JOURNAL OF GEOSCIENCE EDUCATION **65**, 455–472 (2017)

Learning About Spatial and Temporal Scale: Current Research, Psychological Processes, and Classroom Implications

Kim A. Cheek,[1,a] Nicole D. LaDue,[2] and Thomas F. Shipley[3]

ABSTRACT

Geoscientists analyze and integrate spatial and temporal information at a range of scales to understand Earth processes. Despite this, the concept of scale is ill defined and taught unevenly across the K–16 continuum. This literature review focuses on two meanings of scale: one as the magnitude of the extent of a dimension and the other as a relationship between objects or events. We review 42 papers from science education and discipline-based education research (DBER) literature on students' conceptions related to one or both meanings of scale. Analysis of this prior work reveals a broader (though still limited) research base on domain general concepts of scale as magnitude and scant research on scale as a relationship. Learners begin reasoning about spatial and temporal magnitudes categorically by working with scales based on standard units and nonmetric values, such as body length. Concepts of scale magnitudes outside human experience are nonlinear. Facility with fractions and proportional reasoning are positively associated with the ability to reason about scale as a relationship. Two constructs from the psychological literature, structure mapping and the category adjustment model, offer theoretical accounts for these findings. We borrow a typology from the psychological literature to frame common geoscience instructional models in the context of spatial and temporal scale and suggest how instructors might facilitate students' reasoning about scale models. We identify a number of avenues for possible future research, including a critical need to understand how conceptual understanding of scale develops across the K–16 continuum. © *2017 National Association of Geoscience Teachers*. [DOI: 10.5408/16-213.1]

Key words: scale, spatial cognition, magnitude, proportion

LEARNING ABOUT SPATIAL AND TEMPORAL SCALE: CURRENT RESEARCH, PSYCHOLOGICAL PROCESSES, AND CLASSROOM IMPLICATIONS

Modern technologies, including geospatial analysis, enable scientists to study processes and discern patterns in the natural world at spatial and temporal scales previously impossible. Scale and scalar relationships are integral to many disciplines, especially geoscience. Many spatial thinking skills, including reasoning about scale, are important in science, technology, engineering, and mathematics fields (Uttal et al., 2013). Some, such as the relationship between mental rotation and visual penetrative ability (Ormand et al., 2014; Atit et al., 2015), have been researched more extensively than scale. One reason for a lack of focused research on scale may be that the term *scale* is polysemous. Scale can refer to an instrument to measure mass, a standard of measurement, a proportional relationship, a set of musical notes, or a small outer covering on an organism, to name a few. We focus on two ways scale is used throughout geoscience in this paper. The first refers to the spatial or temporal extent of an object or event. By this definition, scale is the magnitude of a dimension that can be measured using standard or nonstandard units. Scalar intervals can be expressed linearly or logarithmically. Adjectives to describe scale as magnitude, include large, small, long, or short, but the magnitudes they denote vary across contexts. The second meaning of scale refers to a relationship between objects or events. Scale models or maps are examples of relational uses of scale in which the ratio of the size of an object in a model, drawing, or map is proportional to the size of the actual object. In context, scale as a relationship can be described using relational terms, such as larger, smaller, longer, or shorter.

We begin by describing why both meanings of scale are important in geoscience and how learning about scale can serve as a bridge between K–12 and undergraduate education. We then survey a sample of the science education and discipline-based education research (DBER) on learning about scale in science, highlighting insights from that literature and gaps that remain. Next, we briefly describe two areas of psychological research that offer a theoretical lens through which to view the science education and DBER findings. We propose a framework that can help geoscience educators use familiar models to improve students' understanding of scale. Insights from the research described herein are used to make pedagogical and research recommendations to help geoscience learners better comprehend scale.

THE IMPORTANCE OF SCALE IN GEOSCIENCE AND RELATED DISCIPLINES

Scientists use scale to denote both magnitudes and relationships. Measurable magnitude scales at which scientists work can be roughly grouped into three categories:

Received 31 August 2016; revised 16 March 2017; accepted 9 April 2017; published online 16 November 2017.

[1]Department of Childhood Education, Literacy, and TESOL, University of North Florida, 1 UNF Drive, Jacksonville, Florida 32224, USA
[2]Department of Geology and Environmental Geosciences, Northern Illinois University, 302C Davis Hall, Dekalb, IL, 60115, USA
[3]Department of Psychology, Temple University, Weiss Hall 1701 North 13th St, Philadelphia, PA 17112, USA
[a]Author to whom correspondence should be addressed. Electronic mail: k.cheek@unf.edu. Tel: 904-620-2794. Fax: 904-620-1025

scales directly observable by humans, those that are too small or occur too quickly for human observation, and those that are too large or occur too slowly for human observation. Geoscientists work at a variety of spatial and temporal scales and thus must reason about a variety of magnitudes. Scale as magnitude is represented on a numeric scale with endpoints that delimit the lower and upper bounds of the scale's extent, like the divisions on the geological timescale. Like geoscience, atmospheric science bases scalar intervals on the spatial or temporal magnitude of significant events, but boundaries between those intervals have not been consistently defined, which could create some uncertainty for learners (Markowski and Richardson, 2010). In general, spatial and temporal scales are related (e.g., Wilson cycles occur over long periods). Yet the relationship is not always straightforward. For example, some small-scale molecular changes, like carbonate dissolution, can have large-scale impacts.

Scalar relationships among variables may change with scalar magnitude, and geological properties may change with scale (e.g., electrical, magnetic, and physical properties of substances differ at the nanoscale versus larger scales). Experts engage in proportional, scalar transformations to easily move from human scales to nonhuman scales in their work (Jones and Taylor, 2009). Geological mapping provides one example of how many practitioners integrate the two meanings of scale in the context of their work. Topographic maps are used to orient and plan routes through a landscape. Geoscientists transfer scalar relationships on the map to relationships in the field. They move mentally back and forth from parts to whole, making observations at various locations and noting relationships to map spatially separated outcrops that are part of the same formation. Measurements and observations at different magnitude scales are used to draw geological maps and to reconstruct the geological history of the area. Later, thin sections of collected samples can be analyzed to provide further data on the lithology of the formation. Rock samples that appear homogeneous to the human eye may be quite heterogeneous when viewed as a thin section. Geological features that appear static at observed scales may be undergoing deformation at larger spatial or longer temporal scales (Shipley et al., 2013).

CONNECTING K–12 AND UNDERGRADUATE EDUCATION

If scale is an important concept for geoscientists, then it is important for learners (Dickerson et al., 2005; Johnson et al., 2013). Therefore, geoscience instructors should explicitly teach the two meanings of scale. The 2014 National Science Foundation-sponsored Summit on the Future of Geoscience Education and the 2015 Geoscience Employers Workshop emphasized that geoscience graduates must be able to engage in systems of thinking to understand how the various spheres of the Earth system interact at diverse spatial and temporal scales (Geoscience Employers Workshop Outcomes, 2015; Summit on Future of Undergraduate Geoscience Education, 2014). These reports highlight a number of concepts related to scale that graduates should grasp: deep time and its significance, the relationship between various Earth processes and the spatial and temporal scales at which they occur, feedback loops that occur at different scales, and how varying rates of change can have major impacts on Earth's systems, as is the case with anthropogenic climate change. In addition, graduates need to show proficiency with quantitative reasoning skills, including the ability to engage in proportional reasoning to understand scalar relationships (Geoscience Employers Workshop Outcomes, 2015). Both reports are summaries, which are necessarily terse. Neither distinguishes between the two meanings of scale in the way we do in this paper, but both meanings can be inferred from the concepts highlighted above.

The geoscience summit and employers' documents indicate that geoscientists, academic and professional, believe that scale permeates the discipline and should permeate the curriculum. Undergraduate geoscience educators prepare students for future careers in geoscience and/or graduate school, but that is not their only mission. Several geoscience literacy documents—the Atmospheric Science Literacy (Atmospheric Science Literacy Framework, 2007), Climate Literacy (U.S. Global Change Research Project, 2009), Earth Science Literacy Principles (Earth Science Literacy Initiative, 2010), and Ocean Literacy (National Oceanic and Atmospheric Administration, 2013)—developed through broad community consensus (Wysession et al., 2012) indicate that an understanding of scale is essential for scientifically literate adults. Although no individual would be expected to fully understand all concepts in the literacy documents (U.S. Global Change Research Project, 2009, p. 6), when viewed in aggregate, these documents emphasize that a scientifically literate citizenry should display proficiency with the same broad categories of scalar concepts as geoscience majors. Therefore, introductory geoscience courses that are taken to fulfill general education requirements, as well as upper division courses, should include a thorough treatment of scale. Despite this, none of the literacy documents distinguish between the two meanings of scale: magnitude and relationship.

Introductory geoscience courses also provide an opportunity to improve the knowledge and skills of future K–12 teachers. One way to do so is to incorporate the Next Generation Science Standards (NGSS) (NGSS Lead States, 2013) into course curricula and outcomes (Summit on Future of Undergraduate Geoscience Education, 2014). The NGSS articulate how science learning in three dimensions (disciplinary core ideas, science and engineering practices, and crosscutting concepts) can progress across the K–12 grade span. The NGSS and the Framework for K–12 Science Education (National Research Council, 2012), on which the NGSS are based, list *scale, proportion,* and *quantity* among seven crosscutting concepts. As is the case for the Summary Report on the Future of Undergraduate Geoscience Education, the Geoscience Employers Workshop Outcomes, and the various geoscience literacy documents discussed above, the Framework and the NGSS do not distinguish between the two meanings of scale as explicitly as we do in this paper. By including the terms *proportion* and *quantity,* they link learning about scale to formal mathematics instruction and echo the importance employers place on quantitative reasoning skills for geoscience graduates (Geoscience Employers Workshop Outcomes, 2015).

The Framework and the NGSS envision that learning about scale in K–12 should progress from concepts of relative scale in the earliest grades to a feel for quantity

(National Research Council, 2012, p. 90) at very small and very large magnitudes. To achieve those learning goals, crosscutting concepts like scale, proportion, and quantity must be revisited in increasingly sophisticated ways as learners move through K–12 education. The difficulty is that there is currently no learning progression for scale as magnitude or relationship across the K–12 span. To our knowledge, no such progression exists for undergraduate geoscience students either.

It may appear to be a simple task for learners to develop approximate, categorical ideas about scale as magnitude. One problem is that nominal descriptors for scale as magnitude (e.g., small and large) do not have the same meaning across contexts. Large-scale flooding does not imply the same spatial extent as large-scale glaciation. Large-scale geoscience phenomena occur over thousands of kilometers and millions of years, while an extratropical cyclone (large-scale atmospheric science event) can be 200 km in diameter and have a life span of about a week. Geoscience, chemistry, and physics all use *microscale* to denote the same spatial magnitude. By contrast, *microscale* in atmospheric science refers to atmospheric processes that occur on a spatial scale of at least 1 km. Geoscientists and atmospheric scientists know how terms are defined in their disciplines, so terminology is unlikely to be a problem for them. It could be confusing to K–12 students, their teachers, and undergraduate nonmajors taking Earth Science who finish an instructional unit on running water's role in surface processes one week and begin a weather unit the next.

Students may also find it challenging to develop quantitative understanding of scale as a relationship. Fractions, ratios, and proportional reasoning are difficult for many students. Factors within a problem, such as its context, can influence performance on proportional reasoning tasks (Tourniaire and Pulos, 1985). Faulty ideas about fractions and ratios could make it difficult for learners to grasp the difference between a large-scale and a small-scale map. A 1:125,000 map is a larger scale than a 1:625,000 map, because the former depicts a smaller geographic area than the latter. The ratios make clear which one is the large-scale map and which one is the small-scale map, but only if the student knows how to read and interpret a ratio. The words *large* and *small* used to describe the map's scale are consistent with fraction terminology (e.g., ninths are smaller than fourths). The difficulty is that fractions are problematic for many learners due to their tendency to treat fractions like whole numbers (Ni and Zhou, 2005). One might expect difficulties with fractions, ratios, and proportional reasoning would only be issues in K–12 education. Although proportional reasoning is taught in upper elementary and middle school classrooms, university students are often less adept at it than their instructors assume (Glaser and Riegler, 2015).

