Educ Psychol Rev
DOI 10.1007/s10648-016-9371-5

REVIEW ARTICLE

Using Relational Reasoning to Learn About Scientific Phenomena at Unfamiliar Scales

Ilyse Resnick [1] **· Alexandra Davatzes** [2] **· Nora S. Newcombe** [3] **· Thomas F. Shipley** [3]

Abstract Many scientific theories and discoveries involve reasoning about extreme scales, removed from human experience, such as time in geology and size in nanoscience. Thus, understanding scale is central to science, technology, engineering, and mathematics. Unfortunately, novices have trouble understanding and comparing sizes of unfamiliar large and small magnitudes. Relational reasoning is a promising tool to bridge the gap between direct experience and phenomena at extreme scales. However, instruction does not always improve understanding, and analogies can fail to bring about conceptual change, and even mislead students. Here, we review how people reason about phenomena across scales, in three sections: (a) we develop a framework for how relational reasoning supports understanding extreme scales; (b) we identify cognitive barriers to aligning human and extreme scales; and (c) we outline a theory-based approach to teaching scale information using relational reasoning, present two successful learning activities, and consider the role of a unified scale instruction across STEM education.

Keywords Size and scale · Relational reasoning · Analogy · Progressive alignment · Corrective feedback

Being able to reason about size and scale is central to performance in science, technology, engineering, and mathematics (STEM; Hawkins 1978; Tretter et al. 2006) and, as such, are suggested as a unifying theme in science education (National Research Council 2011; American Association for the Advancement of Science 1993). Size refers to an absolute magnitude or value. Scale refers to systems of measurement, which allow for the comparison of relative sizes. For example, years and kilometers are both conventional scales to measure

✉ Ilyse Resnick
iresnick@udel.edu

[1] School of Education, University of Delaware, 211C Willard Hall, Newark DE 19716, USA

[2] Department of Earth and Environmental Sciences, Temple University, Philadelphia, PA 19122, USA

[3] Department of Psychology, Temple University, Philadelphia, PA 19122, USA

Published online: 19 May 2016

temporal duration and spatial distance, respectively. Knowing that the Earth formed 4.6 billion years ago provides a scale, or measure, to understand that while there is approximately 225 million years between when dinosaurs and humans appear, these two events occurred close together relative to the history of the planet.

Many fundamental scientific concepts and discoveries, such as the example above, are at extreme scales, far removed from human experience. Additional examples include the astronomical distances involved in space exploration and the rapidly developing field of nanotechnology—both are based on phenomena occurring at scales that cannot be directly perceived. Unfortunately, scales outside of human perception can be particularly difficult to comprehend. While novices can typically place events and phenomena in the correct sequential order, they fail to understand the magnitude in between (Jones et al. 2008; Miller and Brewer 2010; Trend 2001; Tretter et al. 2006). In particular, novices have trouble assigning, comprehending, and comparing absolute sizes, especially at extreme scales (e.g., Delgado et al. 2007; Jones et al. 2008; Libarkin et al. 2005).

Because scales outside of human perception cannot be directly experienced, comprehension likely requires relational reasoning. Relational reasoning is a basic cognitive mechanism involved in the formation of conceptual categories and encompasses the ability to detect similarities and differences in patterns among objects, concepts, and situations (James 1890; Alexander and the Disciplined Reading and Learning Research Laboratory (DRLRL) 2012). Below, we present a theoretical framework for how relational reasoning may support understanding scales outside of human perception. We then identify cognitive barriers to the application of relational reasoning in understanding scale. Finally, we outline a theory-based approach to teaching scale information using relational reasoning. We present two successful learning activities and consider the role of unifying scale instruction across STEM education.

Framework for How Relational Reasoning Supports Scale Understanding

Similarities Between Human and Extreme Scales

There is converging cognitive, neurocognitive, developmental, and comparative evidence that suggest reasoning about any type of scale (e.g., temporal, spatial, abstract) uses the same neural and conceptual resources (e.g., Bueti and Walsh 2009; Cantlon et al. 2009; Lourenco and Longo 2011; Walsh 2003; for a review, see Cohen Kadosh et al. 2008). Magnitude information is cognitively stored as a hierarchical combination of both metric and categorical information (Huttenlocher et al. 1988; Huttenlocher et al. 2000). For example, recollection that a wedding took place on June 12, implicitly includes the higher-level categorical information that the wedding took place in the summer. Such use of categorical information is seen even at extremely large and small scales (Landy et al. 2014; Resnick et al. 2016a, b). Variation in estimation thus occurs because of imprecision of category boundaries, with increased variation associated with larger category ranges (Shipley and Zacks 2008; Zacks and Tversky 2001).

Because people reason about scales within and outside of human perception in similar ways, analogical reasoning may be especially relevant in understanding extreme scales (Resnick et al. 2016a, b). Analogical reasoning is a specific kind of relational reasoning, which refers to aligning structural similarities between a base concept and target concept (e.g., Gentner et al. 2007; Markman and Gentner 1993a). In an analogy, a new unfamiliar concept is mapped onto a relevant analog accessed from memory to identify systematic correspondences (Gentner and Holyoak 1997). This mapping process is characterized by a consistent, one-to-one connections between the structural relations of the base and target concepts, and not their features (Gentner 1982, 1983; Gentner & Gentner 1983). The base concept functions to organize and visualize the target concept, which, in turn, increases understanding and recall (Orgill and Bodner 2004; Simons 1984). Thus, the base concept serves as a structure to help explain or clarify the target concept (Gentner 1983; Markman and Gentner 1993a).

One way people may come to understand scales far removed from their personal experience is by aligning the structural similarities of extreme scales with familiar human scales. Landy et al. (2014) suggest that extreme scales are reasoned about by recycling, and not extending, cognitive resources involved in small number processing. In their study, when estimating values on a billion scale, undergraduates first divided the billion scale into smaller subscales, made linear adjustments within each subscale, and then combined the subscales to compromise the original billion scale. Being able to build and use extensive complex units, or categories, is associated with a range of mathematical abilities (Carpenter and Moser 1983; Lamon 1994), because it allows for the comparison of individual items as well as their aggregate (Callanan and Markman 1982). Indeed, unitizing is characteristic of STEM experts' reasoning about scale (Tretter et al. 2006).

