



# Click-On-Diagram Questions: a New Tool to Study Conceptions Using Classroom Response Systems

Nicole D. LaDue<sup>1</sup> · Thomas F. Shipley<sup>2</sup>

Published online: 6 June 2018  
© Springer Science+Business Media, LLC, part of Springer Nature 2018

## Abstract

Geoscience instructors depend upon photos, diagrams, and other visualizations to depict geologic structures and processes that occur over a wide range of temporal and spatial scales. This proof-of-concept study tests click-on-diagram (COD) questions, administered using a classroom response system (CRS), as a research tool for identifying spatial misconceptions. First, we propose a categorization of spatial conceptions associated with geoscience concepts. Second, we implemented the COD questions in an undergraduate introductory geology course. Each question was implemented three times: pre-instruction, post-instruction, and at the end of the course to evaluate the stability of students' conceptual understanding. We classified each instance as (1) a false belief that was easily remediated, (2) a flawed mental model that was not fully transformed, or (3) a robust misconception that persisted despite targeted instruction. Geographic Information System (GIS) software facilitated spatial analysis of students' answers. The COD data confirmed known misconceptions about Earth's structure, geologic time, and base level and revealed a novel robust misconception about hot spot formation. Questions with complex spatial attributes were less likely to change following instruction and more likely to be classified as a robust misconception. COD questions provided efficient access to students' conceptual understanding. CRS-administered COD questions present an opportunity to gather spatial conceptions with large groups of students, immediately, building the knowledge base about students' misconceptions and providing feedback to guide instruction.

**Keywords** Clickers · Diagrams · Conceptions · Mental models · Spatial thinking · Geology

## Introduction

Introductory, general education courses must provide students with the knowledge and skills necessary for civic participation. Students come to class with naïve theories about natural phenomena (Vosniadou 2002; Vosniadou and Brewer 1987; Brewer 2008) and conceptual frameworks that diverge from the scientific consensus (Chi 2008). Changing students' conceptions of scientific phenomena can consist of simple belief revisions (i.e., *false beliefs*) that are easily remediated in the presence of new information. The process of conceptual change can also be incremental. Students' mental models may become more scientifically

accurate, while still being incomplete or containing errors (i.e., *flawed mental model*) (Chi 2008). Highly resistant conceptions that do not change despite rigorous contradictions (i.e., *robust misconceptions*) may be so because they require an ontological shift to a different conceptual category with different plausible attributes (Chi 2008; Lombardi et al. 2016). For example, students may have a misconception of heat as an entity (i.e., heat "rises") rather than a process (i.e., heat as the process of molecules moving at various speeds) (Chi 2008). In this example, students must be able to switch the category to which they ascribe the concept of heat in order to hold a scientific conception of heat. To access this level of conceptual understanding, most researchers studying conceptual change use interviews and drawings to capture students' thinking (Cheek 2010; Gurel et al. 2015). While these methods provide detailed information about students' conceptual frameworks and can be used to develop new approaches to instruction, they are too time-intensive to allow large-scale surveys of students' concepts.

Student conceptions in geology are complicated by the unfamiliar spatial and temporal scales on which geologic structures

---

✉ Nicole D. LaDue  
nldue@niu.edu

<sup>1</sup> Geology and Environmental Geosciences Department, Northern Illinois University, 312 Davis Hall, DeKalb, IL 60115, USA

<sup>2</sup> Department of Psychology, Temple University, 325 Weiss Hall, Philadelphia, PA 19122, USA

and processes occur (Tretter et al. 2006; Jones et al. 2008; Cheek 2013). Learning materials for geology mirror this spatial complexity by integrating a variety of diagrams (Cheek et al. 2017; Kastens and Rivet 2010; LaDue et al. 2015; Manduca and Kastens 2012). Diagrams can be more efficient for learning than text descriptions (Larkin and Simon 1987; Sweller 1994), particularly if they are aligned with the task they serve (Schnotz and Bannert 2003; Hegarty et al. 2010). Nevertheless, interpreting geology diagrams can range from simple identification tasks to more complex spatial tasks (Kastens and Rivet 2010). Visualizing a three-dimensional structure, where the viewer must make sense of relationships within or between pieces of information contained in the diagram (Newcombe and Shipley 2015), and mentally animating or unraveling a sequence of events (Kastens et al. 2014) are both important geoscience thinking skills.