What do we know about science students' understanding of scale as magnitude or a relationship? What strategies might instructors use to help students develop more accurate conceptions of scale? To answer those questions, we began by reviewing studies in science education and DBER that either investigated students' current conceptions about scale or tested an intervention to improve those conceptions. Due to space limitations, we focused solely on papers that have implications for geoscience researchers and practitioners and address one or both of the definitions of scale used in this paper:

- Magnitude: spatial or temporal extent of an object or event that can be measured using standard or nonstandard units
- Relationship: spatial or temporal relationship between objects or events, which could be relative (e.g., larger or represented mathematically with a ratio, such as 1:125,000)

SCIENCE EDUCATION AND DISCIPLINE-BASED RESEARCH LITERATURE ON LEARNERS AND SCALE

We began to synthesize the literature on students' conceptions of scale in science by reviewing published peer-reviewed papers (including early views) from science education and DBER journals. We used Web of Science and searched for the terms *scale* and *student conceptions*. The term *scale* presents a challenge for any systematic review because instruments used to measure a variety of constructs are referred to as scales, despite not relating to scale as either the magnitude of a measurable dimension or a relationship. From the initial search, 24 references were retained. An additional 27 references were added based on citations within those papers. We excluded 9 of the 51 papers from our review because they were not focused on student conceptions of scale as magnitude or a relationship or did not report on an empirical study. This review of 42 papers related to learners' conceptions of spatial and temporal scale in science is not exhaustive but offers a broad view of recent research on students' sense of scale. Several of the studies retained for this review were not specifically designed to address scale concepts, but they report findings about how learners understand unfamiliar scales, so they were included.

Table I lists the studies reviewed for this paper by discipline and whether the study focused primarily on scale as magnitude, as a relationship, or both. The relative sizes of cells within Table I and the presence of several empty cells highlight the uneven approach to scale in the literature. Seventy-one percent of the papers solely or primarily investigated conceptions of scale as magnitude, though a number also reported on students' understanding of qualitative scalar relationships. Fourteen percent exclusively or primarily studied conceptions of scale as a relationship, and 14% devoted roughly equal attention to both meanings of scale. Forty-three percent of the studies came from a geoscience perspective. Seventeen percent were from a chemistry perspective, 7% were biology, 5% ($N = 2$) were astronomy, 2% ($N = 1$) were mathematics, and 2% ($N = 1$) were engineering. Twenty-six percent of the studies did not connect to an individual science discipline, so they were categorized as domain general. All of the latter were published in science education journals, where domain general research is common. Sixty-seven percent of the studies were geoscience or domain general, and in both cases, magnitude studies predominated. Fifty-seven percent focused on spatial scale, and all but one dealt only with one-dimensional magnitude. Forty-three percent focused on temporal scale. Seventeen percent tested a specific curricular or pedagogical intervention, while 83% investigated conceptual understanding. Samples for most studies were students in grades 6–8, those in grades 9–12, and lower or upper division undergraduates. Fifty-seven percent of the studies had samples of less than or equal to 100, and 79%

TABLE I: Scale emphasis (magnitude or relationship) and disciplinary focus of science education and DBER studies reviewed.[1]

Discipline	Exclusively or Primarily Magnitude	Exclusively or Primarily Relationship	Both Magnitude and Relationship
Geoscience	Cheek, 2016 (T, C, $N = 39$) Cheek, 2013a (T, C, $N = 35$) Cheek, 2013b (T, C, $N = 17$) Cheek, 2012 (T, C, $N = 35$) Clary et al., 2009 (T, I, $N = 25$) Delgado, 2013b (T, C, $N = 64$) Dickerson et al., 2005 (S, C, $N = 73$) Johnson et al., 2013 (S, C, $N = 63$) Jolley et al., 2013 (T, C, $N = 96$) Teed and Slattery, 2011 (T, C, $N = 108$)	Dodick and Orion, 2003 (T, C, $N = 509$) Lin et al., 2012 (T, I, $N = 69$) Trend, 1998 (T, C, $N = 189$)	Hidalgo et al., 2004 (T, C, $N = 169$) Libarkin et al., 2007 (T, C, $N = 63$) Nam et al., 2016 (T, C, $N = 43$) Trend, 2001 (T, C, $N = 51$) Trend, 2000 (T, C, $N = 179$)
Astronomy		Miller and Brewer, 2010 (S, C, $N = 83$) Schneps et al., 2014 (S, I, $N = 152$)	
Biology	Catley and Novick, 2009 (T, C, $N = 126$) Jones et al., 2003 (S, I, $N = 50$) Rundgren et al., 2010 (S, C, $N = 175$)		
Chemistry	Kruse and Roehrig, 2005 (S, C, $N = 45$) Nicoll, 2003 (S, C, $N = 56$) Yezierski and Birk, 2006 (S, I, $N = 719$)	Gerlach et al., 2014a (S, I, $N = 8{,}163$) Gerlach et al., 2014b (S, C, $N = 56$)	Johnson and Tymms, 2011 (S, C, $N = 4{,}450$)
Engineering		Swarat et al., 2011 (S, C, $N = 143$)	
Mathematics		Cox, 2013 (S, C, $N = 21$)	
Interdisciplinary	Delgado, 2013a (S, C, $N = 64$) Frändberg et al., 2013 (S, C, $N = 954$) Jones et al., 2013 (S, C, $N = 226$) Jones and Taylor, 2009 (S, C, $N = 50$) Jones et al., 2009a (S, I, $N = 19$) Jones et al., 2009b (S, C, $N = 17$) Jones et al., 2008 (S, C, $N = 66$) Lee et al., 2011 (T, C, $N = 514$) Tretter et al., 2006a (S, C, $N = 215$) Tretter et al., 2006b (S, C, $N = 215$)		Jones et al., 2011 (S, C, $N = 46$)

[1]First letter in parentheses indicates whether the study focused on spatial (S) or temporal (T) scale. Second letter denotes whether it investigated students' conceptions (C) or tested an intervention (I).

had samples sizes of less than or equal to 200. Only 5% ($N = 2$) employed samples more than 1,000.

Findings from science education and DBER studies are organized thematically, first for scale as magnitude and then for scale as a relationship.

Scale as Magnitude

Conceptions of Scale as Magnitude

Students from elementary through PhD, including those with visual impairments, estimate linear spatial magnitude better at human scales and worse at the extremes of scale, nanoscale to astronomical (Tretter et al., 2006a, 2006b; Jones et al., 2008, 2009a, 2009b). Typically, the magnitudes of small-scale objects are overestimated, while the sizes of large-scale ones are underestimated; however, one study found that college students both under- and overestimate the sizes of small objects (Gerlach et al., 2014a). Learners may misapply size categories from familiar to unfamiliar domains. For example, students often think groundwater exists in underground pools or lakes (Dickerson et al., 2005). Even those who are aware that groundwater can be found in pores and cracks sometimes describe those pores and cracks as basketball size or larger.

Accuracy of object-size estimation decreases smoothly as size increases to a very large spatial scale. In contrast, estimation accuracy for objects at the microscopic level or smaller is discontinuous. There is a propensity to judge objects at scalar extremes as more similar in size than they actually are (Tretter et al., 2006a, 2006b; Jones et al., 2009b, 2008). Those error patterns persist even when students use a model to support their explanations, specifically a scale model to illustrate astronomical distances. The farther an object is from Earth, the greater the estimation error (Miller and Brewer, 2010). Thus, models, in the absence of specific instruction about scalar relationships, may not support the development of concepts of scale. For the geoscience instructor who works to help students connect micro and macro phenomena, misconceptions about the magnitude of phenomena may occur at both ends of the spatial scale. Of particular concern would be cases in which errors cross important conceptual boundaries (e.g., believing that atomic packing determines sediment porosity).

Photographs of outcrops typically include objects, such as hammers, as spatial scale cues. Cues can also include size information that is not embedded in the photograph, such as a caption stating the height of the outcrop is 25 m and its base is 20 m from the camera. Adults (university undergraduates, graduate students, and faculty members) who are provided with similar nonembedded scaling cues more accurately estimate the sizes of features in digital images of geological outcrops than similar individuals who are not given scaling cues (Johnson et al., 2013). However, there is far greater variance in the size estimations by those who receive scaling cues. Underestimation errors are more common for those who do not receive scaling cues, while those who receive cues are more likely to overestimate the size of spatial features than individuals for whom no scaling cues are provided.

Temporal magnitude estimation studies deal primarily with participants' understanding of the geological timescale. Learners and teachers are inaccurate when assigning absolute ages (magnitudes) for past geological or biological events and can be off by several orders of magnitude (Trend, 1998, 2000, 2001; Catley and Novick, 2009; Delgado, 2013b). As is true for unfamiliar spatial scales, events further in the past are perceived as temporally closer to one another than they actually are and are plotted close together on a timeline (Libarkin et al., 2007; Catley and Novick, 2009). This trend is accounted for by the psychological theories discussed below but may reflect other influences on students' understanding of temporal scale, such as media portrayals of dinosaurs and humans coexisting or young Earth creationist worldviews (Libarkin et al., 2007).

Standard and Nonstandard Units as Magnitude Categories

Young children's early exposure to measurement often includes nonstandard units like blocks to measure linear space. Adults continue to use familiar referents as nonstandard measurement units to describe spatial scale, especially when measuring tools are not at hand. Units such as "room size," "me," or "small" serve as separate spatial categories along a linear dimension in which objects and distances can be grouped (Tretter et al., 2006a; Jones et al., 2008; Gerlach et al., 2014a). The self or human body remains an important linear spatial metric for all ages, which can be iterated to determine an object's length. Increasing age and expertise are associated with a greater number of size categories with more precise definitions, like "room size" or "microscopic" as opposed to simply "small" or "big." Large-scale categories are differentiated before small-scale ones. Even high school students include objects ranging in size from a pencil to the nucleus of an atom in the same spatial size group (Tretter et al., 2006a).

As is true for spatial scale, older participants in temporal scale magnitude studies report more categories than younger ones (Trend, 1998, 2000; Nam et al., 2016). In contrast to the spatial scale literature, however, temporal categories remain fairly broad (extremely ancient, moderately ancient, and less ancient) even for adults. The difference in category specificity in spatial versus temporal scale studies could be due to the nature of the samples in the studies we reviewed. Samples in several spatial scale studies included PhD science or science education students. In contrast, adults in Trend's (2000, 2001) studies were preservice or in-service elementary (primary) teachers. Doctoral geoscience or biology students might use well-defined temporal categories from their discipline to locate nondisciplinary events. To our knowledge, this question of whether disciplinary spatial and temporal categories are generalized to other domains has not been studied.

Perhaps temporal scale magnitude is simply more difficult to understand than spatial magnitude. Unlike the spatial magnitude studies, there is no good empirical evidence of the existence of a common temporal metric akin to the human body (Delgado, 2013b). Only a few temporal scale magnitude studies include events that occur on human timescales (Lee et al., 2011; Delgado, 2013b; Jolley et al., 2013; Cheek, 2012, 2013a, 2013b, 2016), so the potential value of any familiar, embodied event, such as day, year, or generations, is unclear.

Impact of Formal Mathematics Instruction on Magnitude Estimation

Ideas about scalar magnitude are closely linked to an understanding of the base-ten system, measurement, and proportional reasoning. Familiarity with linear metric measurement units is associated with better spatial size estimation by students and teachers (Tretter et al., 2006b; Jones et al., 2008, 2013; Delgado, 2013a), perhaps putting Americans, who must learn two measurement systems, at a disadvantage. It is unknown whether facility with the metric system aids learning about scale because it relies on base-ten proportional relationships, thus connecting the metric system to broader numeracy, or whether people who are asked to estimate spatial magnitude using metric units simply do better if they are familiar with the metric system. We did not find studies that specifically instructed students to estimate sizes using the U.S. customary system, so this is an open question.