Dissimilarities Between Human and Extreme Scales

Despite similarities in reasoning at scales within and outside of human perception, there are some notable differences. The same observation at one scale can have a completely different meaning at another scale. For example, while fluctuations in weather indicate something about daily atmospheric conditions, the same weather fluctuations over much longer periods of time can indicate changes in climate. Further, even very small changes in Earth's average temperature over long periods of time can have very large effects, even though temperature can change dramatically within a 24-h period with no noticeable effect the following day (Riebeek 2010).

Phenomena at different scales can also behave very differently, and, as such, properties of phenomena at a given scale may not be true of other scales. A classic example of scaling effects can be found in the observation of structural strength; after falling eight feet a horse's bones will break whereas a dog falling, the same amount would be able to walk away (Galileo 1638). This particular scaling effect is largely due to the ratio of surface area to volume varying as the scale of the object changes and has serious implications in construction.

Human scales and extreme scales may also be formally conceptualized using mutually exclusive, or antinomic, organizational structures. For example, while time at human scales is based on temporal durations (60 s = 1 min, 60 min = 1 h), time at geologic scales is based on the occurrence of important events (the Mesozoic = age of reptiles; the Cenozoic = age of mammals). Thus, human temporal scales are equally spaced whereas geologic temporal scales are not.

Alexander and DRLRL (2012) have identified three ways in which people reason relationally about differences. Anomalous reasoning is the ability to identify deviations from an established pattern (Dumas et al. 2013). Antithetical and antinomic reasoning both involve being able to reconcile conflicting information (Alexander et al. 2016), with antithesis pertaining to identifying oppositional relations and antinomy pertaining to incompatible relations (Dumas et al. 2013). Reasoning about dissimilarities requires being able to first detect a pattern and then identify subsequent deviations from that pattern (Dumas et al. 2013) and has been implicated as a mechanism for conceptual change (Gentner 1983; Hummel and Holyoak 2003; Vosniadou and Mason 2012). Conceptual change can occur because the process of mapping a base concept and target concept not only creates new knowledge, but can also cause the evaluation and reconceptualization of existing conceptual categories and schemas (Hatford 1993).

People may come to understand scales far removed from their personal experience by identifying and reasoning about observed differences between different scales. Fractions, for example, typically represents the first real opportunity for students to reason about numbers other than whole numbers (Siegler and Lortie-Forgues 2014; Siegler et al. 2011). Importantly, the properties of whole numbers are not true of all numbers; while the multiplication of whole numbers always results in a larger number, the same is not true of fractions. Learning about fractions often results in a deeper understanding of numerical properties because students must reconcile properties of fractions with their previous understanding of number (Siegler and Lortie-Forgues 2014; Siegler et al. 2011). While fractions can be located interstitially between whole numbers, given their different properties, fractions and whole numbers can be categorized as separate scales.

Finally, being able to flexibly structure scales with different anchors, or opposing ends of a continuum, is likely an important part of reasoning about scales both within and outside of human experience. There is nothing about a given anchor point that necessarily defines that value as the exact opposite of another anchor point; however, positioning them as such is a form of antithetical reasoning that can allow for the visualization of relative magnitude. There is convergent evidence human's internal representation of number is along a spatialized number line (Dehaene et al. 1993; de Havia and Spelke 2010). Having an accurate, and thus linear, mental number line supports a range of mathematical reasoning (Booth and Siegler 2008; Laski and Siegler 2007; Thompson and Siegler 2010). Children begin with a compressed representation of number, with smaller numbers occupying a relatively greater proportion of their mental number line and larger unfamiliar numbers occupying less (see Barth and Paladino (2011) and Opfer et al. (2011) for a discussion on mental models of magnitude representation). Linearity develops over time for an increasingly wider range of scales (Booth and Siegler 2006; Siegler and Booth 2004; Siegler and Opfer 2003), though some may never come to accurately reason about extremely small and large scales (Landy et al. 2013; Schneider and Siegler 2010).

Summary of Theoretical Framework

Above, we describe a theoretical framework for how relational reasoning supports reasoning about scale. We identified representational similarities between human and extreme scales and suggested that conceptual resources in small number processing are recycled in an analogous fashion to support reasoning about extreme scales. Dissimilarities were also identified, suggesting that extreme scales also require anomalous, antinomous, and antithetical reasoning. We suggested such reasoning functions to create conceptual change, with antithetical reasoning being particularly important in developing flexibility in representation between scales.

Barriers to Relational Reasoning

In the previous section, we present a theoretical framework for how relational reasoning may be involved in understanding scales far removed from human experience. Unfortunately, relational reasoning can fail to bring about conceptual change (Brown and Salter 2010; Duit 1991) and can often mislead students, resulting in misconceptions about the target concept that are hard to identify and resolve (Brown and Clement 1989; Duit 1991; Zook 1991; Zook and DiVesta 1991; Thagard 1992; Clement 1993; Zook and Maier 1994; Glynn 1995; Kaufman et al. 1996). In the current section, we identify barriers to alignment and discuss how they relate to reasoning about scales outside of human perception. While much of the literature reviewed here is on analogy, findings likely apply to all types of relational reasoning because anomaly, antinomy, and antithesis also require the alignment of a base and target concept (Alexander and DRLRL 2012).

There are a number of potential barriers in the alignment of human scales and extreme magnitude. Without a familiar base concept, for example, it can be difficult to identify the structure, making it all but impossible to make relevant connection to another concept (Gentner 1983; Kotovsky and Gentner 1996). While people may have experience with human scales, that does not ensure they will have a firm linear representation of the scale's magnitude. For example, while most humans likely encounter buildings of different sizes every day, it might not be clear how tall each building actually is.

It can also be difficult to identify relevant features to align if the base and target concepts have multiple differences, because the two concepts can be aligned in many different ways (Gentner and Gunn 2001; Kokinov and French 2003; Markman and Gentner 1996, 1997). This is particularly problematic when the unrelated features are more salient than the underlying structure. For example, a common analogy when explaining the geologic time scale is to map geologic time onto a 24-hour clock. The geologic time scale is a system of chronological measurement of Earth's history. Divisions of time are hierarchically organized based on major geologic events. The geologic time scale is conventionally depicted as a spatial representation, with Earth's formation (4.6 billion years ago) located at the bottom of a column(s) and present day located at the top (Fig. 1). However, there are a number of salient differences between the geologic time scale and a clock (Fig. 2). One salient difference is the temporally equal divisions of the clock (60 s = 1 min, 60 min = 1 h), which may lead novices to erroneously believe that the periods of Earth's history are also evenly spaced (which they are not). In this example, students are focusing on making an analogy between the distribution of divisions of time, and, thus failing to make an analogy between the relative magnitudes of time between events (e.g., to understand humans appeared relatively recently).