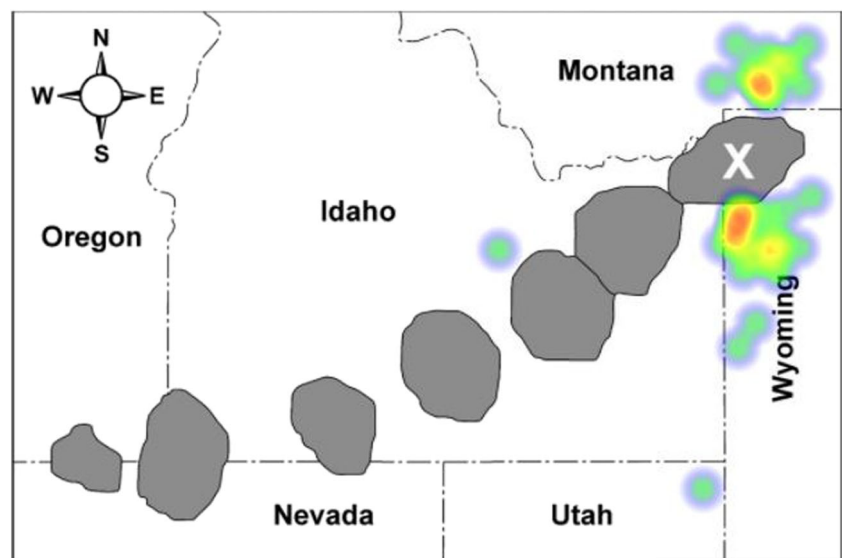
Drawing, sketching, or diagram labeling tasks can be a useful assessment of students' internal mental representations of a structure or process in science (Forbus et al. 2017). However, student-generated sketches still require researchers to make a significant time investment to analyze the data (Clark and Libarkin 2011). A new application of classroom response systems (CRS) offers instructors and science education researchers a spatial tool to access students' conceptions in geology. CRS, also known as clickers, are now offering a greater variety of response formats (Shea 2016) that can be exploited to investigate spatial aspects of students' conceptions. Although adoption by undergraduate instructors has been slow (Emenike and Holme 2012), CRS technology offers the opportunity for the educational researcher to study the spatial nature of students' conceptions.

Modern CRS technology places fewer limitations on students because web-based and app-based interfaces allow students to participate on smart devices, such as cell phones and tablets. Students can indicate their answer by clicking on their

device screen. Instructors can now ask students click-on-diagram (COD) questions to identify objects within an image. The COD question format allows instructors to view the spatial distribution of responses as a smoothed probability density function, referred to as a heat map, where the color represents the relative concentration of responses (Fig. 1). This immediate feedback tells students and instructors, alike, a great deal about the distribution and nature of students' understanding. Gagnier et al. (2017) showed that providing students with spatial feedback on the nature of their error supported the development of students' mental models. CODs are also an open-ended response option that does not require researchers to construct multiple-choice options based on hunches about students' alternative conceptions.

The present study explores the opportunity to identify novel misconceptions using COD questions. We situate this study within Chi's (2008) framework of conceptual change because we observed that some spatial concepts were easily adjusted (*false belief*) and others were resistant to multiple iterations of instruction and displayed systematic errors (*robust misconception*). Central to the efficacy in revealing student concepts is the spatial information provided by CODs. The examples presented below demonstrate the unique utility of this tool to engage students in spatial feedback (Gagnier et al. 2017). Deploying CODs over the course of instruction can reveal particularly robust spatial misconceptions that warrant further attention by both instructor and researcher. We present examples of spatial reasoning problems associated with time, scale, and multiple frames-of-reference. The focus of this study is the utility of COD questions as a research tool. Researchers need expedient ways to identify previously unrecognized errors that are present in students' mental models, evaluate their prevalence and impact on learning, and develop new fruitful lines of research on teaching geoscience concepts.