In-service teachers (Jones et al., 2008) and visually impaired teenagers (Jones et al., 2009b) are more accurate at estimating sizes of objects at the extremes of scale, both large and small, when using metric units than when using nonstandard units like body lengths. When comparing across studies, the authors found that visually impaired students are better at the extremes of scale than their sighted counterparts. Jones et al. (2009b) hypothesize the latter finding is because visual information about very large and very small objects doesn't help learners develop an idea about their actual sizes because there is no scale referent, a point well understood by geoscientists who routinely include a scale referent in photographs.

Scientists often express very large and very small magnitudes with logarithmic scales, but engineering undergraduates are not uniformly adept at creating or interpreting logarithmic scales to display size data across a range of spatial scales. Even those who are able to construct log scales sometimes describe the linear scale representation as the real one and the log scale as a distortion (Swarat et al., 2011).

At a more fundamental level, some authors note that confusion about the numbers used in unfamiliar scales (Trend, 1998, 2000; Teed and Slattery, 2011; Cheek, 2012, 2013b) can account for some of the errors people make when estimating spatial or temporal magnitude. Learners who have experience with large magnitudes in one context may be able to transfer ideas about the relationship between them to other situations (Cheek, 2016).

Impact of Subject Matter Knowledge on Magnitude Estimation

Subject matter knowledge is related to better performance on tasks assessing understanding of scale as magnitude, but the relationship is not straightforward. High school biology students who are taught to use a remotely operated atomic force microscope to look at viruses show some increase in their understanding of microscale objects, as well as a shift from two-dimensional to three-dimensional representations of viruses, after a single week of instruction (Jones et al., 2003). Landscape architecture students who engage in a design task to create an informal space to illustrate geological time improve their knowledge of a timeline of Earth's events from pre- to posttest (Clary et al., 2009). Similarly, preservice teachers enrolled in an undergraduate introductory geoscience course show some gains in their understanding of geological time from the beginning to the end of the course, as measured by several items on the Geoscience Concept Inventory (Teed and Slattery, 2011). Chemistry undergraduates' scores on a magnitude scale literacy instrument and a chemistry content assessment are positively correlated (Gerlach et al., 2014b). However, the small-scale items on the scale literacy instrument were all ones that would be encountered in a chemistry course, which might mean the correlation reflects differences in overall disciplinary knowledge.

Conversely, Lee et al. (2011) found that the ability to estimate an event's temporal duration (magnitude) is weakly correlated with whether students said they knew the event's duration or guessed. Upper division undergraduate majors in biology or geology are only marginally better than lower division undergraduates (Catley and Novick, 2009) at estimating the absolute placement of biological and geological events in Earth's history, even though the nature of their estimation errors differs. Lower division undergraduates are just as likely to overestimate as underestimate the age of an event, consistent with Gerlach et al.'s (2014a) findings for microscopic objects. In contrast, upper division students tend to underestimate an event's age, similar to many of spatial magnitude studies described above (Tretter et al., 2006a, 2006b; Jones et al., 2008, 2009b). It is also akin to what occurs when students are asked to estimate vertical scale. High school students, who presumably possess less subject matter knowledge of geology than undergraduate and graduate students, tend to overestimate the typical depth of U.S. wells. Undergraduates and graduate students are more likely to underestimate their depth (Dickerson et al., 2005). Johnson et al. (2013) note the opposite tendency when adults are and are not, respectively, provided with relevant spatial information to help them estimate the size of specific features in digital images of outcrops. Those who are not given spatial cues tend to underestimate a feature's size, while overestimation errors are more common among those given spatial cues. It is unclear why greater knowledge would be associated with a propensity to underestimate magnitude in some instances but not others.

Experts in a range of fields (e.g., scientists, engineers, chefs, mechanics, and artists) report that a combination of in-school and out-of-school experiences helped them develop their understanding of scale as magnitude (Jones and Taylor, 2009). Space navigation, model creation, measurement activities, and unit conversions are all cited as activities that helped experts learn about scale. The ability to mentally convert units within the metric system appears to be useful, such as converting 1 million m to 1,000 km. There are inherent difficulties in asking people to reflect on past experiences that facilitated learning about scale, which the authors acknowledge. It is useful to know which activities aided people who now have a good conception of scale as magnitude to develop their understanding. Yet all of the experts had classmates who created similar models or completed the same unit conversion exercises but did not universally develop the same level of understanding about scale as magnitude. Why might that be so? We did not find any studies that addressed this question.

Connecting Spatial and Temporal Magnitude Scales: Temporal Duration

Spatial and temporal magnitude scales are related (National Research Council, 2012), but the relationship is not always as clear-cut as many learners think it is. Earlier, we noted that an understanding of spatial scale is important for understanding topics like groundwater (Dickerson et al., 2005). We argue that appreciating the temporal duration of a geoscience process is, in some cases, inextricable from understanding the process itself. Learning about temporal duration in stratigraphy is complicated due to variable depositional rates and diagenetic changes in sediments from the time of deposition to lithification. Learners often equate spatial size with the temporal duration of an event (Dodick and Orion, 2003; Cheek, 2012, 2013a, 2013b). In some cases, that thinking does not change even after being shown animations demonstrating various depositional rates but equal bed thickness (Cheek, 2013a, 2013b). Just as students often underestimate the age of an event on the geological timescale, they also underestimate its temporal duration and view nearly instantaneous catastrophic events, such as volcanic eruptions or earthquakes, as responsible for nearly all geomorphic features (Hidalgo et al., 2004). This might be because media reports of geological events focus primarily on those of short duration. It may also be because it is simply impossible to experience many geological processes in real time, so students have no meaningful referent for events of long duration analogous to the human body for estimating spatial size.

As is true for spatial size estimation or placing events on the geological timescale, university undergraduates overestimate the durations of events that occur on timescales too short to be perceived by humans (e.g., typical neuron response to stimulus) and underestimate the durations of events with long time spans (e.g., mountain formation) (Lee et al., 2011). While they generally do best at estimating temporal durations of events at human timescales, they are not universally good at those scales. Estimating the amount of time for a sunspot cycle appears to be more difficult than either the time for the fastest chemical reactions or the time since dinosaurs became extinct to the present, which could be explained by how easy or difficult it might be to use logical reasoning to deduce durations of particular events.

Similarly, lower and upper division undergraduate students accurately estimate the durations of geomorphic events that occur over short, human timescales (Jolley et al., 2013). They also correctly estimate durations of events that occur over long time spans, such as mountain building. However, they struggle with intermediate duration events (e.g., hoodoos and alluvial fans). Intermediate time span events are also the ones about which there is more likely to be expert disagreement about the time necessary for their formation.

Scale as a Relationship

Conceptions of Relative Scalar Relationships

Students from elementary school through undergraduates and in-service teachers do reasonably well placing a series of geological and biological events in correct temporal order and place the oldest events the most accurately (Trend, 1998, 2000, 2001; Hidalgo et al., 2004; Libarkin et al., 2007; Catley and Novick, 2009; Nam et al., 2016). It is unclear whether this represents actual knowledge of the temporal relationships between events or logical reasoning, i.e., Earth had to form before the first life could appear on Earth's surface. Events that do not lend themselves easily to logical reasoning, such as the extinction of trilobites, tend to show far greater variability in their relative placements. The Framework states students' ideas about scale should progress from contextual generalities of magnitude and scalar relationships to quantitative conceptions as they move through K–12 education. The research discussed above indicates conceptions of spatial magnitude seem to progress over time, but we did not find a similar trend in ideas about scale as a relationship for either space or time.

Impact of Formal Mathematics on Understanding of Scale as a Relationship

Scaling tasks, like constructing and reading maps or creating scale models, where there is a focus on the scalar relationships within and between objects, are ways to learn about scale as a relationship. Scaling problems in mathematics are often taught numerically using scale factors. In order to successfully complete scaling problems, students must realize they need to multiply or divide by the scale factor rather than add or subtract, but this concept is not easily mastered by all when completing geometric (Cox, 2013) or temporal (Cheek, 2012, 2013b) scaling tasks. Working with scale factors in the metric system is more straightforward than it is in the U.S. customary system (Delgado, 2013a). Scores for middle school students on an instrument that tests their ability to engage in scaling problems are highly correlated with their scores on an instrument assessing their ability to engage in logical thinking, including proportional reasoning (Jones et al., 2011).

Geometric scaling problems for complicated figures where one shape is embedded in another are more difficult for learners than typical scale factor problems, because the learner must preserve not only the proportional sizes of the objects but also the scalar relationships of the spaces between the figures (Cox, 2013). Thus, teaching proportional reasoning solely through the use of simple scale factor problems may be insufficient for the type of scaling problems encountered by geoscience learners.

Impact of Subject Matter Knowledge on Understanding Scale as a Relationship

We noted above that greater subject matter knowledge is related to better performance on tasks assessing understanding of scale as magnitude. The same can be said of scale as a relationship. Adults do better ordering temporal events (Trend, 1998, 2000, 2001) and the sizes of objects (Tretter et al., 2006a, 2006b) than children, which Trend hypothesizes is due to the former's greater knowledge and experience. After using a computer simulation to learn about relative geological dating, high school students improved their ability to use principles of relative dating to order geological events (Lin et al., 2012). The ability to construct correct explanations of geological processes appears to be facilitated by time spent comparing micro- and macroscopic images embedded in the software. Earlier, we noted that upper division undergraduate majors in biology or geology are only marginally better when estimating the absolute placement of biological and geological events in Earth's history (Catley and Novick, 2009); they are also only slightly better at estimating relative placements of those events.

Experts report mentally transporting themselves to the spatial scale at which they work and mentally moving back and forth across scales (Tretter et al., 2006b; Jones and Taylor, 2009). Learners who can visualize chemical phenomena at atomic or molecular levels appear to have better conceptual understanding of those phenomena (Yezierski and Birk, 2006). Yet people who work at the nanoscale are not particularly good at estimating the sizes of objects at astronomical scales, suggesting little transfer across scales (Jones and Taylor, 2009). Chemistry students and teachers appear to have difficulty moving back and forth from macro- to microscopic levels, and students may not progress much in their understanding of submicroscopic phenomena from lower to upper division undergraduate courses (Nicoll, 2003). Middle school students (Frändberg et al., 2013) and some biochemistry majors (Rundgren et al., 2010) employ macroscopic explanations for chemistry concepts like conservation of matter or osmosis through a cell membrane. Even science teachers (Kruse and Roehrig, 2005) have trouble translating between the macro- and the microscopic properties of substances.

The way in which information about scalar relationship is conveyed seems to affect conceptual understanding. After using a virtual three-dimensional simulation in which they could pinch and zoom to explore the solar system, high school students who use a true-to-size simulation make greater gains on a concept inventory that includes scale questions than students who use a version of the simulation that exaggerates planetary sizes relative to their orbits (Schneps et al., 2014). These results raise a question about whether the use of vertical exaggeration of topographic profiles in lower division undergraduate geoscience labs helps or hinders the development of students' ideas about spatial scale.

When, or how, learners integrate their subject matter knowledge with ideas about scale as a relationship is an open question. Johnson and Tymms (2011) report on the development of a learning progression for the concept of a substance in chemistry. In their view, learners do not develop robust ideas about scale in chemistry, particularly that the particles in a substance do not display macroscopic properties, until fairly late in the learning progression. It is not even clear when learners begin to attend to scale when learning science (Nicoll, 2003). The extent to which scalar concepts might develop later rather than earlier in subjects like geoscience is simply unknown.