There may also be psychological barriers to alignment based on pre-existing spatial or functional characteristics. For example, when mapping an extreme scale to a cross-country road trip, categorization of state properties may influence magnitude recall in unintended ways (e.g., Friedman and Brown 2000; Stevens and Coupe 1978). Practical constraints on the classroom may also create barriers to understanding. For example, if aligning an extreme magnitude to a roll of toilet paper, the physical size of the classroom may necessitate bending the roll of toilet paper. In this case, it can be difficult to accurately represent the magnitude of the base concept, and, subsequently, be able to map magnitude relations between scales.

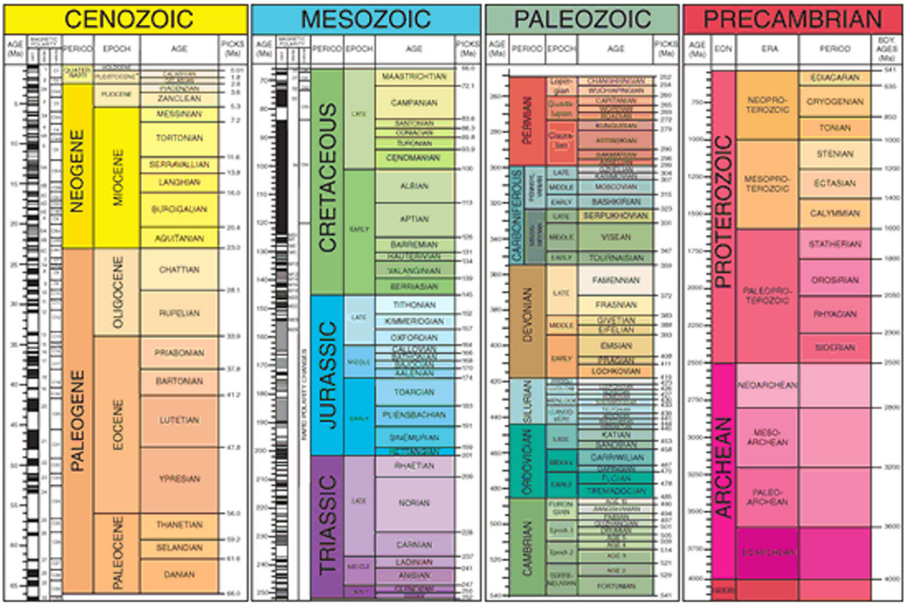

Fig. 1 The geologic time scale. Image from the Geological Society of America

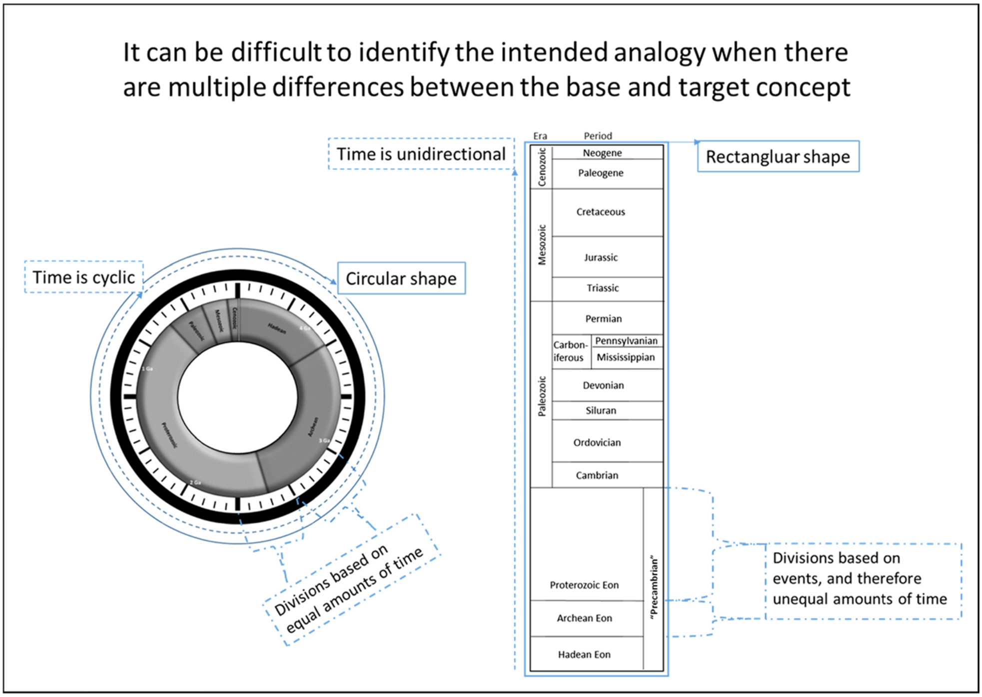

Fig. 2 Comparison between a spatial representation of the geologic time scale and the geologic time scale mapped onto a clock. In this example, we highlight three differences between both representations that may interfere with aligning relative magnitudes between the two scales

Scaffolding Relational Reasoning to Support Scale Learning: a Unified Scale Curriculum

To address the challenge of teaching magnitudes and phenomena outside of human perception, teachers have adopted a wide range of approaches. For example, one-part-per-billion has been represented as one hamburger in a chain of hamburgers circling the Earth's equator two and a half times (Kamrin et al. 1995). Geologic time has been mapped onto spatial structures, such as a roll of toilet paper, the Eiffel Tower, or a football field (Clary and Wandersee 2009; Wheeling Jesuit University 2004). The relative size of the universe has been depicted in visualizations like the famous "Powers of Ten" video (Eames and Eames 1968) and, more recently, websites like "The Scale of the Universe 2" (Huang and Huang 2012), which allow users to compare the sizes of different objects at each scale. Unfortunately, instruction does not always improve understanding. Even with these analogies, students continue to struggle comprehending phenomena and magnitudes at extreme scales (e.g., Delgado et al. 2007; Jones et al. 2008; Libarkin et al. 2005).

However, there are a number of approaches to scaffold relational reasoning. The act of making comparisons provides a path for experiential learning of relational structure (Gentner and Namy 1999). The more commonalities that exist between the base and target concepts, and if these commonalities are highlighted, the more salient corresponding relations will be (Gentner and Namy 2006). Highly similar base and target concepts also facilitate identifying how concepts are different (Goldstone 1994; Markman and Gentner 1993a, 1993b; Medin et al. 1993).