**Fig. 1** Top Hat generated heat map of students' post-instruction responses to a question about where the next hot spot will form if the plate started moving to the North ( $N = 33$ )



## The Structure of Students' Conceptions

Students' conceptions may take the form of mental models, or frameworks, representing the relationships between ideas in science (Vosniadou and Brewer 1992; Vosniadou et al. 2008). These mental models change gradually in the presence of new information (i.e., assimilation) or are replaced when conceptual conflicts with existing knowledge are great (i.e., accommodation) (Posner et al. 1982). A useful example from Vosniadou and Brewer (1992) describes how young children who start with a mental model that the earth is flat describe a variety of other shapes (i.e., rectangular solid, flattened disc) while trying to reconcile their models with scientific representations of Earth as a sphere. Following Chi (2008), we use the term *mental models* to indicate coherent explanations of a phenomena that can be probed with predictable results. This framework is a good fit for the geology conceptions we are targeting because the questions are based in diagrams, which are coherent, visual models that may influence the structure of students' conceptions.

Misconceptions are characterized by a "systematicity in the pattern of [students'] responses" (Chi 2008, p.68). Chi's (2008) framework of conceptual change provides suggestions for how one might categorize conceptions that are easily revised versus those that are more resistant to change. *False beliefs* are described as missing or incomplete conceptions about a single idea (Chi 2008, p.66). When confronted with information about the scientific model, simple false beliefs will be easily revised. For example, students believe the center of the Earth is made of hot liquid. Later in this study, we demonstrate how this stand-alone concept is easily revised to the scientific conception of a liquid outer core surrounding a solid inner core following instruction on the Earth's interior. Chi (2008) distinguishes a *robust misconception* from a *flawed mental model* such that correcting a robust misconception requires people to re-categorize something into a different ontological category (i.e., a conceptual category with different attributes, such as matter versus process; Chi 1997; Libarkin and Kurdziel 2006). Our goal is to develop a technique for researchers to identify the difficulties related to the spatial nature of students' geoscience conceptions; therefore, the present study is situated using Chi's (2008) framework

because we hypothesize that spatial conceptions in geology will fall into distinct ontological categories.

## Conceptions in Geology

Several reports have synthesized the extant literature on geology misconceptions (Cheek 2010; Francek 2013; Dove 1998; King 2008), yet there are substantially fewer analyses of conceptual understanding in geology compared to physics and chemistry (Cheek 2010; National Research Council 2012). The majority of peer-reviewed studies on misconceptions in geology have focused on plate tectonics, weathering and erosion, rocks and minerals, and geologic time (Francek 2013). In some cases, students' misconceptions appear to stem from early learning experiences in and out of school. The language (King 2010) and diagrams (Clark et al. 2011) used in common textbooks are also a source of misconceptions. Dolphin and Benoit (2016) explored the metaphor of tectonic plates and found that every day knowledge about dinner plates interferes with students' conceptual model of plate tectonics. Finally, the difficulty in knowing how to help students apply what they know about common events to the events on the spatial and temporal scales of geosciences means that conceptions of geologic processes are relatively understudied (Cheek 2010).

One difficulty developing robust conceptual understanding of geology is that geologic processes are frequently unobservable on the spatial and temporal scales on which they occur. Drawing from existing frameworks for spatial thinking, we propose three distinct spatial ontological categories that drive students' misconceptions in geology (Table 1). First, to predict how a geologic structure originally formed people must integrate multiple distinct spatial patterns to infer a geologic process (Shiple et al. 2013). When a geologist looks at a rock outcrop, they may engage in *retrodiction* (Trend 2000; Ault 1998), or infer past events from a present situation. For example, a geologist who sees folded rock layers may mentally unfold the layers to evaluate the sequence of events that led to the present appearance of the rocks. This spatial visualization achievement requires mental animation, or mentally transforming objects in rigid and non-rigid ways (Resnick and Shiple 2013; Atit et al. 2013). Second, students struggle