Summary of Significant Findings From Science Education and DBER Literature on How Learners Understand Scale

- Learners' concepts of scale are nonlinear when outside the realm of their experience. Errors may differ from scientific consensus by several orders of magnitude.
- The sizes of objects and durations of events that are smaller or shorter than human scale are typically overestimated. The sizes of objects and length of events larger or longer than human scale are generally underestimated. However, contrary findings have been reported.
- People sort the magnitudes of objects or events into categories, either standard measurement units or linguistic descriptors that they treat as nonstandard units. Educational level is associated with a greater number of more precise categories. Temporal categories are less well defined for adults than linear, spatial size categories.
- While relevant educative experiences are positively associated with a better understanding of scale as magnitude and as a relationship, more coursework does not necessarily result in better concepts of scale as magnitude or a relationship.
- Facility with the metric system, proportional reasoning, and basic numeracy affect learners' ability to understand scale as magnitude and as a relationship.
- Many learners see the relationship between spatial and temporal scales as simpler than is usually the case.
- The literature provides virtually no insight into how or when learners develop ideas about scale as they move through the educational system.

The science education and DBER literature provides a useful starting point for geoscience (and other science) educators, but it may not provide the kind of data needed to help learners acquire the skills necessary to do more complex tasks, such as field mapping. We acknowledge the challenge in designing research that investigates how students learn about scale in the context of disciplinary core ideas and provide specific suggestions for future research later in this manuscript.

RELEVANT INSIGHTS FROM PSYCHOLOGICAL RESEARCH ON LEARNING ABOUT SCALE

Research in developmental and cognitive psychology provides a perspective on how humans learn about scale as magnitude and as a relationship across the life span. This section is not a comprehensive review of all psychological literature regarding how people learn about scale. The task would be far beyond the scope of this paper. Instead, we highlight psychological literature and theoretical perspectives that we consider most germane to learning about scale in geoscience. Where appropriate, we note alternate theoretical perspectives, along with a few key references for readers who wish to investigate further.

Psychological Perspectives on Scale as Magnitude

As the science education and DBER literature surveyed indicates, humans are good at reasoning about scales of objects and events that are familiar, but they have trouble with unfamiliar magnitudes. A search of psychological research about scale as magnitude yields many studies on how people estimate number magnitude. Error patterns in the science education and DBER studies described above mirror the types of errors seen in number magnitude estimation research. When reasoning about relative values of numbers with unfamiliar magnitudes, both children and adults tend to overestimate the value of smaller numbers relative to larger ones (for example locating 230 million roughly halfway between 0 and 1 billion). Children typically show this pattern for smaller magnitudes (thousands) (Siegler and Opfer, 2003), some adults show the pattern for larger magnitudes (millions or billions) (Landy et al., 2013). Experience with large numbers appears to be more related to improvement in number magnitude estimation than age (Dehaene et al., 2008). Non-Western indigene adults whose language does not contain number words for large quantities exhibit number mapping that is more similar to that of children than to that of adults from cultures whose language contains words for large numbers.

Multiple theoretical accounts have been suggested that offer contrasting explanations for these error patterns. A theoretical model that is useful to geoscience educators wanting to teach about scale should provide concrete suggestions instructors can use with their students. The category adjustment model (Huttenlocher et al., 1988, 1991) does so and is discussed more fully below. Two other theoretical models are sometimes cited in science education or DBER literature about scale. Space does not permit discussion of these theories or their relative merits, but readers are referred to the relevant citations as starting points. One of these models posits that number magnitudes are represented using two systems: logarithmic and linear (Dehaene, 2003; Siegler et al., 2009). Magnitudes that are familiar are represented linearly, while those in unfamiliar ranges are represented logarithmically. The second hypothesizes that magnitude estimation is essentially a proportion judgment task (Barth and Paladino, 2011). Large magnitudes are undervalued because the proportional relationships between quantities are unknown.

While the logarithmic to linear shift and the proportion judgment theoretical models predict behavior, we do not think they provide the kind of assistance that will help geoscience educators who want to teach students about scale. The category adjustment model, developed by Huttenlocher et al. (1991, 1988), hypothesizes that humans combine categorical and metric information to recall spatial location. For example, suppose one is trying to locate the car keys. One may not recall exactly where in space they were placed, but one may remember being in the dining room (category) when the keys were put down. Remembering that the keys are in the house (a larger category) is better than nothing, but does not help as much as remembering they are in the dining room. Thus, precision of recall depends heavily on the relative size of the categories, an important point for a geoscientist wanting to recall the location of specific features, such as faults in the field. The category adjustment model accounts for many of the magnitude estimation errors cited in science education and DBER studies. With unfamiliar numbers, times, and distances, people generally possess few categories, and those categories encompass a large extent of a dimension (e.g., "really long ago"). Thus, it is unsurprising that students consider objects ranging in size from the nucleus of an atom to a pencil to all be small objects (Tretter et al., 2006a) or adults view time periods across many orders of magnitude as all extremely ancient (Trend, 2000). Category adjustment predicts older learners would estimate very large and very small spatial and temporal magnitudes more accurately than children since they possess a greater number of metric categories that are more precisely defined, which is confirmed by the science education and DBER literature. Category adjustment could also account for why estimating temporal magnitude appears more difficult than spatial magnitude. There does not appear to be a common temporal analog to body length, and it is impossible to experience very large temporal magnitudes other than to mentally imagine them. That could explain why temporal categories are less well defined for adults than spatial ones (Trend, 2000; Cheek, 2012; Nam et al., 2016; Resnick et al., 2017).

Psychological Perspectives on Scale as a Relationship

In the psychological literature, scale as a relationship is more commonly referred to as a proportional transformation. We continue to use scale as a relationship to make clear that we are referring to how learners reason both formally and informally about scalar relationships. Students are normally introduced to formal proportional reasoning in middle school but are expected to reason informally about scale relationships at younger ages (National Research Council, 2012). People typically learn about scale through analogies, which are indirect ways to acquire concepts that cannot be directly experienced.

Analogical reasoning refers to the process of identifying and learning by using representations and structures shared between two domains (Markman and Gentner, 1993; Gentner and Smith, 2012). Within geoscience, as noted above, much learning involves magnitudes and proportional relationships that are outside learners' experience or cannot easily be brought into the classroom. Thus, geoscience instructors frequently use physical models that require students to engage in analogical reasoning about scale.

An effective analogy enables learners to draw inferences about a new domain from what they know about a familiar one (Gentner, 1983; Gentner and Smith, 2012). Learning in this way requires that the two domains be aligned so that what is known in one domain can be applied to the other. For example, one might be able to make inferences about a new type of animal using what is known about other similar animals. To make those inferences, learners compare and align the relational structure of the familiar (base) domain with the new (target) domain. This core cognitive process of analogical reasoning is known as structure mapping (Gentner and Smith, 2012). Suppose a young child sees an elephant for the first time. The child can reason that the large objects on the side of the elephant's head are ears because there are two of them, they look like mirror images of each other, and they are in roughly the same location as the family dog's ears. Structure mapping provides a well-developed account of how humans align domains, which is a powerful learning mechanism for spatial thinking, including reasoning about scale (Gentner, 1989; Markman and Gentner, 1993; Pozzer and Roth, 2003; Gentner et al., 2007; Gentner et al., 2011). Learners can use a mental model from the base domain to simulate events in the target domain to develop their understanding, make predictions (Holland et al., 1986), and create new mental models of the target domain (Collins and Gentner, 1987).

Using Analogical Reasoning to Learn Scale in Geoscience

Scale as Magnitude

To understand an object or event that is too big, long, small, or fast to allow direct experience, teachers often employ an object or event that can be brought into the classroom and directly experienced. The use of such scaled models requires students to draw an analogy by relating (mapping) the model object or event to the object or event that exists in the world at a different scale. Students use the scale model or map to represent a large object or event with something smaller, such as a clay model of the relative thickness of Earth's layers or a topographic map. Conversely, a small object or event could be represented with something larger, such as a ball-and-stick model of atomic packing in a crystal. Models preserve some properties of the physical objects they represent but not others, an important consideration when using them in the classroom. For example, a satellite image preserves the optical structure of the imaged region but not its material properties. The relationships that are preserved from the model (familiar domain) to the target (real-world domain) are the ones learners can use to make inferences about the new domain (Gentner, 1989).

Research on analogical reasoning and the category adjustment model may be combined to provide suggestions for how instructors can help geoscience learners reason about previously unfamiliar spatial and temporal magnitudes by drawing on what has been learned about smaller (familiar) magnitudes (Boyer et al., 2008; Thompson and Opfer, 2010). By appropriately aligning the elements of knowledge in two domains (say, two timescales), humans are able to use structure mapping (see, e.g., Gentner, 1989) to make inferences by using what is known in one domain (a familiar, human timescale) to make predictions in the other domain (an unfamiliar timescale). Using analogies enables a learner to extend a scale by aligning familiar portions of the scale to the novel magnitude and thereby adjust the magnitude categories to be more distinct (rather than simply "a really long time ago").

Learners are better able to engage in analogical mapping when instructors highlight correspondences across the two domains (Gentner et al., 2016), and such guidance appears to help students learn unfamiliar number magnitudes (Thompson and Opfer, 2010). The more similar the two domains, the easier and faster the analogical mapping (Gentner and Markman, 1994, 1997). Large or small magnitudes may be difficult for students because they are so dissimilar to students' everyday experiences with magnitude. This implies that starting with small differences and then moving to larger ones is likely to lead to better learning than jumping immediately from human scale to very large or very small magnitude scales. Students also benefit from multiple opportunities to align scales. Resnick et al. (2017) found that an approach based on successive scale increases was beneficial to students learning to reason about geological time magnitudes. Students given multiple opportunities in class to align the traditional geological timeline with a linear scale showed significantly reduced errors in locating numbers on the billions scale, suggesting the practice helped them form a rich representation of the magnitude of billions that could be generalized across scales. It is not known whether a similar approach would help students more accurately estimate objects and times at the micro- and nanoscales.

Scale as a Relationship

How do these constructs apply to learning about scale as a relationship? To use a map, one must align the map to the territory by relating elements of the map such as mountains and rivers to the corresponding topographic features, sometimes by turning the map to correspond to one's current spatial perspective. Having achieved the alignment, one may then make inferences about the physical world from the map. Spatial relations of objects in a territory to locations on the map are related by a specific proportion, one often set by convention. In order to use a map to reason about a territory, learners need to understand which spatial relations are preserved on the map (e.g., angles defined by three locations), which ones are related by a scaling transformation (e.g., metric distance between two locations), and which are not preserved (e.g., surface texture at a location).

Consistent with the science education and DBER literature reviewed above, proportional reasoning is associated with mathematical fraction understanding (Möhring et al., 2016) and is not universally mastered by adults (Ni and Zhou, 2005; Glaser and Riegler, 2015). Recent research on proportional reasoning offers some insights into how the mind may make the mental transformation required to understand scale relationships between models, including maps, and objects or events they represent. When using maps to reason about physical space, learners demonstrate linear increases in errors and response times with increasing scaling magnitude, suggesting an analoglike imagery strategy (Möhring et al., 2014, 2015). An analoglike imagery structure could also account for difficulties with temporal scaling transformations (Cheek, 2012, 2016; Delgado, 2013a). The pattern of data (linear increase in errors and response times) is similar to that seen for mental rotation. Although the two skills appear to be correlated (Frick and Möhring, 2016), the extent of overlap in underlying cognitive processes is unclear. These findings highlight the potential need to adapt supports for students in a single classroom with varying skills in proportional reasoning, even at the undergraduate level.