Using similar base and target concepts is not always possible, such as aligning human and extreme scales. The progressive alignment (Kotovsky and Gentner 1996; Thompson and Opfer 2010) of scales may alleviate the conceptual dissimilarity between human scales and extreme scales by providing more structural alignment across smaller increases of scale. Alignment of highly similar base and target concepts can increase the uniformity between the two representations, and, thus, helps to extend reasoning to more unfamiliar concepts (Gentner and Namy 2006). Specifically, the process of comparing two highly similar concepts can make higher-order relations more salient, which promotes identifying the same higher-order relations within unfamiliar concepts (Kotovsky and Gentner 1996). Progressive alignment also has the added benefit of providing learners with repeated opportunity to practice making relevant alignments (Resnick et al. 2016a, b).

A reason why novices may be less accurate when estimating magnitudes at extreme scales compared to human scales is because they have far fewer conceptual categories and less organization (Trend 2001). Thus, the hierarchical alignment of scale information can provide structure to populate the unfamiliar scale with salient category boundaries, or landmarks, and to emphasize the relation between scales (Resnick et al. 2016a, b). Hierarchical alignment refers to identifying the relation between all previous concepts when progressively aligning concepts. Seeing the causal relation between phenomena at multiple levels within and between scales, the common characteristics that hold together conceptual categories, and where they exist in the hierarchy can help develop an interconnected internal organization of scale (Resnick et al. 2012).

Successful Learning Activities

Below, we present two learning activities that use a spatial analogy to align the geologic time scale with a smaller more manageable scale—a number line. While the use of a spatial analogy to represent the geologic time scale is common (Libarkin et al. 2007), it is not always sufficient in fostering a sense of linear scale. For example, having repeated exposure to a conventional spatial representation of the geologic time scale (Resnick et al. 2013), mapping the geologic time scale onto a human lifespan (Petcovic and Ruhf 2008), or mapping the geologic time scale to a physical space that learners traverse (Semken et al. 2009) have not been effective in learning the magnitude involved in the geologic time scale (see Fig. 3 for examples why some common spatial analogies for geologic time may fail). The activities below incorporate the theoretical framework presented above for how relational reasoning supports understanding scale along with techniques for overcoming barriers to alignment.

Hierarchical Alignment Activity In the hierarchical alignment activity, learners begin by mapping a familiar scale onto a linear number line. The learners then progressively map successively larger scales onto the same amount of space. Using the same amount of space for the analogy provides structural alignment, highlighting one alignable difference—the magnitude of each scale. In each analogical step, the learners identify the relative locations of all previous scales, hierarchically organizing the scales (Fig. 4). For example, when learning geologic time, learners may construct ten separate timelines on to a 1-m space: personal history, an average human lifespan (from 75 years ago), American history (520 years ago), recorded history (5512 years ago), human evolution (6 million years ago (Ma)), Cenozoic Period (65 Ma), Phanerozoic Eon (542 Ma), Proterozoic Eon (2.5 billion years ago (Ga)), Archean Eon (3.8 Ga), and Hadean Eon (4.6 Ga—the full geologic time scale). Here, each timeline is based on conventionally defined boundaries (e.g., the Archean, Proterozoic, and Cenozoic are all divisions in the geologic time scale) that differ by orders of magnitude. In order to hierarchically populate each timeline with events and the relation between scales, the learner would locate all previous scales for each timeline, by calculating how many years each centimeter represented and then how many centimeters were required to represent each scale. The hierarchical alignment activity has been successful in reducing magnitude-based errors in estimation of both geologic events and astronomical distances, which also transferred to more accurate estimations of extreme abstract magnitude (Resnick et al. 2013).

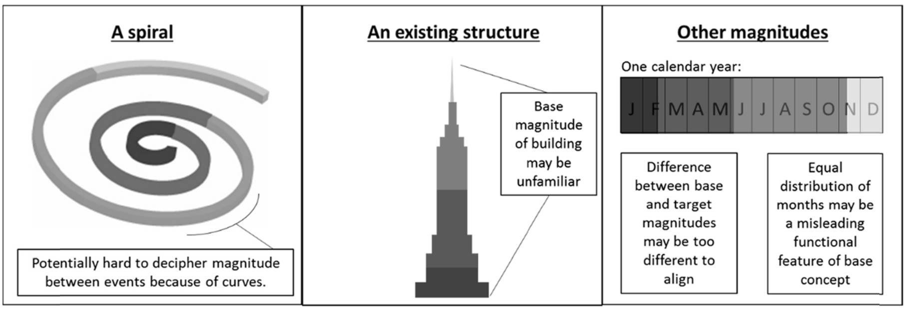

Fig. 3 Examples of analogies for the geologic time scale and potential barriers to alignment. *Lightest gray* = Phanerozoic, *light gray* = Proterozoic, *dark gray* = Archean, *darkest gray* = Hadean

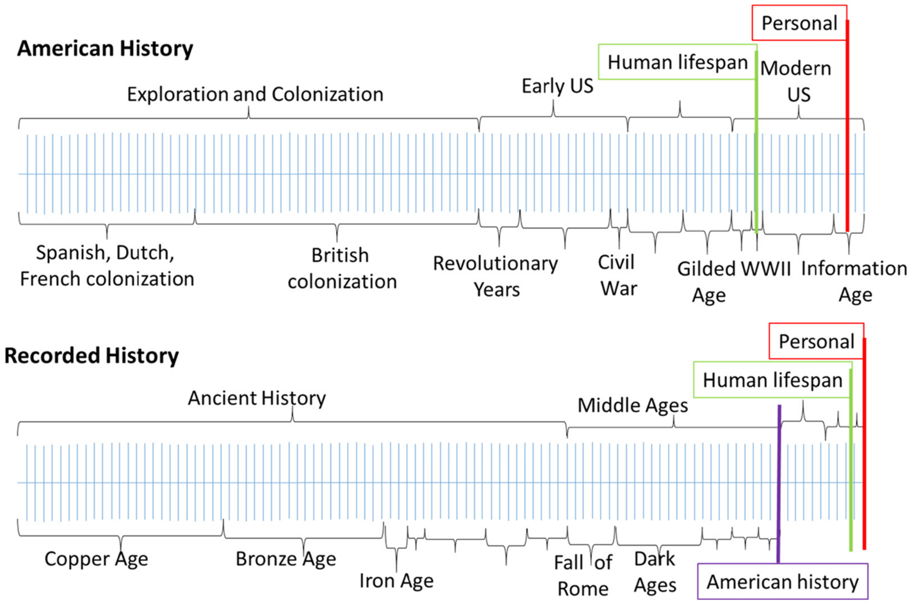

Fig. 4 Timelines at the hundred and thousand scale in the hierarchical alignment activity. The three previous timelines are located relative to the current scale