**Table 1** Categorizing the ontology of spatial conceptions in geology with examples and existing spatial processes

Spatial category	Example	Spatial cognitive process
Spatial integration	Combining field observations to construct a geologic explanation (Shiple et al. 2013)	Spatial visualization (Newcombe and Shiple 2015)
Scaler relationships	Spacing of geologic events in deep time (Libarkin et al. 2007)	Spatial visualization (Newcombe and Shiple 2015)
Spatial reference frame	Mentally animating the geologic processes forming a static structure (Shiple et al. 2013)	Mental animation (Newcombe and Shiple 2015) Perspective-taking (Hegarty and Waller 2004) Retrodiction (Ault 1998; Trend 2000)

with the scale of deep geologic time. Understanding the magnitude and relationship of extremely old events or extremely large structures, beyond the scope of the average human being, is especially challenging (Cheek et al. 2017; Tretter et al. 2006; Jones et al. 2008). Students have a sense of the relative order of events in geologic time, but struggle to place those events in an accurate spacing along a geologic timeline (Libarkin et al. 2007). This may be, in part, caused by the non-linear representations of geologic time that are used in geology instruction and in part student's categorical representation of time (Resnick et al. 2017). Third, understanding many geologic processes requires the observer to imagine multiple frames of reference (Wade 1996; Wade and Swanston 2013). People approaching a problem are likely to use human-centric reference frames (Carlson 1999), while geologic processes require alternative reference frames. For example, many geologic hot spots are laterally stationary relative to the Earth's mantle (Konrad et al. 2018). As a tectonic plate moves over the hot spot, it leaves a chain of successively younger volcanic calderas. To understand the past from the present, a geologist has to imagine the ground beneath them is movable relative to the stationary mantle reference frame beneath the lithosphere. This brief review is not intended to be exhaustive, but to ground this study in the extant literature on the spatial demands of geoscience understanding. Not surprisingly, geologists self-report high levels of spatial skill in reasoning about both small-scale objects and large-scale navigation (Hegarty et al. 2010).

Overall, spatial and temporal reasoning in the geosciences remains understudied in the research literature (National Research Council 2012). For a discipline so enmeshed in diagrams (LaDue et al. 2015) and spatial concepts (Shipley et al. 2013), there is troublingly scant research on students' conceptions that utilizes diagrams and spatial reasoning tasks (Cheek 2010). The present study explores a tool that may accelerate research on students' spatial thinking and conceptions of diagrams associated with geologic structures and processes.

## Methods for Studying Geology Conceptions

In a review of geoscience learning, Cheek (2010) offered preliminary categorization of both geologic concepts and the methods researchers have used to study them. The majority of the studies were descriptive, focusing on interviews to elicit students' conceptual frameworks (Cheek 2010). Only 12 of the 79 studies reviewed by Cheek (2010) required students to draw, and an additional 8 studies asked students to engage in a visually mediated, physical task (i.e., card sorting, mineral identification, concept mapping). In the broader education literature, studies of students' science conceptions primarily utilize interviews, open-ended tests, and multiple-choice tests (Gurel et al. 2015; Cheek 2010; Mills et al. 2016). Often, the

goal of such efforts is to develop concept inventories (Hestenes et al. 1992; Libarkin and Anderson 2005; Garvin-Doxas and Klymkowsky 2008). Interviews and open-ended questions provide in-depth information about the students' conceptions and flexibility for any given specific question, but such methods are time-consuming both for the participant and researcher. To address this challenge, researchers have developed multiple-tier misconceptions instruments to merge the strengths of qualitative methods and mitigate the weaknesses of quantitative measures (Gurel et al. 2015). Despite this recent effort to balance research method efficiency and data quality, using these approaches to uncover the spatial nature of students' conceptions is limited by the spatial language available to the participants or interpretation by the researchers. Given the demands of time, it is understandable that few studies in the geosciences have utilized visual methods (Cheek 2010).

## A New Tool to Study Conceptions and Diagrams

COD questions have similar affordances to traditional data collection, where students label diagrams (Libarkin et al. 2007) or sketch (Gagnier et al. 2017; Libarkin 2006), but offer greater efficiency for gathering large data sets. COD questions can be administered in large or small classes alike with the same time to prepare and analyze data; they can yield a large volume of research-quality data from a broad range of students. More importantly, research using COD questions target gaps in the geoscience conceptions literature on diagrams (King 2010) and spatial thinking (Cheek 2010). Since they are open-ended, and students can click anywhere on the image, they do not require anticipating which, if any, misconceptions students may have. This response style has existed for a few years, yet few studies have combined CRS technology with research on diagrams or spatial thinking.