Successful proportional reasoning requires successful structure mapping. One of the key findings in this area is that proportional reasoning is best supported when the proportions are presented in such a way that the part–whole relationship in the proportion is salient (Möhring et al., 2016) and the parts of the whole do not have visual subdivisions (Boyer et al., 2008). For example, a stacked bar of one-third red and two-thirds green emphasizes that red is one-third of the whole, whereas side-by-side red and green bars make it less clear there is a 1:3 ratio of red to the whole. Similarly, a problem can be challenging when individual units are indicated (e.g., 1 unit of red and 2 units of green), because students are apt to attend to the part–part (red to green) relationship rather than the part–whole one. Unfortunately, mathematics worksheets for students who are learning fractions often contain images that highlight part–part relationships, possibly making the learning process more difficult (Möhring et al., 2016). Furthermore, the finding that humans misapply whole-number equivalence concepts to proportional reasoning problems (Boyer et al., 2008) suggests that structure mapping matters because people reason about magnitudes in two fundamentally different ways. An approximate number system that processes relationships among multiple objects (extrinsic relationships) appears to facilitate reasoning about whole numbers (Dehaene, 1997; Feigenson et al., 2004). In contrast, proportional reasoning considers relationships within a single object (intrinsic relationships) (Matthews et al., 2016). When learners apply an approximate number system for whole numbers to problems better served by an approximate ratio system, errors may occur. Anecdotally, one author has witnessed many K–12 students (and preservice elementary teachers) say 5/7 is equivalent to 7/9 (because $5 + 2 = 7$ and $7 + 2 = 9$. i.e., "do the same thing to the top and the bottom"). If proportional reasoning or scaling problems are presented in a way that makes salient the individual objects in one set or the other rather than the part–whole relationship, they may invoke reasoning better suited for whole numbers than for ratios. Both points raise questions about how students interpret typical diagrams in introductory textbooks that show Earth's layers in cross-section if the diagram indicates the depth of each layer. It raises the possibility that students who have not fully mastered proportional reasoning will notice the relative thickness of Earth's layers but may not attend to how the thickness of a particular layer relates to the whole. To our knowledge, this has not been studied.

A Typology of Classes of Spatial Reasoning Skills

Recent research in neuroscience and psychology suggests that differences in learning about scale as magnitude and scale as a relationship (proportion), as well as variance in learning about spatial and temporal scales, may reflect two underlying dissimilarities in how people reason about spatial relations (Chatterjee, 2008; Newcombe and Shipley, 2014). A typology, based on that research, hypothesizes that spatial skills and concepts can be categorized into four types based on two dichotomies. One dichotomy contrasts intrinsic object relations (spatial properties of a single object) with extrinsic object relations (spatial properties among multiple objects). The second dichotomy deals with whether the spatial relationship is considered static or dynamically transformed (e.g., when mentally rotating an object in space or mentally navigating through a scene).

This typology is useful for understanding how learners reason about scale, because it provides a framework within which to situate the distinction between scale as magnitude and scale as a relationship. Static intrinsic spatial representations apply to objects with measurable dimensions. These spatial properties define the object and enable us to distinguish one object from another. In contrast, static extrinsic spatial relationships situate multiple objects in a particular spatial location, highly useful for navigation. Dynamic intrinsic representations involve events defined by motions within objects, such as the change in shape of rock layers. Dynamic extrinsic representations include events in which the movement involves multiple objects, i.e., changes in perspective. We can use this dichotomy to consider research on learning scale as magnitude as the intrinsic relationships inherent within an object and scale as a relationship as the extrinsic relationship between objects, either of which could be static or dynamic. In considering this distinction, it is important to keep in mind that the way an object is defined depends upon the observer—a person may be viewed as one complete organism or a spatial array of organs. Instructors can use the typology to design and choose appropriate models that will facilitate analogical reasoning and learners' ability to refine their own spatial and temporal categories to better understand scale. We discuss how to do so in the next section.

Summary of Insights From Psychological Literature on Learning Scale

- The category adjustment model indicates that instruction on scale as magnitude should progress from familiar portions of a scale to larger or smaller unfamiliar magnitudes and that learners benefit from multiple opportunities to reason about new portions of a magnitude scale. Learning is enhanced when students are given information about new portions of the scale that can serve as categories (e.g., events on a timescale or locations on a distance scale).
- Analogical reasoning involves the process of mapping important structural aspects of a model to the actual thing being modeled. Students are not equally adept at structure mapping or the requisite mathematical skills (especially fractions) to use maps or models to engage in proportional reasoning. Some students may need more support aligning the model and referent than others to learn scale as a relationship.
- The structure mapping that a learner uses to solve a scaling problem affects the likelihood that the problem will be solved correctly. If scaling problems are presented in a way that emphasizes individual objects rather than the part–whole relationship, errors are more apt to occur.
- Objects and events have both intrinsic and extrinsic spatial properties that are reasoned about differently. Differences in learning about the two meanings of scale, or spatial versus temporal scales, may reflect differences in underlying cognitive processes.

USING RELEVANT RESEARCH TO IMPROVE TEACHING ABOUT SCALE

The science education, DBER, and psychological literature highlighted above indicates learners benefit from explicit support to acquire concepts of scale as magnitude and as a relationship. The challenge is how to accomplish that within an already crowded curriculum. We think K–16 instructors would do well to follow the approach advocated in the NGSS (NGSS Lead States, 2013) and infuse concepts like scale throughout instruction. This means emphasizing scale in the context of many geoscience topics and not solely when discussing geological time. This can be done through the use of well-chosen physical models that facilitate analogical reasoning and help learners refine the categories they use to estimate spatial and temporal magnitude.

Physical models are ubiquitous in geoscience education precisely because many natural processes occur at spatial and temporal scales that cannot be experienced in the classroom (Kastens and Rivet, 2008). The ways learners use physical models to construct understanding may be different from expert model use. The cognitive science literature on analogical reasoning discussed above should guide instructional use of models. The primary instructional goal of some physical models in geoscience education, such as those for geological time, is to teach scale as magnitude, but understanding geological time also requires understanding scale as a relationship. Other common geoscience models are not typically used to teach scale, but they could be. Where there is relevant research, we discuss how physical models of both types might be improved in light of the science education, DBER, and psychological literature on scale learning.

An Object–Event Framework for Characterizing Physical Models to Teach Scale Concepts

In considering the variety of physical models employed in geoscience education, we hypothesize, based on the spatial reasoning dichotomies described above, that different models may have different pedagogical value and may result in different learning due to the different psychological processes likely to be engaged in reasoning about the model and real-world referent. We focus here on using models to learn about static and dynamic phenomena, because we find this dichotomy particularly germane to the use of physical models in geoscience education. Four fundamentally different types of models can be described within this framework that distinguishes between static and dynamic models and static and dynamic referents (real-world phenomena). For this paper, we use the terms *object* to refer to a physical model or referent that is static and *event* to denote a model or referent that changes over time (is dynamic). At present, the relative merits of the different categories are unknown. Thus, we offer this typology as a potential guide for a research program to put learning with physical models generally, and learning about scale specifically, on a sound theoretical foundation.

From the perspective of analogical learning, models may be powerful ways to develop base knowledge that can be extended to real-world domains (Gentner, 1983). Furthermore, better learning will occur by supporting students aligning the model and referent (Gentner, 1989). Thus, one might expect objects to serve as physical models for real-world objects and, conversely, events to serve as models for real-world events. However, an object can also serve as a physical model for an event and an event can serve as a model for an object. In some cases, a physical model may be used as either an event or an object, depending upon the instructional goals and how the model is used (see discussion of event models below). When a model and its referent do not match, one might anticipate additional cognitive difficulty in aligning model and referent due to their dissimilarity (Gentner, 1989). Furthermore, some have

TABLE II: A 2 × 2 matrix of object–event models and their object–event referents.[1]

	Object Referent	Event Referent
Object Model	Scaled distances of planets from the Sun and one another marked on adding machine tape (MS-ESS1-3) (Flynn, 2007) http://serc.carleton.edu/sp/mnstep/activities/planets.html	Toilet paper, rope, or paper timeline of geological events to show positions of significant events on the geological timescale (MS-ESS1-4) (Atkins, 2013). http://serc.carleton.edu/sp/mnstep/activities/19783.html
Event Model	Walking to predetermined distances from one another to show relative distances of planets from one another and the Sun (MS-ESS1-3) (Muhammed, 2016) http://serc.carleton.edu/sp/mnstep/activities/35914.html	Running a stream table to model erosion and deposition processes (2-ESS2-1, 4-ESS2-1, MS-ESS2-2) (Friesen, 2005) http://nagt.org/nagt/teaching_resources/teachingmaterials/9271.html For a stream table model appropriate for elementary school, see Cheek (2013c)

[1]We have noted in parentheses which NGSS performance expectations (PEs) are addressed by each model at the elementary school (K–5) and middle school (MS) levels. There are no PEs at the high school level related to these models (NGSS Lead States, 2013).

argued that events are particularly difficult to align with one another due to their transient nature and/or the difficulty in reliably deconstructing events (Zachs and Tyversky, 2001; Gentner, 2006; Maguire et al., 2011). Thus, one might hypothesize that events would be harder to learn from and learn about. For example, to reconstruct the geological history of a location, experts and students must use observations about static present-day objects to make inferences about dynamic geological events.

Table II shows common physical models used in K–12 and undergraduate geoscience classrooms. Models have been sorted into cells based on the object or event categorization of the model and the referent. The object referent, the solar system, is commonly encountered in science in the context of large distances, and the event referent, geological time (events from Earth's formation to present day), is commonly encountered in the context of large temporal intervals. For all four types of models, the primary instructional goal is to teach scale as magnitude, though instructors may also encourage generalizations about scalar relationships, which could be used to teach proportional reasoning for scale as a relationship.

Using Object Models to Teach About Scale

We provide examples of two object models, one to teach about an object, the solar system, and a second to teach about an event, geological time. For each model, we briefly describe how instructors might employ the physical model (for more details, the references provided with each model in Table II include materials developed to help teachers use these physical models in their classrooms), discuss why the models may be challenging for learners, and then suggest ways geoscience educators might improve instruction when using the models based on the science education, DBER, and psychological literature discussed above.

Object–object scale models generally use long objects, such as toilet paper, beads on a string, or adding machine tape, to help students grasp the large sizes and distances in space. For example, an undergraduate Earth Science lab instructor might provide student dyads with a 4-m long piece of adding machine tape, meter sticks, and a data table listing the eight planets and average distances from the Sun. Students are instructed to mark the Sun at a predetermined location on the adding machine tape (e.g., 10 cm from one end). They use a data table with average distances and a predetermined scale (e.g., 1 mm = 100 million miles) to decide how far to place each planet, the asteroid belt, and the Kuiper belt from the Sun. Students calculate the scalar transformation from actual solar system distances to distances on the adding machine tape. A middle school Earth Science teacher might use the same model in the same fashion, or the teacher may decide to calculate the scaled distances in the model for students. This removes the initial challenge of converting units within the metric system, which could help students who struggle with either metric measurement units or unit conversions.

The same approach has been used to convey the magnitude of time since Earth's formation, except the long objects are related to the event: the history of Earth. A high school Earth Science teacher who wishes to help students comprehend the immensity of geological time might have students make their own timeline using strips of adding machine paper. Groups of students measure off a 4.5-m length of paper and mark the location of significant geological and biological events on the paper, with the present at one end and the formation of Earth at the other. As students work, the teacher circulates the room to make sure students have figured out that 1 m = 1 billion years. Later, students transition from measuring in units of billions of years to million-year units and ultimately to thousand-year units, which are barely visible on this scale. The final product is an object–event model relating the locations on an object to events that occurred in geological time.

Students face a number of challenges when using these models. There are the mathematical challenges with student-created models of both space and time. In the U.S., facility with unit conversions within the metric system may be difficult for students and their teachers (Jones et al., 2008). Representing large magnitude differences on the same scale can also be challenging for younger learners (Harel and Confrey, 1994) and even some undergraduates (Cheek, 2012; Landy et al., 2016). Because some people mistakenly treat thousand, million, and billion as an additive count sequence, it may be important to specifically discuss how those numbers are mathematically related (Landy et al., 2013). Finally, interpolation within base units is needed (e.g., where to place 230 million years on a 0–1 billion year timescale).