Corrective Feedback Activity In the corrective feedback activity, learners are presented with scale information, are asked to make predictions about the location of that scale information on a linear scale, and are then immediately provided with corrective feedback. Immediate (Coulter and Grossen 1997) and corrective (Sharpe et al. 1997) feedback promotes student understanding and is specifically effective for learning about unfamiliar magnitudes (Thompson and Opfer 2010). Learners should be provided with multiple opportunities to make different alignments, which can be implemented across several learning sessions. Spaced practice is beneficial, as students learn better when concepts are taught spaced across learning sessions as opposed to taught repeatedly within a shorter time span (e.g., Pashler et al. 2007).

When using the corrective feedback activity to teach geologic time, for example, students were first presented with an image of the geologic time scale, which conventionally is nonlinear, and directly alongside a blank linear timeline of equal length (Fig. 5). Students were asked where a particular division highlighted on the geologic time scale would be located on the linear scale and provided with four response options on the blank timeline. The students used a clicker response system, which is a handheld electronic device that allows entire classes to answer a multiple choice question simultaneously, to make their prediction. Clicker response systems can improve learning and engagement, particularly when paired with immediate corrective feedback (Kay and LeSage 2009). Students were then provided with immediate corrective feedback on a new slide (Fig. 6). The corrective feedback slide showed the conventional geologic time scale alongside a linear timeline. In both representations, the major categories of time were depicted, with the estimated division of time highlighted in color and connected by an arrow. Structural alignment is attained by having the conventional geologic time scale, the blank linear timeline, and the corrective feedback timeline all being

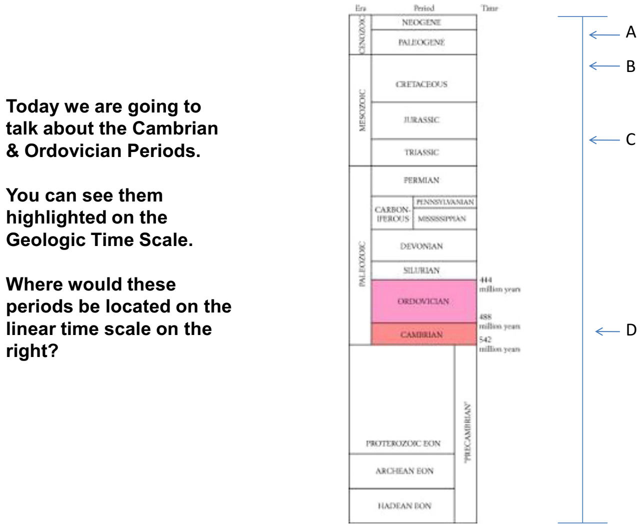

Fig. 5 Example of a clicker slide. Students are asked to locate where the highlighted divisions on the geologic time scale would be located on the linear timeline

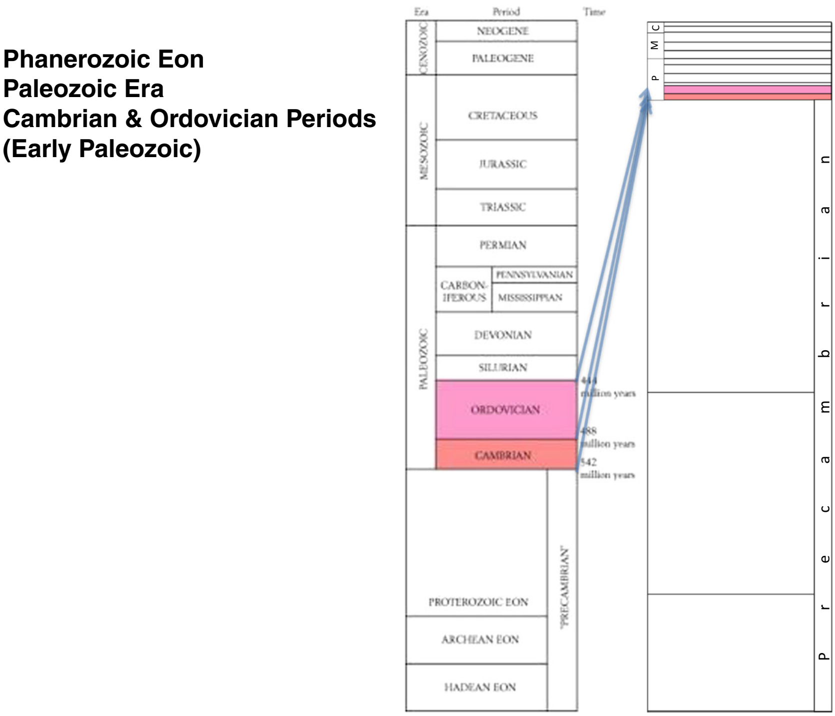

Fig. 6 Example of slide containing corrective feedback. The conventional geologic time scale is aligned with the linear time scale

of equal length and directly next to one another, showing the corresponding categories of time in the conventional geologic time scale and the corrective feedback timeline, and highlighting the estimated division in both representations using color and connecting them with arrows. The instructor completed ten different alignments over ten classes, with each alignment representing a major category of geologic time. Lecture material was framed around each temporal category.

The corrective feedback activity was associated with improved exam scores, more accurate estimations of extreme temporal magnitude, and transfer to more accurate estimations of extreme abstract magnitude (Resnick et al. 2013). Importantly, when just the slide aligning the geologic time scale with the linear timescale was presented, and students thus did not actively make a prediction about where scale information would be located on a linear scale, there was no improvement in temporal or abstract magnitude estimations and exam scores decreased (Resnick et al. 2013).

While any effect at the exam level may be surprising, it is important to note that the course content was structured around the geologic time scale, and, as such, a better understanding of how categories of geologic time relate to one another is directly relevant for exam performance. Further, the relation between exam performance and the corrective feedback activity was found across multiple semesters. It is hypothesized that when students understand the scale of geologic time, they are better able to place key course concepts (such as major evolutionary radiations, plate tectonics, and climate changes) into context, and, therefore, understand their underlying processes and systems.