Two recent studies have included images to investigate students' conceptions associated with astronomy. Lee and Schneider (2015) developed a set of clicker questions with accompanying photographs that may be used in large lecture. Students were asked to view a photo while answering the question, but were not asked to click on the photo (Lee and Schneider 2015). The questions required students to retrieve knowledge about astronomical structure and features, not decipher spatial relations between the structures. This type of question is an identification question and has the power to reveal a *false belief* (Chi 2008). However, the information obtained from the question does not provide further insight into students' mental models.

A second astronomy study by Lee and Feldman (2015) used the Technology-Enhanced Formative Assessment (TEFA; Beatty and Gerace 2009) model to engage middle

school students in collaborative reasoning associated with CRS questions. The questions related to the time of day and year associated with photos of the sun's position relative to the horizon. This type of question asks students to consider the spatial relationship between actual or inferred objects in the image. In contrast to Lee and Schneider (2015), Lee and Feldman (2015) did have students' reason about spatial information using images (i.e., photos). The limitation of this approach is in choosing a finite set of response options based on intuition about the common errors.

COD questions have the potential for a more efficient approach to identifying *robust misconceptions*. They require students to respond to a spatially open-ended question, since they click directly on a diagram. For example, one of Lee and Schneider's (2015) questions asked students about the dark space on a picture of a spiral galaxy. Students were presented with a photo of the galaxy and asked, "what are the dark features at B?" (Lee and Schneider 2015, p.6). The correct multiple-choice option indicated the dark feature is a relatively cool region of the galaxy. With a COD question, students could be asked to click on an area that is relatively cool. This open-ended version requires students to reflect on what they know about the structure and spatial distribution of energy in the galaxy. Therefore, the COD question can be more conceptually revealing than a multiple-choice concept inventory question but equally efficient, instructionally. The CRS software constructs a heat map, where density of student clicks is represented with a range of colors, giving the instructor immediate feedback about the most common student errors (Fig. 1). Thus, COD questions reveal more than the number of incorrect responses, provide insights about *how* and *how much* students mental model may differ from the scientific consensus, and thus offer clues to improving students' understanding. Similarly, the open-ended nature of the COD questions allows researchers to circumvent the first, time-consuming step of collecting and coding written open-response answers. Researchers can selectively target students with specific, common incorrect answers for interviews or more fully developed explanations. This efficiency could facilitate rapid data collection to fill gaps in the research base on geoscience misconceptions (National Research Council 2012).

## Present Study

The purpose of this study is to test COD questions as a method for uncovering students' conceptions. The research questions focusing this study are as follows: (1) Are COD questions an effective technique for identifying students' conceptions? and (2) Can COD questions be used to identify novel, robust misconceptions? Our hypothesis is that the open-ended nature of students' responses to COD questions will be especially effective for identifying robust spatial misconceptions because they

will reveal *how* and *how much* students' answers differ from the scientific consensus. This rich information will reveal students' understanding of scalar relationships, spatial integration, and spatial reference frame. Four content topics were investigated (e.g., Earth's interior, base level, geologic time, and hot spot formation) within an introductory geology class to demonstrate that COD questions can detect false beliefs, flawed mental models, and robust conceptual errors that warrant additional research (Chi 2008). We apply a variety of analytical techniques that build on the basic distribution visualization provided by the software (i.e., heat map) to provide the reader with an understanding of the complexity of measuring spatial and diagrammatic reasoning and to offer suggestions for future work using COD diagrams in research.

## Methods

The study includes student data collected in two sections of an introductory geology course at a 4-year, public university in the US Midwest region. The undergraduate population is 57.0% white, 15.9% black, 15.4% Latino/Hispanic, and 5.0% Asian. The ACT score for incoming freshman is 22; however, over 40% of the undergraduates transfer from another institution. Introductory geology is a general education course primarily serving non-majors, although five geology majors enrolled in the course during the data collection in spring 2017. The one-credit laboratory associated with the course is not required. Less than half of the students included in this study were concurrently enrolled in the laboratory. The two course sections had a total enrollment of 32 and 39 students. Both sections were taught by the same tenure-track geology professor with over 10 years of experience implementing active learning pedagogies. The study protocol was approved by the Institutional Review Board and was determined exempt by the human-subjects institution review board under protocol #HS16-0409.