Learners must also be able to use part–whole reasoning to divide the space to represent the entire scale. Both models use linear space to represent either large spatial or long temporal scales, a common practice in many of the science education and DBER studies reviewed for this paper. Learners face some similar analogical reasoning challenges when interpreting the models, even though one model deals with spatial scale and the other deals with temporal scale. One is aligning familiar scales to unfamiliar (astronomical or geological) scales. Spatial and temporal patterns in the small-scale model are analogous to the patterns in the large-scale referent, as long as students think about part–whole relationships. On a scale in which 1 m = 1 billion years, a learner must be able to see a physical distance that is only a fraction of the length of the classroom as representing an incredibly long time period. Meanings for terms used to describe patterns in the physical model may be different from meanings ascribed to those same terms in other courses. Events described as ancient history in a World History class barely differ from present day in a geological time model. For instructors, this may be as simple as noting that the categories ancient human history and ancient geological history are vastly different.

Aligning scales is more challenging with increasingly greater scalar differences between the physical model and the referent space (Möhring et al., 2014). Viewing the imperceptible difference between today and the appearance of modern humans on the adding machine paper model as a similarly tiny difference on the geological timescale is challenging. Rather than expecting students to make sense of large spatial or long temporal scales using a single model, research suggests that it would be more helpful to use a series of bridging analogies. As noted in a previous section, Resnick et al. (2017) applied the category adjustment model (Huttenlocher et al., 1991) to improve undergraduates' understanding of the geological timescale. Students who aligned a series of temporal scales onto a timeline that progressed from short, familiar scales to longer geological ones improved their ability to represent geological time on a linear scale. Nesting all previous scales within the next new scale may have enabled students to see scalar relationships. Improvements in magnitude estimation were seen after a single intervention and were durable. Resnick et al. (2017) used domain-specific scalar categories (divisions on the geological timescale). Numbers alone could also serve as categories (Landy et al., 2016), which would be similar to zooming in on or out from a scale (Jones et al., 2011), as in the *Powers of 10* film (Jones et al., 2007). Might instruction designed to improve general understanding of the magnitude and relationships between large numbers help learners align familiar spatial and temporal scales to unfamiliar ones in both astronomy and historical geology? It is a question worth investigating.

Using Event Models to Teach About Scale

Any dynamic model that is used to represent a real-world object or event is an event model. Runnable physical models and embodied activities (e.g., gesturing with one's body or hands to represent an object or event) (Atit et al., 2014) are common examples of dynamic analog modeling for geological processes. The event models in Table II illustrate how an embodied activity is an event model analogous to the object models of astronomical distances and geological time. An elementary school teacher may want to help the children in her class develop some understanding of relative distances in the solar system. She takes the class outside to the playground, where children assume the role of various solar system objects. One student is the Sun. Others are each of the eight planets. Each planet walks a specified number of baby steps from the previous planet and stops. The children who are not part of the model stand near the Sun with the teacher.

A middle school Earth Science teacher may choose to model the geological timescale using a 1,000-sheet roll of toilet paper. Prior to class, he marks off significant events in Earth's history on the toilet paper and then rerolls it. Because this is a teacher-created model, students do not have to do any mathematical computations to create the model. One student takes the loose end, and another holds the roll. The first student walks around the room, unrolling the toilet paper and reading off significant events in Earth's history as the paper is unrolled.

These are event models because the pedagogy focuses attention on the students' bodies physically moving through space, as opposed to focusing on the static marks on adding machine tape or toilet paper. Both models remove the metric measurement challenges inherent in the object models described above. The only mathematical challenge in the solar system model is maintaining an accurate count of the number of steps taken. The elementary school model shows magnitude imprecisely, because it depends upon children's interpretation of baby steps. It does, however, enable children to make generalizations about how distances among objects vary in the solar system. In both models, participants' perspectives can vary widely. In the toilet paper geological time model, generally only one student walks through space to simulate the passage of time while classmates watch. The solar system model is experienced quite differently by children who are part of the model than those who are not or even by those who take the role of inner versus outer planets. If being able to flexibly move back and forth from an Earth-based to a space-based perspective facilitates conceptual understanding of astronomy (Plummer, 2014), then running the model several times so that children have the opportunity to take different roles may increase its effectiveness. Although challenging for students to mentally move to the astronomical or geological scale, experts report they do so when reasoning about disciplinary problems (Jones and Taylor, 2009). Spatially, the alignment task when using these models is similar to the mental transformation required when using a map to determine the location of an object in a larger, referent space. The difference may be that learners cannot physically enter the referent space as they can when navigating. We hypothesize that such activities are popular with teachers because they are seen as an opportunity for active learning (Bonwell and Eison, 1991) where actions embody events or objects, but their usefulness for teaching scale is unknown. Research is needed to identify the relative value of action versus the potential cognitive hurdle of the mismatch between an event model, in which children are moving, and an object referent.

Informal science education also employs walking-scale models for astronomical distances and for geological time. Several U.S. cities, including Washington, DC, have scale models of the solar system that visitors to the city can walk to better comprehend relative spatial distances. The Trail of Time at the Grand Canyon (Karlstrom et al., 2008) and the area by Dinosaur Ridge in Red Rocks Park, near Denver, CO, are examples of informal science education models for geological time. As visitors vertically descend into the Grand Canyon, they travel backward in time at the rate of 1 million years per meter walked. The models offer the opportunity to connect standing in one spatial location and movement through space to a referent so that they may be conceived of as both object and event models. It is not known whether geological time models that connect time to actual movement in space are better at helping students learn about long time periods than the object model described earlier. However, recent research uncovered statistically significant relationships among motor control, spatial reasoning, and proportional reasoning in young children (Frick and Möhring, 2016), suggesting a potential relationship between movement in space and learning about scalar relationships. Future studies could disentangle whether the statistical relationships between these cognitive factors persist when tested with content-specific embodied activities, like walking a scale model.

Using Other Runnable Physical Models to Teach Scale: The Example of a Stream Table

Some runnable event models do not explicitly include spatial or temporal scale as an instructional goal; they are designed to teach about spatial patterns that are the product of events (Kastens and Rivet, 2008). Because those patterns reveal spatiotemporal geological processes acting over a range of spatial and temporal scales, they can be leveraged to teach learners about scale as both magnitude and relationship. Here, we illustrate the educational opportunities and challenges of doing so by using a stream table, an example of the broad class of runnable physical models used to demonstrate Earth processes.

Middle school, high school, and university instructors often use a stream table as an event–event model of fluvial erosion and deposition (Friesen, 2005). The model may be used to achieve multiple learning outcomes across K–16 contexts, including showing students the following: a complex process with clear regularities, such as flow velocity differences at different locations within the channel; the characteristic traces left by the event, such as greater sediment deposition in regions of slow water flow; and the repeatable patterns that occur as a result of manipulating variables, like table slope.

To highlight the importance of sediment size in depositional patterns, the table might be filled with a mix of sand and gravel. The teacher may guide students to notice similarities in sediment distribution from one run to the next, how those patterns change if the energy of the water increases or decreases, and how sediment distribution changes for simple laminar flow in a straight stream compared to turbulent flow in a curved stream. A teacher could draw attention to the rate of water flow in different parts of the stream around a curve to try to relate the sediment pattern to the event of flowing water carrying sediments. Students may be instructed to note similarities between patterns seen in the model and patterns seen on Earth. For example, sediment bars on the table resemble nearby islands, or when a lot of sediment is added to the flowing water, the stream shape on the table is analogous to a sinuous river.

Elementary teachers also use stream tables, usually straight stream models with a single type of Earth material for each run (Cheek, 2013c). Students investigate how vegetation can mitigate erosion, thereby linking the model to environmental science. Children might generalize about the spatial extent of the erosion or the volume of the deposited material when one of the variables described above is manipulated. The goal is for students to infer the relationship between a manipulated variable like water volume and the erosion rate. The analogy between processes in the model and actual Earth processes is strong enough for such models to serve as disciplinary research tools, such as St. Anthony Falls Laboratory's Jurassic Tank (Paola et al., 2001).

A stream table is a way to demonstrate how fluvial processes sculpt Earth's surface across temporal and spatial scales, but learners may not grasp scalar properties of the model in the absence of specific instruction (Miller and Brewer, 2010). Students face scaling challenges when aligning familiar scales in the model to geological ones. One is the relationships among rate, duration, and spatial extent. When stream discharge and velocity are high, significant erosion can occur on relatively short timescales because stream competence and capacity are increased. Conversely, slow temporal rates can produce large spatial scale changes. For practical purposes, event model rates may not match Earth rates. This could lead students to think that erosion happens more quickly than is the case, analogous to the catastrophist views about geological time Hidalgo et al. (2004) found among high schoolers and lower division undergraduates. We do not know whether textbook images and media portrayals of storm-related erosion or the speed with which the model is usually run contributes to this view. One way to deal with the misconception is to design the apparatus so that velocity and discharge are both low but the stream flows over an extended period (several days).

Older students face additional scalar challenges with the model. Proportional relationships are difficult to measure in the stream table but are important for undergraduates in geomorphology courses. A stream table introduces multidimensional scale relationships such as the channel width-to-depth ratio, if only in approximate terms. Scalar relationships between objects in the model may be challenging (Cox, 2013), and learners must be able to determine whether particular spatial relationships that exist in the referent are preserved or transformed in the stream table. Sediments in the model have less scalar variability than those in the referent, and stream tables use only unconsolidated sediments for pragmatic reasons. Both affect erosional rates and volume of alluvium transported. Models in which one dimension is exaggerated relative to others may make it more difficult for learners to grasp scalar relationships (Schneps et al., 2014). If students are struggling to map the event model to the event referent, intermediate versions of the model that are more similar to the referent may be helpful (Gentner and Markman, 1994, 1997; Resnick et al., 2017).

We do not know whether learners think about spatial and temporal scale aspects of fluvial processes when their instructor uses a stream table in class. The science education and DBER literature reviewed for this paper provides no evidence of how people learn about scale in a contextualized way; however, two chemistry education studies indicate that teachers and students both find it difficult to determine which variables are important at particular scales (Kruse and

Roehrig, 2005; Frändberg et al., 2013). These findings suggest that explicitly directing students' attention to the features of the model that align with features of the real-world referent would be helpful.

RESEARCH IMPLICATIONS

As Table I indicates, there is relatively little science education or DBER on students' conceptions about spatial and temporal scale and even fewer studies on how to improve students' conceptions of scale. This is unfortunate given their importance in all science disciplines. The research that does exist is not as well connected to the ways practicing geoscientists use scale as is probably needed (e.g., reasoning across regional and thin-section scales to develop a compelling kinematic account of a region's deformational history). We have identified a number of questions for future research throughout this paper. We summarize them here:

- In what ways do (or should) concepts of scale as magnitude or as a relationship develop across K–16 schooling? Is there evidence of a learning progression for scale as the Framework envisions there should be (National Research Council, 2012)? When do learners integrate ideas about scale with subject matter knowledge? Do concepts of scale develop fairly late in geoscience, as they do in chemistry (Johnson and Tymms, 2011)? If that is the case, is that normative, or is it because we have not been purposeful about teaching scale? Why have different studies found contradictory error patterns related to students' subject matter knowledge?
- How closely is a feel for quantity (National Research Council, 2012) tied to students' conceptions of scale as magnitude? Are disciplinary spatial and temporal magnitude categories generalized to other domains? If learners have a feel for the magnitude of $1 billion, does that facilitate a feel for the magnitude of 1 billion years, for example? Does competence with the metric system help students better understand scale as magnitude and as a relationship? If so, what does this mean for American students who have little everyday experience with metric measurement units? Would interventions designed to improve mathematical concepts, such as part–whole reasoning, also improve scalar understanding across contexts?
- What is the relationship between reasoning about scale and other spatial reasoning skills, such as mental rotation?
- Is the object–event dichotomy useful for instructors who want to improve how they teach about scale? Would emphasizing scale when using models like a stream table produce gains in students' ability to reason about scale? We have introduced a 2×2 matrix of common geoscience models that we argue can be leveraged to better integrate concepts of scale within geoscience instruction (Table II). At present, we have no data to indicate whether our assertion is accurate due to the paucity of science education and DBER on this point.
- What pedagogical interventions facilitate the development of concepts of scale across K–16? We found few in our review, and only one collected the data needed to provide evidence of durable results (Resnick et al., 2017). Would a similar intervention that targets small-scale magnitudes be as effective?
- To what extent do event models that employ active, embodied learning help students improve their ability to reason about scale? Such activities are popular with instructors throughout K–16 contexts, but their value for teaching scale has not been established.
- Why do common teaching activities related to measurement seem to help some learners improve their ability to reason about scale (Jones and Taylor, 2009) but not others? Is it related to the cognitive approach different learners employ to solve scaling problems?