More unexpected is that just seeing the alignment, without actively making a prediction, actually interfered with learning. Prediction may serve to engage the students into correctly aligning both representations. If students failed to understand the relation between the two representations, they would be left with multiple representations of the same content, which can interfere with understanding (Ainsworth 1999). Students often have difficulty aligning multiple representations (de Jong et al. 1998), a characteristic of expert understanding (Kozma et al. 2000).

Unification of Scaling Instruction Across STEM

Unifying themes across science education has been proposed as a mechanism to foster a more coherent curriculum (Tretter et al. 2006). By engaging in the core ideas and practices of science in different contexts, students are able to leverage a common language and familiarity to improve overall understanding (AAAS 1993). While learning from connected and integrated materials leads to improved retention, critical thinking, and problem solving (e.g., Ellis and Fouts 2001), US curriculum has unfortunately been characterized as being fractured and incoherent (Schmidt et al. 2002).

Scale is particularly poised to serve as an important unifying theme in STEM education; size, and scale represent foundational concepts across the STEM disciplines. One way to unify a scale curriculum is to develop cohesive techniques that help students align the vast set of scales across the sciences, thus building a foundation of scale understanding. The use of relational reasoning to teach concepts at extreme scales is particularly important, as very small and very large scales cannot be directly experienced (Jones et al. 2009).

Using our theoretical framework for how relational reasoning supports understanding of extreme scales, aligned with how people naturally reason about magnitude, is essential in developing an effective scale curriculum. This review found that a scale curriculum should provide learners with multiple opportunities to align magnitude to a spatial linear representation, using the same amount of space for each alignment (structural alignment). Students should also be engaged in actively making predictions about where scale information would be located on a linear scale; the use of a clicker response system may be a useful technique. Immediate corrective feedback and spaced instruction are also useful in establishing salient landmarks. Finally, progressive and hierarchical alignments appear to have an additive benefit on learning scale information (Resnick et al. 2016a, b).

The hierarchical alignment activity and the corrective feedback activity are examples of how a scale curriculum might be designed. Importantly, both activities are designed to be transposable for teaching any magnitude-based context. Future investigations may examine the benefit of science of learning principles, such as active prediction, immediate feedback, spaced learning, and the use of visual representations. However, it is clear that relational reasoning and analogy are paramount in understanding many foundational scientific concepts and phenomena that take place at extreme scales far removed from human experience.

Acknowledgments This research was supported by the National Science Foundation Grants SBE-0541957 and SBE-1041707 which support the NSF funded Spatial Intelligence Learning Center and the Institute of Education Sciences Grant R305B130012 as part of the Postdoctoral Research Training Program in the Education Sciences.

References

Ainsworth, S. E. (1999). The functions of multiple representations. *Computers & Education, 33*, 131–152.

Alexander, P. A., & the Disciplined Reading and Learning Research Laboratory. (2012). Reading into the future: competence for the 21st century. *Educational Psychologist, 47*, 259–28.

Alexander, P. A., Jablansky, S., Singer, L. M., & Dumas, D. (2016). Relational reasoning: what we know and why it matters. *Policy Insights from the Behavioral and Brain Sciences, 3*(1), 36–44. doi:10.1177/2372732215622029.

American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. New York: Oxford University Press.

Barth, H. C., & Paladino, A. M. (2011). The development of numerical estimation: evidence against a representational shift. *Developmental Science, 14*(1), 125–135. doi:10.1111/j.1467-7687.2010.00962.x.

Booth, J. L., & Siegler, R. S. (2006). Developmental and individual differences in pure numerical estimation. *Developmental Psychology, 42*, 189–201. doi:10.1037/0012-1649.41.6.189.

Booth, J. L., & Siegler, R. (2008). Numerical magnitude representations influence arithmetic learning. *Child Development, 79*(4), 1016–1031.

Brown, D., & Clement, J. (1989). Overcoming misconceptions via analogical reasoning: abstract transfer versus explanatory model construction. *Instructional Science, 18*, 237–261.

Brown, S., & Salter, S. (2010). Analogies in science and science teaching. *Advanced Physiological Education, 34*, 167–169. doi:10.1152/advan.00022.2010.

Bueti, D., & Walsh, V. (2009). The parietal cortex and the representation of time, space, number and other magnitudes. *Philosophical Transactions of the Royal Society B, 364*, 1831–1840.

Callanan, M. A., & Markman, E. M. (1982). Principles of organization in young children's natural language hierarchies. *Child Development, 53*, 1093–1101.

Cantlon, J. F., Platt, M. L., & Brannon, E. M. (2009). Beyond the number domain. *Trends in Cognitive Science, 13*, 83–91.

Carpenter, T. P., & Moser, J. M. (1983). The acquisition of addition and subtraction concepts. In R. Lesh & M. Landau (Eds.), *Acquisition of mathematics concepts and processes*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Clary, R. M., & Wandersee, J. H. (2009). Tried and true: How Old? Tested and trouble-free ways to convey geologic time. *ScienceScope, 33*(4), 62–66.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30*, 1241–1257.