**Participants** Students enrolled in introductory geology gave consent to the investigator to gather data throughout the semester for this study. Students who did not give consent ( $n = 1$ ), withdrew from the course ( $n = 3$ ), or had incomplete data for the pre-instruction and end-of-course administration of the questions were removed ( $n = 28$ ) from the data set. There was a high rate of incomplete data due to late enrollments caused by the State of Illinois budgetary challenges that disproportionately impacted low-income students. Nevertheless, the pattern in the incomplete data was consistent with all conclusions presented below. Each COD item presented in this study has a different number of students included in the analysis due to variations in which individuals were in attendance on the day the post-instruction data were collected. The post-instruction questions were administered on the day the specific material was covered

in lecture to insure all data for the question was collected from students who had been present for the relevant instruction. We included data for students who completed all three administrations for a specific question, and report that number in the narrative that follows. Demographic information was not collected as a part of this study. Students enrolled in the course were predominantly freshman and sophomores seeking general education credit. Most students in the course were pursuing majors in early childhood education, elementary education, or communications.

**Implementation** Diagrams that address the key content goals of the course were selected from textbooks, google image searches, and the New York State Regents Exam (New York State Education Department [NYSED] 2018). The researchers developed questions that required students to identify the location of an object or predict the location of something based on a process. The diagrams were then re-drawn based on cognitive principles to maximize the salience of the target concept (Hegarty 2011) and eliminate unnecessary detail. Specifically, the diagrams were all drawn using simple lines, with minimal shading in grayscale and representing two dimensions. Superfluous patterns, structures, and process symbols were eliminated to support students’ attention to the essential information in the diagram. An example of a diagram and the redrawn version we used for the study are found in Fig. 2.

Data for this study was collected during classes that use Top Hat, a web-based CRS software, as a formative assessment tool. Students were required to participate in Top Hat throughout the course as an active learning strategy to improve their attention in class and retention of the course material. The instructor would lecture for 10–15 min, then use Top Hat to pose multiple-choice and COD questions on the material covered in lecture. After each question, the instructor displayed the answer and re-explained the content if students did not converge on the correct answer. Students earned class

participation points, which were 20% of their final course grade, for participating in the Top Hat questions throughout the semester. Students knew they received this credit regardless of whether their responses were correct or incorrect, and that they are essentially an attendance incentive. Over the course of the semester, students answered more than 150 questions using Top Hat, the majority of which were traditional multiple-choice questions. The present study examines responses from a selection of the 23 COD questions that were administered three times during the course: pre-instruction, immediately post-instruction, and at the end of the course. Students were pre-assessed on the 23 COD questions in the first 2 weeks of the course. Questions were administered a second time following instruction on the relevant course material. The timing of the post-instruction administration ranged within the semester from the first to 14th week of the semester. The third and final administration was at the end of the course when students were offered extra credit to answer the click-on-diagram questions outside of class time, prior to their final exam. The minimum delay between a post-instruction question and the end-of-course questions was 5 weeks.

**Data Analysis** Top Hat is presently the only CRS software that allows instructors to download the coordinates of each student click. The x-coordinate is the left-to-right position within the image and the y-coordinate is the top-to-bottom position within the image. The coordinates are provided as proportions relative to the top left corner (0,0), such that the top right corner is (1,0), the bottom left corner is (0,1) and bottom right corner is (1,1). CRS software like Top Hat use students’ smart devices (i.e., cell phones, tablets, and laptops) and therefore precision will be higher on devices with a larger screen and lower on those with a smaller screen. We tested the spatial precision of students’ responses ( $N= 35$ ) by asking them to click on a dot within a larger square diagram. Figure 3 is a plot of the student clicks. The mean  $X$  value is within 1% screen