Many of the science education and DBER studies reviewed for this paper are only tangentially connected to the psychological literature, which we see as a drawback. Those that address scale as a relationship do so mainly through ordering tasks. Students can often complete those tasks successfully on the basis of logical reasoning alone, so it is difficult to say what they understand about scale as a relationship from that research. Future research needs to be well connected to theoretical frameworks in cognitive science on learning about scale as magnitude and as a relationship. Interdisciplinary research teams with geoscience education researchers, cognitive scientists, and perhaps mathematics education researchers are well suited for these types of studies.

Study authors need to explicitly state when they are referring to scale as magnitude or as a relationship. This would make it easier for readers to compare findings across studies. Future studies need to address the imbalance between the amount of research devoted to scale as magnitude and that devoted to scale as a relationship. We posed several potential research questions related to intervention studies above. Overall, we can say that more intervention studies are needed. Some of the science education and DBER studies reviewed for this paper were cross-age. No studies were longitudinal. Any attempt to develop a learning progression for how people learn to reason about scale will benefit from both types of studies.

TEACHING IMPLICATIONS

We laid out a number of teaching implications in preceding sections. In summary, instructors should be intentional about integrating instruction on spatial and temporal scale throughout the curriculum, an approach promoted by the NGSS (NGSS Lead States, 2013). We think this practice is as important for undergraduate instructors as it is for K–12 teachers. We believe the framework we have proposed can be useful for teachers as they plan instruction related to scale but acknowledge that its usefulness is yet to be tested. When teaching about scale as magnitude, we advocate the use of bridging analogies that will enable learners to progressively align scales from the very familiar to less familiar, rather than making a single cognitive leap to a magnitude scale that is highly dissimilar to students' experiences. Drawing students' attention to part–whole relationships in a model can help students learning about scale as a relationship. Constructing runnable models like stream tables so that they run with slow processes (e.g., low velocity) but over several days can help students discern how what appear to be small-scale changes can have large impacts. Directing students' attention to features of the model that align with the referent can help them improve their capacity to reason with the model. Similarly, encouraging students to evaluate the benefits and limitations of models, one of the eight science and engineering practices in the NGSS (NGSS Lead States, 2013) for spatial and temporal scale, can also be beneficial. Noting a model's benefits and limitations requires a learner to consider which features of the model align with the referent, which ones do not, and the implications for the model's usefulness.

CONCLUSIONS

We often refer to scale as if it is a singular concept whose meaning is transparent, but that is not so. Scale can be thought of in at least two ways in geoscience: as magnitude and as a relationship. Both are important for the integrated way scale is used by geoscientists when engaged in their work, including field mapping. Integrating magnitude and relational information across multiple scales may also be necessary in other fields, such as particle physics and nanotechnology.

Therefore, scale (with its various meanings) is incredibly important in geoscience education. We cannot expect that all learners will construct an understanding of scale on their own. We see a focus on scale as timely given the development of the NGSS and its emphasis on three-dimensional learning that incorporates concepts like scale into content instruction. While we have advocated for teaching scale in a way that is integrated across a range of geoscience topics, we only have limited understanding of how to help learners construct understanding of the magnitude of the spatial and temporal scales of objects and processes in the Earth system or the relationships between them. Our community needs research-based instructional practices that geoscience instructors can use to meaningfully integrate spatial and temporal scale into their teaching.

Acknowledgment

Preparation of this manuscript was supported in part by National Science Foundation grant 640800 to TFS and NDL. We thank Mia Velazquez for her thoughtful suggestions regarding the structure of Table II.

REFERENCES

Atit, K., Garnier, K., and Shipley, T.F. 2015. Student gestures aid penetrative thinking. *Journal of Geoscience Education,* 63:66–72.

Atit, K., Shipley, T.F., and Tikoff, B. 2014. A framework for classifying spatial gestures. *In* Montello, D.R., Grossner, K., and Janelle, D.J., eds., Space in mind: Concepts for spatial learning and education. Cambridge, MA: MIT Press, p. 173–194.

Atkins, K. 2013. Investigating the geologic time scale: Creating posters to display trends in geologic time. Available at http://serc.carleton.edu/sp/mnstep/activities/19783.html (accessed 10 March 2017).

Atmospheric Science Literacy Framework. 2007. Atmospheric science literacy: The essential principles and fundamental concepts of atmospheric science. Available at http://eo.ucar.edu/asl/ (accessed 10 February 2016).

Barth, H.C., and Paladino, A.M. 2011. The development of numerical estimation: Evidence against a representational shift. *Developmental Science,* 14:125–135.

Bonwell, C.C., and Eison, J.A. 1991. Active learning: Creating excitement in the classroom. ASHE-ERIC Higher Education Report. Washington, DC: George Washington University School of Education and Human Development.

Boyer, T.W., Levine, S.C., and Huttenlocher, J. 2008. Development of proportional reasoning: Where young children go wrong. *Developmental Psychology,* 44:1478–1490.

Catley, K., and Novick, L. 2009. Digging deep: Exploring college students' knowledge of macroevolutionary time. *Journal of Research in Science Teaching,* 46:311–332.

Chatterjee, A. 2008. The neural organization of spatial thought and language. *Seminars in Speech and Language,* 29:226–238.

Cheek, K.A. 2012. Students' understanding of large numbers as a key factor in their understanding of geologic time. *International Journal of Science and Mathematics Education,* 10:1047–1069.

Cheek, K.A. 2013a. Exploring the relationship between students' understanding of conventional time and geologic time. *International Journal of Science Education,* 35:1925–1945.

Cheek, K.A. 2013b. How geoscience novices reason about temporal duration: The role of spatial thinking and large numbers. *Journal of Geoscience Education,* 61:334–348.

Cheek, K.A. 2013c. Washed away! *Science and Children,* 50:52–56.

Cheek, K.A. 2016. Early view. Do Indonesian children's experiences with large currency units facilitate magnitude estimation of long temporal periods? *Research in Science Education,* 1–23.

Clary, R.M., Brzuszek, R.F., and Wandersee, J.H. 2009. Students' geocognition of deep time, conceptualized in an informal educational setting. *Journal of Geoscience Education,* 57:275–285.

Collins, A., and Gentner, D. 1987. How people construct mental models. *In* Holland, D., and Quinn, N., eds., Cultural models in language and thought. Cambridge, UK: Cambridge University Press, p. 243–265.

Cox, D.C. 2013. Similarity in middle school mathematics: At the crossroad of geometry and number. *Mathematical Thinking and Learning,* 15:3–23.

Dehaene, S. 1997. The number sense: How the mind creates mathematics. Oxford, UK: Oxford University Press.

Dehaene, S. 2003. The neural basis of the Weber-Fechner law: A logarithmic mental number line. *Trends in Cognitive Sciences,* 7:145–147.

Dehaene, S., Izard, V., Spelke, E., and Pica, P. 2008. Log or linear? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. *Science,* 320:1217–1220.

Delgado, C. 2013a. Cross-cultural study of understanding of scale and measurement: Does the everyday use of U.S. customary units disadvantage U.S. students? *International Journal of Science Education,* 35:1277–1298.

Delgado, C. 2013b. Navigating deep time: Landmarks for time from the big bang to the present. *Journal of Geoscience Education,* 61:103–112.

Dickerson, D., Callahan, T., Van Sickle, M., and Hay, G. 2005. Students' conceptions of scale regarding groundwater. *Journal of Geoscience Education,* 53:374–380.

Dodick, J., and Orion, N. 2003. Cognitive factors affecting student understanding of geologic time. *Journal of Research in Science Teaching,* 40:415–442.

Earth Science Literacy Initiative. 2010. Earth Science literacy principles: The big ideas and supporting concepts of Earth Science. Available at http://www.earthscienceliteracy.org/index.html (accessed 10 February 2016).

Feigenson, L., Dehaene, S., and Spelke, E. 2004. Core systems of number. *Trends in Cognitive Sciences,* 8:307–314.

Flynn, A.E. 2007. Investigating ordering planets: Math connections and number sense in science. Available at http://serc.carleton.edu/sp/mnstep/activities/planets.html (accessed 10 March 2017).

Frändberg, B., Lincoln, P., and Wallin, A. 2013. Linguistic resources used in grade 8 students' submicro level explanations—Science items from TIMSS 2007. *Research in Science Education,* 43:2387–2406.

Frick, A., and Möhring, W. 2016. A matter of balance: Motor control is related to children's spatial and proportional reasoning skills. *Frontiers in Psychology,* 6:1–10.

Friesen, B. 2005. Build your own stream table. Available at http://nagt.org/nagt/teaching_resources/teachingmaterials/9271.html (accessed 10 February 2016).

Gentner, D. 1983. Structure mapping: A theoretical framework for analogy. *Cognitive Science,* 7:155–170.

Gentner, D. 1989. The mechanisms of analogical learning. *In* Vosniadou, S., and Ortony, A., eds., Similarity and analogical reasoning. Cambridge, UK: Cambridge University Press, p. 199–241.

Gentner, D. 2006. Why verbs are hard to learn. *In* Hirsh-Pasek, K., and Golinkoff, R., eds., Action meets word: How children learn verbs. Oxford, UK: Oxford University Press, p. 544–564.

Gentner, D., Anggoro, F.K., and Kilbanoff, R.S. 2011. Structure mapping and relational language support children's learning of relational categories. *Child Development,* 82:1173–1188.

Gentner, D., Levine, S.C., Ping, R., Isaia, A., Dhillon, S., Bradley, C., and Honke, G. 2016. Rapid learning in a children's museum via analogical comparison. *Cognitive Science,* 40:224–240.

Gentner, D., Loewenstein, J., and Hung, B. 2007. Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development,* 8:285–307.

Gentner, D., and Markman, A.B. 1994. Structural alignment in comparison: No difference without similarity. *Psychological Science,* 5:152–158.

Gentner, D., and Markman, A.B. 1997. Structure mapping in analogy and similarity. *American Psychologist,* 52:45–56.

Gentner, D., and Smith, L. 2012. Analogical reasoning. *In* Ramachandran. V.S., ed., Encyclopedia of human behavior, 2nd ed. Oxford, UK: Elsevier, p. 130–136.

Geoscience Employers Workshop Outcomes. 2015. EAR-1347209: The future of undergraduate geoscience education. Available at http://www.jsg.utexas.edu/events/files/Employers_Workshop_outcomes.pdf (accessed 16 January 2017).

Gerlach, K., Trate, J., Blecking, A., Geissinger, P., and Murphy, K. 2014a. Investigation of absolute and relative scaling conceptions of students in introductory college chemistry courses. *Journal of Chemical Education,* 91:1526–1537.

Gerlach, K., Trate, J., Blecking, A., Geissinger, P., and Murphy, K. 2014b. Valid and reliable assessments to measure scale literacy of students in introductory college chemistry courses. *Journal of Chemical Education,* 91:1538–1545.

Glaser, K., and Riegler, P. 2015. Beginning students may be less capable of proportional reasoning than they appear to be. *Teaching Mathematics and Its Applications,* 34:26–34.

Harel, G. and Confrey, J. 1994. The development of multiplicative reasoning in the learning of mathematics. Albany, NY: State University of New York Press.