Cohen Kadosh, R., Lammertyn, J., & Izard, V. (2008). Are numbers special? An overview of chronometric, neuroimaging, developmental and comparative studies of magnitude representation. *Progress in Neurobiology, 84*, 132–147.
Coulter, G. A., & Grossen, B. (1997). The effectiveness of in-class instructive feedback versus after-class instructive feedback for teachers learning direct instruction teaching behaviors. *Effective School Practices, 16*, 21–34.
de Havia, M. D., & Spelke, E. S. (2010). Number-space mapping in human infants. *Psychological Science, 21*, 653–660.
de Jong, T., Ainsworth, S., Dobson, M., van der Hulst, A., Levonen, J., Reimann, P., Sime, J., van Someren, M., Spada, H., & Swaak, J. (1998). Acquiring knowledge in science and math: the use of multiple representations in technology based learning environments. In M. W. Van Someren, P. Reimann, H. Bozhimen, & T. de Jong (Eds.), *Learning with multiple representations* (pp. 9–40). Amsterdam: Elsevier.
Dehaene, S., Bossini, S., & Giraux, P. (1993). The mental representation of parity and numerical magnitude. *Journal of Experimental Psychology: General, 122*, 371–396.
Delgado, C., Stevens, S., Shin, N., Yunker, M., & Krajcik, J. (2007). *The development of Students' conceptions of size*. New Orleans, LA: National Association of Research in Science Teaching.
Duit, R. (1991). On the role of analogies and metaphors in learning science. *Science Education, 30*, 1241–1257.
Dumas, D., Alexander, P. A., & Grossnickle, E. M. (2013). Relational reasoning and its manifestations in the educational context: a systematic review of the literature. *Educational Psychology Review, 25*, 391–427.
Eames, C. & Eames, R. (1968). *Powers of Ten* [Motion picture]. USA: IBM.
Ellis, A. K., & Fouts, J. T. (2001). Interdisciplinary curriculum: the research base. *Music Educators Journal, 87*, 22–27.
Friedman, A., & Brown, N. R. (2000). Reasoning about geography. *Journal of Experimental Psychology. General, 129*, 193–219.
Galilei, G. (1638). *Two new sciences*. Madison, WI: University of Wisconsin Press.
Gentner, D. (1982). Are scientific analogies metaphors? In D. S. Miall (Ed.), *Metaphor: Problems and perspectives* (pp. 106–132). Brighton, England: Harvester Press Ltd.
Gentner, D. (1983). Structure-mapping: a theoretical framework for analogy. *Cognitive Science, 7*, 155–170.
Gentner, D., & Gentner, D. R. (1983). Flowing waters or teeming crowds: Mental models of electricity. In D. Gentner & A. L. Stevens (Eds.), Mental models (pp. 99–129). Hillsdale, NJ: Lawrence Erlbaum Associates. (Reprinted in M. J. Brosnan (Ed.), *Cognitive functions: Classic readings in representation and reasoning*. Eltham, London: Greenwich University Press).
Gentner, D., & Gunn, V. (2001). Structural alignment facilitates the noticing of differences. *Memory and Cognition, 29*(4), 565–577.
Gentner, D., & Holyoak, K. J. (1997). Reasoning and learning by analogy: introduction. *American Psychologist, 52*, 32–34.
Gentner, D., & Namy, L. (1999). Comparison in the development of categories. *Cognitive Development, 14*, 487–513.
Gentner & Namy. (2006). Analogical Processes in Language Learning. *Association for Psychological Science, 15*(6).
Gentner, D., Loewenstein, J., & Hung, B. (2007). Comparison facilitates children's learning of names for parts. *Journal of Cognition and Development, 8*(3), 285–307.
Glynn, S. (1995). Conceptual bridges: using analogies to explain scientific concepts. *Science Teacher, 62*, 24–27.
Goldstone, R. L. (1994). The role of similarity in categorization: Providing a groundwork. *Cognition, 52*, 125–157.
Halford, G. S. (1993) *Children's understanding: the development of mental models*. Erlbaum. [DBB, DG, arGSH].
Hawkins, D. (1978), Critical barriers to science learning, *Outlook, 29.*
Huang, C., & Huang, M. (2012). The scale of the universe 2. Copyright: Cary and Michael Huang. Retrieved from http://htwins.net/scale2/.
Hummel, J. E., & Holyoak, K. J. (2003). A symbolic-connectionist theory of relational inference and generalization. *Psychological Review, 110*, 220–264.
Huttenlocher, J., Hedges, L., & Prohaska, V. (1988). Hierarchical organization in ordered domains: Estimating the dates of events. *Psychological Review, 95*, 471–484.
Huttenlocher, J. E., Hedges, L. V., & Vevea, J. L. (2000). Why do categories affect stimulus judgment? *Journal of Experimental Psychology. General, 129*, 220–241.
James, W. (1890). *The principles of psychology.* New York, NY: Henry Holt.
Jones, M. G., & Taylor, A. R. (2009). Developing a sense of scale: looking backward. *Journal of Research in Science Teaching, 46*(4), 460–475.
Jones, M. G., Tretter, T., Taylor, A., & Oppewal, T. (2008). Experienced and novice teachers' concepts of spatial scale. *International Journal of Science Education, 30*(3), 409–429.

Kamrin, M. A., Katz, D. J., & Walter, M. L. (1995). *Reporting on risk. A Journalist's handbook. Ann Arbor: Michigan sea grant college program, 1994* (2nd ed.). Los Angeles: Foundation for American Communications and National Sea Grant College Program.

Kaufman, D. R., Patel, V. L., & Magder, S. A. (1996). The explanatory role of spontaneously generated analogies in reasoning about physiological concepts. *International Journal of Science Education, 18*, 369–386.

Kay, R. H., & LeSage, A. (2009). A strategic assessment of audience response systems used in higher education. *Australian Journal of Educational Technology, 25*(2), 235-249.

Kokinov, B., & French, R. M. (2003). Computational models of analogy making. In L. Nadel (Ed.), *Encyclopedia of cognitive science* (pp. 113–118). London: MacMillan.

Kotovsky, L., & Gentner, D. (1996). Comparison and categorization in the development of relational similarity. *Child Development, 67*, 2797–2822.

Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The role of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*(3), 105–144.

Lamon, S. (1994). Ratio and proportion: cognitive foundations in unitizing and norming. In G. J. Harel Confrey (Ed.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 89–120). Albany, NY: State University of New York Press.

Landy, D., Silbert, N., & Goldin, A. (2013). Estimating large numbers. *Cognitive Science, 37*, 775–799.

Landy, D., Charlesworth, A., & Ottmar, E. (2014). Cutting in line: discontinuities in the use of large numbers in adults. *Proceedings of the 36th Annual Conference of the Cognitive Science Society.* Quebec City, Quebec: Cognitive Science Society.

Laski, E., & Siegler, R. (2007). Is 27 a Big number? correlational and causal connections among numerical categorization, number line estimation, and numerical magnitude comparison. *Child Development, 78*(6), 1723–1743.

Libarkin, J. C., Anderson, S. W., Dahl, J., Beilfuss, M., & Boone, W. (2005). Qualitative analysis of college Students' ideas about the earth: interviews and open-ended questionnaires. *Journal of Geoscience Education, 53*(1), 17–26.

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

Lourenco, S. F., & Longo, M. R. (2011). Origins and development of generalized magnitude representation. In S. Dehaene & E. M. Brannon (Eds.), *Space, time and number in the brain: Searching for the foundations of mathematical thought* (pp. 225–244). London: Elsevier.

Markman, A. B., & Gentner, D. (1993a). Structural alignment during similarity comparisons. *Cognitive Psychology, 25*, 431–467.

Markman, A. B., & Gentner, D. (1993b). Splitting the differences: A structural alignment view of similarity. *Journal of Memory and Language, 32*, 517–535.