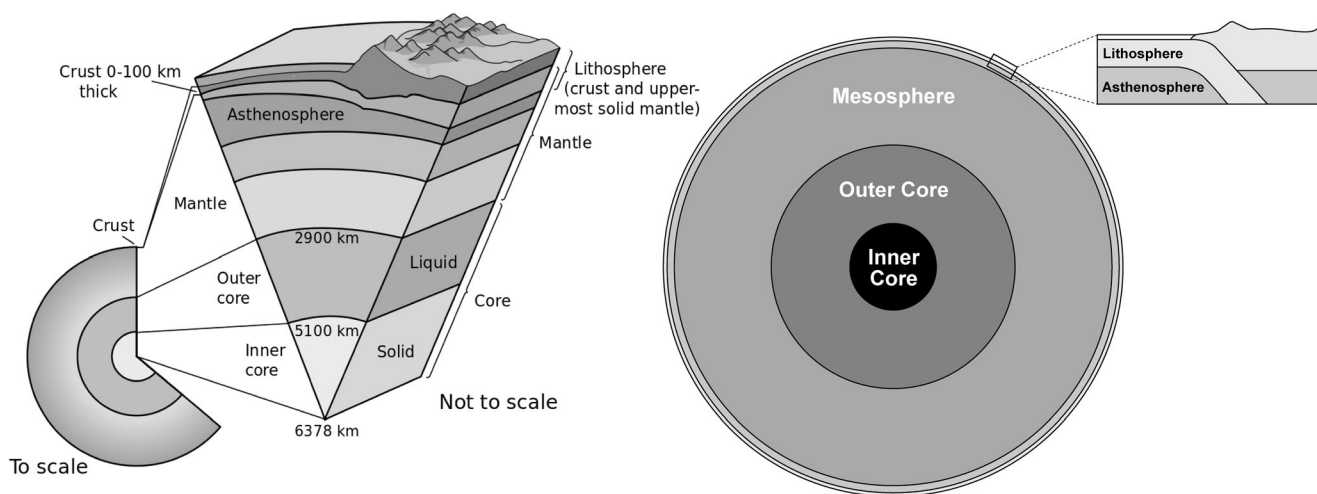
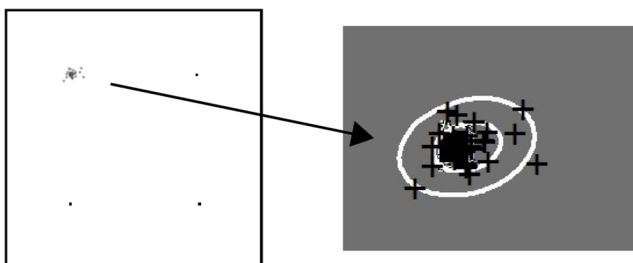


Fig. 2 Example diagrams of a typical diagram (left) (USGS 2017) and the redrawn diagram employed in the present study (right)

width ( $SD = 0.02$ ) of the true location and the mean of the  $Y$  value is within 0.1% of the screen height ( $SD = 0.01$ ) of the true location. To characterize the dispersion of points around the mean, we calculated the root mean square distance and evaluated how many students fell within one or two standard deviations of the mean. The standard deviation ellipses for one and two standard deviations are plotted along with the individual click locations in Fig. 3. Approximately 70% of students' clicks fell within the first standard deviation ellipse ( $n = 24$  of 34) and 97% of students' clicks were within two standard deviations ( $n = 33$  of 34), meaning the distribution was approximately normal with few outliers. There was a slight bias in the student data trending toward the center and top of the image. Students used a variety of devices, including Apple and android smart phones, tablets, and laptops. We did not collect information about each students' device, but a laboratory test of a Lenovo laptop ( $SD_x = 0.001$ ,  $SD_y = 0.002$ ), Surface Pro tablet ( $SD_x = 0.002$ ,  $SD_y = 0.005$ ), and iPhone 6 ( $SD_x = 0.003$ ,  $SD_y = 0.006$ ) demonstrated consistently high precision across devices.

The data for each questions were analyzed using statistical techniques that were appropriate for the type of data in each example (ex. chi-square,  $t$  test). Since the students' responses were distributed in a two-dimensional space, we plotted and analyzed data using ArcGIS. The ArcGIS tools allowed us to quickly count which students clicked in correct versus incorrect regions of a diagram using polygons and plot lines that demonstrate shifts in the locations of answers from pre- to post-instruction.

**Categorizing Conceptions** The questions focus on topics for which students have well-documented misconceptions (ex. Earth's interior, geologic time) and novel conceptual challenges observed in the classroom (ex. hot spot formation, erosion to base level). We selected COD questions that are representative of the different aspects of Chi's (2008) conceptual change framework. Chi's (2008) three categories of conceptions are (1) false beliefs, (2) flawed mental models, and (3) robust misconceptions. For our purposes, *false beliefs* were identified as cases where a change was evident immediately following instruction that remained stable at the end of the course. We grouped the latter two categories due to the small



**Fig. 3** Standard deviation ellipses for one and two standard deviations from the mean of students' clicks on an in-class calibration test ( $N = 34$ )

sample size of this study, though we observed systematic and persistent errors from a subset of students on COD questions categorized into the *misconceptions* category. The narrative below describes four examples of COD question that are categorized using Chi's framework.