Hidalgo, A., San Fernando, I.E.S., and Otero, I.C.E.J. 2004. An analysis of the understanding of geological time by students at secondary and post-secondary level. *International Journal of Science Education,* 26:845–857.

Holland, J.H., Holyoak, K.J., Nisbett, R.E., and Thagard, P.R. 1986. Induction: Processes of inference, learning, and discovery. Cambridge, MA: Bradford Books/MIT Press.

Huttenlocher, J., Hedges, L.V., and Duncan, S. 1991. Categories and particulars: Prototype effects in estimating spatial location. *Psychological Review,* 98:352–376.

Huttenlocher, J., Hedges, L.V., and Prohaska, V. 1988. Hierarchical organization in ordered domains: Estimating the date of events. *Psychological Review,* 95:471–484.

Johnson, C.L., Semple, I.L., and Creem-Regehr, S.H. 2013. The effects of scaling cues and interactivity on a viewer's ability to estimate the size of features shown on outcrop imagery. *Journal of Geoscience Education,* 61:68–80.

Johnson, P., and Tymms, P. 2011. The emergence of a learning progression in middle school chemistry. *Journal of Research in Science Teaching,* 48:849–877.

Jolley, A., Jones, F., and Harris, S. 2013. Measuring student knowledge of landscapes and their formation timespans. *Journal of Geoscience Education,* 61:240–251.

Jones, M.G., Gardner, G., Taylor, A.R., Wiebe, E., and Forrester, J. 2011. Conceptualizing magnification and scale: The roles of spatial visualization and logical thinking. *Research in Science Education,* 41:357–368.

Jones, M.G., Paechter, M., Yen, C.-F., Gardner, G., Taylor, A., and Tretter, T. 2013. Teachers' concepts of spatial scale: An international comparison. *International Journal of Science Education,* 35:2462–2482.

Jones, M.G., and Taylor, A. 2009. Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching,* 46:460–475.

Jones, M.G., Taylor, A., and Broadwell, B. 2009a. Concepts of scale held by students with visual impairment. *Journal of Research in Science Teaching,* 46:506–519.

Jones, M.G., Taylor, A., and Broadwell, B. 2009b. Estimating linear size and scale: Body rulers. *International Journal of Science Education*, 31:1495–1509.

Jones, M.G., Taylor, A., Minogue, J., Broadwell, B., Wiebe, E., and Carter, G. 2007. Understanding scale: Powers of ten. *Journal of Science Education and Technology,* 16:191–202.

Jones, M.G., Thomas, A., Superfine, R., and Taylor, R. 2003. Learning at the nanoscale: The impact of students' use of remote microscopy on concepts of viruses, scale, and microscopy. *Journal of Research in Science Teaching,* 40:303–322.

Jones, M.G., Tretter, T., Taylor, A., and Oppewal, T. 2008. Experienced and novice teachers' concepts of spatial scale. *International Journal of Science Education,* 30:409–429.

Karlstrom, K., Semken, S., Crossey, L., Perry, D., Gyllenhaal, E.D., Dodick, J., Williams, M., Hellmich-Bryan, J., Crow, R. Watts, N.B., and Ault, C. 2008. Informal geoscience education on a grand scale: The trail of time exhibition at Grand Canyon. *Journal of Geoscience Education,* 56:354–361.

Kastens, K.A., and Rivet, A. 2008. Multiple modes of inquiry in Earth Science. *Science Teacher,* 75:26–31.

Kruse, R.A., and Roehrig, G.H. 2005. A comparison study: Assessing teachers' conceptions with the Chemistry Concepts Inventory. *Journal of Chemical Education,* 82:1246–1250.

Landy, D., Charlesworth, A., and Ottmar, E. 2016. Early view. Categories of large numbers in line estimation. *Cognitive Science,* 1–28.

Landy, D., Silbert, N., and Goldin, A. 2013. Estimating large numbers. *Cognitive Science,* 37:775–799.

Lee, H.-S., Liu, O.L., Price, C.A., and Kendall, A.L.M. 2011. College students' temporal-magnitude recognition ability associated with durations of scientific changes. *Journal of Research in Science Teaching,* 48:317–335.

Libarkin, J., Kurdziel, J., and Anderson, S. 2007. College student conceptions of geological time and the disconnect between ordering and scale. *Journal of Geoscience Education,* 55:413–422.

Lin, L.-F., Hsu, Y.-S., and Yeh, Y.-F. 2012. The role of computer simulation in an inquiry-based learning environment: Reconstructing geological events as geologists. *Journal of Science Education and Technology,* 21:370–383.

Maguire, M., Shipley, T.F., Brumberg, J., and Ennis, M. 2011. Similarities in object and event segmentation: A geometric approach to spatiotemporal segmentation. *Spatial Cognition and Computation,* 11:254–279.

Markman, A.B., and Gentner, D. 1993. Structural alignment during similarity comparisons. *Cognitive Psychology,* 25:431–467.

Markowski, P., and Richardson, Y. 2010. Mesoscale meteorology in midlatitudes. Chichester, UK: John Wiley & Sons.

Matthews, P.G., Lewis, M.R., and Hubbard, E.M. 2016. Individual differences in nonsymbolic ratio processing predict symbolic math performance. *Psychological Science,* 27:191–202.

Miller, B.W., and Brewer, W.F. 2010. Misconceptions of astronomical distances. *International Journal of Science Education,* 32:1549–1560.

Möhring, W., Newcombe, N.S., and Frick, A. 2014. Zooming in on spatial scaling: Preschool children and adults use mental transformations to scale spaces. *Developmental Psychology,* 50:1614–1619.

Möhring, W., Newcombe, N.S., and Frick, A. 2015. The relation between spatial thinking and proportional reasoning in preschoolers. *Journal of Experimental Child Psychology,* 132:213–220.

Möhring, W., Newcombe, N.S., Levine, S.C., and Frick, A. 2016. Spatial proportional reasoning is associated with formal knowledge about fractions. *Journal of Cognition and Development,* 17:67–84.

Muhammed, A.R. 2016. What sizes are the planets and how do they move around the sun? Available at http://serc.carleton.edu/sp/mnstep/activities/35914.html (accessed 10 March 2017).

Nam, Y., Karahan, E., and Roehrig, G. 2016. Native American students' understanding of geologic time scale: 4th–8th grade Ojibwe students' understanding of Earth's geologic history. *International Journal of Environmental and Science Education,* 11:485–503.

National Research Council. 2012. A framework for K–12 science education: Practices, crosscutting concepts, and core ideas. Washington, DC: National Academies Press.

Newcombe, N.S., and Shipley, T.F. 2014. Thinking about spatial thinking: New typology, new assessments. *In* Geto, J.S., ed., Studying visual and spatial reasoning for design creativity. New York: Springer, p. 179–192.

Next Generation Science Standards (NGSS) Lead States. 2013. Next Generation Science Standards: For states by states. Washington, DC: National Academies Press.

Ni, Y., and Zhou, Y.-D. 2005. Teaching and learning fraction and rational numbers: The origin and implications of whole number bias. *Educational Psychologist,* 40:27–52.

Nicoll, G. 2003. A qualitative investigation of undergraduate chemistry students' macroscopic interpretations of the submicroscopic structure of molecules. *Journal of Chemical Education,* 80:205–213.

National Oceanic and Atmospheric Administration (NOAA), Office of Education and National Marine Sanctuary. 2013. Ocean literacy: The essential principles and fundamental concepts of ocean science for learners of all ages. Available at http://oceanliteracy.wp2.coexploration.org/ocean-literacy-framework/ (access 10 February 2016).

Ormand, C.J., Manduca, C., Shipley, T.F., Tikoff, B., and Harwood, C.L. 2014. Evaluating geoscience students' spatial thinking skills in a multi-institutional classroom study. *Journal of Geoscience Education,* 62:146–154.

Paola, C., Mullin, J., Ellis, C., Mohring, D.C., Swenson, J.B., Parker, G., Hickson, T., Heller, P.L., Pratson, L., Syvitski, J., Sheets, B., and Strong, N. 2001. Experimental stratigraphy. *GSA Today,* 11:4–9.

Plummer, J.D. 2014. Spatial thinking as the dimension of progress in an astronomy learning progression. *Studies in Science Education,* 50:1–45.

Pozzer, L.L., and Roth, W.-M. 2003. Prevalence, function, and structure of photographs in high school biology textbooks. *Journal of Research in Science Teaching,* 40:1089–1114.

Resnick, I., Davatzes, A., Newcombe, N.S., and Shipley, T.F. 2017. Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *Educational Psychology Review* [Special Issue].

Rundgren, C.-J., Rundgren, S.-N.C., and Schönborn, K.J. 2010. Students' conceptions of water transport. *Journal of Biological Education,* 44:129–135.

Schneps, M.H., Ruel, J., Sonnert, G., Dussault, M., Griffin, M., and Sadler, P.M. 2014. Conceptualizing astronomical scale: Virtual simulations on handheld tablet computers reverse misconceptions. *Computers and Education,* 70:269–280.

Shipley, T.F., Tikoff, B., Ormand, C., and Manduca, C. 2013. Structural geology practice and learning, from the perspective of cognitive science. *Journal of Structural Geology,* 54:72–84.

Siegler, R., and Opfer, J. 2003. The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science,* 14:237–243.

Siegler, R., Thompson, C.A., and Opfer, J. 2009. The logarithmic-to-linear shift: One learning sequence, many tasks, many time scales. *Mind, Brain, and Education,* 3:143–150.

Summit on Future of Undergraduate Geoscience Education. 2014. Future of Undergraduate Geoscience Education: Summary report for Summit on Future of Undergraduate Geoscience Education. Available at http://www.jsg.utexas.edu/events/files/Future_Undergrad_Geoscience_Summit_report.pdf (accessed 10 February 2016).

Swarat, S., Light, G., Park, E.J., and Drane, D. 2011. A typology of undergraduate students' conceptions of size and scale: Identifying and characterizing conceptual variation. *Journal of Research in Science Teaching,* 48:512–533.

Teed, R., and Slattery, W. 2011. Changes in geologic time understanding in a class for preservice teachers. *Journal of Geoscience Education,* 59:151–162.

Thompson, C.A., and Opfer, J. 2010. How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development,* 81:1768–1786.

Tourniaire, F., and Pulos, S. 1985. Proportional reasoning: A review of the literature. *Educational Studies in Mathematics,* 16:181–204.

Trend, R. 1998. An investigation into understanding of geological time among 10- and 11-year-old children. *International Journal of Science Education,* 20:973–988.

Trend, R. 2000. Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science. *International Journal of Science Education,* 22:539–555.

Trend, R. 2001. Deep time framework: A preliminary study of U.K. primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching,* 38:191–221.

Tretter, T., Jones, M.G., Andre, T., Negishi, A., and Minogue, J. 2006a. Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching,* 43:282–319.

Tretter, T., Jones, M.G., and Minogue, J. 2006b. Accuracy of scale conceptions in science: Mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching,* 43:1061–1085.

U.S. Global Change Research Project. 2009. Climate literacy: The essential principles of climate science. Available at http://cpo.noaa.gov/OutreachandEducation/ClimateLiteracy.aspx (accessed 10 February 2016).

Uttal, D.H., Miller, D.I., and Newcombe, N.S. 2013. Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science,* 22:367–373.

Wysession, M., LaDue, N., Budd, D.A., Campbell, K., Conklin, M., Kappel, E., Lewis, G., Raynolds, R., Ridky, R.W., Ross, R.M., Taber, J., Tewksbury, B., and Tuddenham, P. 2012. Developing and applying a set of Earth Science literacy principles. *Journal of Geoscience Education*, 60:95–99.

Yezierski, E.J., and Birk, J.P. 2006. Using animations to close the gender gap. *Journal of Chemical Education,* 83:954–960.

Zachs, J.M., and Tyversky, B. 2001. Event structure in perception and conception. *Psychological Bulletin,* 127:3–21.