding, J. Geosci. Educ. **53**, 402 (2005).
[87] I. Heyer, J. S. Slater, and F. T. Slater, Establishing the empirical relationship between non- science majoring undergraduate learners' spatial thinking skills and their conceptual astronomy knowledge, Revista Latino-Americana de Educacao em Astronomia **16**, 45 (2013).
[88] C. Türk, The correlation between pre-service science teachers' astronomy achievement, attitudes towards astronomy, and spatial thinking skills, J. Educ. Learning **5**, 187 (2016).
[89] H. A. Witkin, P. K Oltman, E. Raskin, and S.A Karp, *Group Embedded Figures Test Manual* (Mind Garden, Inc., Redwood City, CA, 2002).
[90] T. F. Slater and S. J. Slater, Development of the Test of Astronomy Standards (TOAST) assessment instrument, Bull. Am. Astron. Soc. **40**, 273 (2008).
[91] C. A. Morrow, Kinesthetic astronomy: The sky time lesson, Phys. Teach. **38**, 252 (2000).
[92] C. Morrow and M. Zawaski, Kinesthetic Astronomy, Space Science Institute: Education: Instructional Materials, Available at: http://www.spacescience.org/education/index.html (2004).
[93] S. J. Parker (Slater), Master's Thesis, Montana State University, Bozeman, MT, 2007.
[94] D. S. Rudman, Solving astronomy problems can be limited by intuited knowledge, spatial ability, or both, *Annual Meeting of the American Educational Research Association, New Orleans, LA* (2002).
[95] D. Heywood, D. J. Parker, and M. Rowlands, Exploring the visuospatial challenge of learning about day and night and the Sun's path, Sci. Educ. **97**, 772 (2013).
[96] R. S. Taylor and E. D. Grundstrom, Diagrammatic representational constraints of spatial scale in Earth–Moon system astronomy instruction, Astron. Educ. Rev. **10**, 010104-1 (2011).
[97] J. D. Plummer, A. Kocareli, and C. Slagle, Learning to explain astronomy across moving frames of reference: Exploring the role of classroom and planetarium-based instructional contexts, Int. J. Sci. Educ. **36**, 1083 (2014).
[98] J. D. Plummer, K. D. Wasko, and C. Slagle, C. Children learning to explain daily celestial motion: Understanding astronomy across moving frames of reference, Int. J. Sci. Educ. **33**, 1963 (2011).
[99] J. Wilhelm, M. Toland, and M. Cole, Evaluating middle school students' spatial-scientific performance within

Earth/space astronomy in terms of gender and race/ethnicity, J. Educ. Sci. Environ. Health **3**, 40 (2017).

[100] C. Jackson, J. A. Wilhelm, M. Lamar, and M. Cole, Gender and racial differences: Development of sixth grade students' geometric spatial visualization within an Earth/Space Unit, School Sci. Math. **115**, 330 (2015).

[101] M Cole, J. Wilhelm, and H. Yang, Student Moon observations and spatial-scientific reasoning, Int. J. Sci. Educ. **37**, 1815 (2015).

[102] S. Padalkar and J. Ramadas, Modeling the round Earth through diagrams, Astron. Educ. Rev. **6**, 54 (2008).

[103] A. Price and H. S. Lee, The effect of two-dimensional and stereoscopic presentation on middle school students' performance of spatial cognition tasks, J. Sci. Educ. Technol. **19**, 90 (2010).

[104] X. Cid and R. Lopez, The impact of stereo display on student understanding of phases of the Moon, Astron. Educ. Rev. **9**, 010105 (2010).

[105] M. H. Schneps, J. Ruel, G. Sonnert, M. Dussault, M. Griffin, and P. M. Sadler, Conceptualizing astronomical scale: Virtual simulations on handheld tablet computers reverse misconceptions, Comput. Educ. **70**, 269 (2014).

[106] A. O. Meyer, M. Mon, and S. Hibbard, The lunar phases project: A mental model-based observational project for undergraduate nonscience majors, Astron. Educ. Rev. **10**, 010203 (2011).

[107] *National Research Council. Taking Science to School: Learning, and Teaching Science in Grades K-8*, edited by R. A. Duschl, H. A. Schweingruber, and A. W. Shouse (National Academies Press, Washington, DC, 2007).

[108] T. Corcoran, F. A. Mosher, and A. Rogat, Learning progressions in science: An evidence-based approach to reform, *Consortium for Policy Research in Education Report #RR-63* (Consortium for Policy Research in Education, Philadelphia, PA, 2009).

[109] J. D. Plummer, Spatial thinking as the dimension of progress in an astronomy learning progression, Studies Sci. Educ. **50**, 1 (2014).

[110] J. D. Plummer and J. Krajcik, Building a learning progression for celestial motion: Elementary levels from an Earth-based perspective, J. Res. Sci. Teach. **47**, 768 (2010).

[111] J. D. Plummer and L. Maynard, Building a learning progression for celestial motion: An exploration of students' reasoning about the seasons, J. Res. Sci. Teach. **51**, 902 (2014).

[112] I. Testa, S. Galano, S. Leccia, and E. Puddu, Development and validation of a learning progression for change of seasons, solar and lunar eclipses, and moon phases, Phys. Rev. ST Phys. Educ. Res. **11**, 020102 (2015).