Markman, A. B., & Gentner, D. (1996). Commonalities and differences in similarity comparisons. *Memory and Cognition, 24*, 235–249.

Markman, A. B., & Gentner, D. (1997). The effects of alignability on memory. *Psychological Science, 8*, 363c367.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*(2), 254–278.

Miller & Brewer. (2010). Misconceptions of Astronomical Distances. *International Journal of Science Education, 32*(12).

National Research Council. (2011). *A framework for K-12 science education* (Committee on a conceptual framework for new K-12 science education standards. Board on science education, DBASSE). Washington, DC: The National Academies Press.

Opfer, J. E., Siegler, R. S., & Young, C. J. (2011). The powers of noise-fitting: reply to Barth and Paladino. *Developmental Science, 14*, 1194–1204.

Orgill, M., & Bodner, G. (2004). What research tells us about using analogies to teach chemistry. *Chemistry Education Research and Practice, 5*, 15–32.

Pashler, H., Bain, P., Bottge, B., Graesser, A., Koedinger, K., McDaniel, M., & Metcalfe, J. (2007). *Organizing instruction and study to improve student learning* (*NCER 2007-2004*). Washington, DC: National Center for Education Research, Institute of Education Sciences, U.S. Department of Education. Retrieved from http://ncer.ed.gov.

Petcovic & Ruhf. (2008). Geoscience conceptual knowledge of preservice elementary teachers: results from the geoscience concept inventory. *Journal of Geoscience Education, 56*(3), 251–260.

Resnick, I., Atit, K., & Shipley, T. F. (2012). Teaching geologic events to understand geologic time. In K. A. Kastens & C. A. Manduca (Eds.), *Earth and mind II: a synthesis of research on thinking and learning in the geosciences: geological society of america special paper 486.* Boulder, Colorado: The Geological Society of America, Inc.

Resnick, I., Davatzes. A., & Shipley, T. F. (2013). *Using analogy to improve understanding of large magnitudes.* Poster presented at the Improving Middle School Science Instruction Using Cognitive Science Conference, Washington D.C.

Resnick, I., Jordan, N. C., Hansen, N., Rajan, V., Rodrigues, J., Siegler, R. S., & Fuchs, L. (2016). Developmental growth trajectories in understanding of fraction magnitude from fourth through sixth grade. *Developmental Psychology*.

Resnick, I., Newcombe, N. S., & Shipley, T. F. (2016). Dealing with big numbers: Representation and understanding of magnitudes outside of human experience. *Cognitive Science*.

Riebeek, H. (2010, June). Global Warming. Retrieved from: http://earthobservatory.nasa.gov/Features/GlobalWarming/.

Schmidt, W. H., Houang, R., & Cogan, L. (2002). A coherent curriculum: The case of mathematics. *American Educator, 26*.

Schneider, M., & Siegler, R. S. (2010). Representations of the magnitudes of fractions. *Journal of Experimental Psychology. Human Perception and Performance, 36*(5), 1227–1238. doi:10.1037/a0018170.

Semken, S., Dodick, J., Ben-David, O., Pineda, M., Bueno Watts, N., & Karlstrom, K. (2009). Timeline and time scale cognition experiments for a geological interpretative exhibit at Grand Canyon. *Proceedings of the National Association for Research in Science Teaching*, Garden Grove, California.

Sharpe, T. L., Lounsbery, M., & Bahls, V. (1997). Description and effects of sequential behavior practice in teacher education. *Research Quarterly for Exercise and Sport, 68*, 222–232.

Shipley, T. F., & Zacks, J. (2008). *Understanding events: from perception to action*. New York: NY, Oxford University Press.

Siegler, R. S., & Booth, J. L. (2004). Development of numerical estimation in young children. *Child Development, 75*, 428–444. doi:10.1111/j.1467-8624.2004.00684.x.

Siegler, R. S., & Lortie-Forgues, H. (2014). An integrative theory of numerical development. *Child Development Perspectives, 8*, 144–150.

Siegler, R. S., & Opfer, J. E. (2003). The development of numerical estimation: Evidence for multiple representations of numerical quantity. *Psychological Science, 14*.

Siegler, R. S., Thompson, C. A., & Schneider, M. (2011). An integrated theory of whole number and fractions development. *Cognitive Psychology, 62*(4), 273–296. doi:10.1016/j.cogpsych.2011.03.001.

Simons, P. R. J. (1984). Instructing with analogies. *Journal of Educational Psychology, 76*, 513–527.

Stevens, A., & Coupe, P. (1978). Distortions in judged spatial relations. *Cognitive Psychology, 10*, 422–437.

Thagard, P. (1992). Analogy, explanation, and education. *Journal of Research in Science Teaching, 29*, 537–544.

Thompson, C., & Opfer, J. (2010). How 15 hundred is like 15 cherries: Effect of progressive alignment on representational changes in numerical cognition. *Child Development*.

Thompson, C. A., & Siegler, R. S. (2010). Linear numerical magnitude representations aid children's memory for numbers. *Psychological Science, 21*, 1274–1281.

Trend, R. D. (2001). Deep time framework: a preliminary study of UK primary teachers' conceptions of geological time and perceptions of geoscience. *Journal of Research in Science Teaching, 38*(2), 191–221.

Tretter, T. R., Jones, M. G., Andre, T., Negishi, A., & Minogue, J. (2006). Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching, 43*(3), 282–319.

Vosniadou, S., & Mason, L. (2012). Conceptual change induced by instruction: a complex interplay of multiple factors. In K. Harris, S. Graham, & T. Urdan (Eds.), *Handbook of educational psychology* (2nd ed., pp. 221–246). Washington, DC: American Psychiatric Association.

Walsh, V. (2003). A theory of magnitude: common cortical metrics of time, space and quantity. *TRENDS in Cognitive Sciences, 7*(11).

Wheeling Jesuit University. (2004). Geologic time activity. Copy right 1997-2004 by Wheeling Jesuit University/NASA-supported Classroom of the Future.

Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin, 127*, 3–21.

Zook, K. B. (1991). Effects of analogical processes on learning and misrepresentation. *Educational Psychology Review, 3*, 41–72.

Zook, K. B., & DiVesta, F. J. (1991). Instructional analogies and conceptual misrepresentations. *Journal of Educational Psychology, 83*, 246–252.

Zook, K. B., & Maier, J. M. (1994). Systematic analysis of variables that contribute to the formation of analogical misconceptions. *Journal of Educational Psychology, 86*, 589–699.