## Results

The goal of this study was to test the value of COD questions as a tool for studying students' geoscience conceptions. For each of the four examples below, we review what, if any, relevant misconceptions were already documented in the literature, present results from the COD questions administered during this study, review how the data inform classification of the misconceptions according to Chi's (2008) framework, and interpret which spatial category the conception represents (Table 1). The four cases are ordered by the pattern of change over the course of the semester, ranging from false belief to robust misconception.

### False Belief

#### Earth's Interior

Libarkin et al. (2005) identified several common erroneous beliefs about the Earth's interior in a study of undergraduate non-science majors. Students are able to identify concentric layers in the Earth's interior but know little detail about the layers. Students struggle to explain why Earth is made of concentric spheres and which layers are solid or liquid. In a study of middle school students, Gobert (2000) found that students identified the center of the Earth as a source of magma for volcanic eruptions. Students were confused about how and where melting occurs within the Earth (Clark et al. 2011). Many textbooks and media represent the mantle as "molten," or liquid (King 2010), despite research using seismic tomography indicating that the mantle is solid with relatively warmer and cooler regions where convection as a solid occurs.

In response to the literature on student conceptions of the Earth's interior, we asked students to click on the layer within the Earth that is entirely liquid. Pre-instruction, most students incorrectly selected the inner core as being liquid (Fig. 4a). Following lecture slides on the process of differentiation and the physical characteristics of Earth's layers, most students adjusted their response to the outer core (Fig. 4b), correctly identifying it as liquid. Students appear to retain this concept because there was little reversion to the pre-conception at the end of the course; the majority of the students at the end identified the outer core as the layer which is entirely liquid (Fig. 4c).

To analyze students' responses, the location of their clicks was plotted in ArcGIS. We were most interested in which students clicked within the outer core layer versus the rest of

a. Pre-instruction

b. Post-instruction

c. End of course

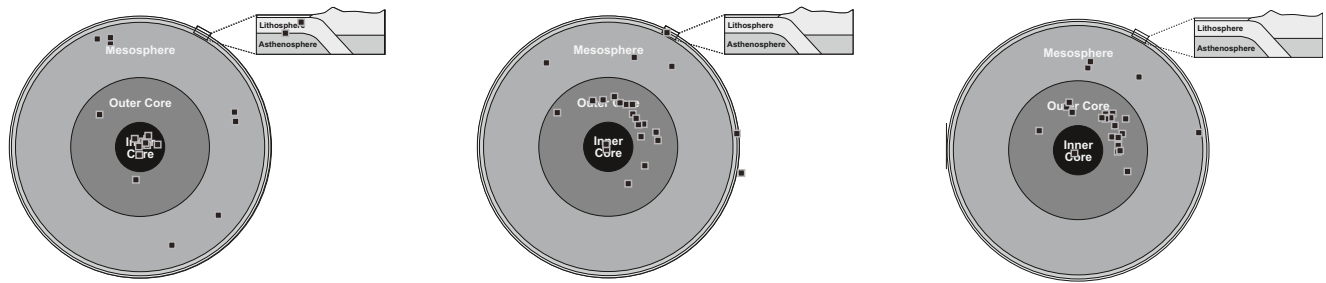


Fig. 4 Student responses to “click on the layer that is entirely liquid” for a pre-instruction, b post-instruction, and c at the end of the course (N=25)

the diagram, so we created a polygon tracing the space that contained the outer core within ArcGIS. Using the “select by location” function, we identified which clicks were within the correct layer versus all other places on the image. Overall, there was clear and relatively stable improvement from pre-instruction (8% correct), to post-instruction (60% correct), and end of the course (76% correct). Counts of correct and incorrect clicks in each condition were analyzed using 2 × 2 contingency tables by calculating Pearson’s chi-square. There was a significant improvement between students’ responses pre-instruction and post-instruction ( $\chi^2(1, N = 25) = 33.59, p < .001, \phi = .82$ ) and post-instruction and end of the course ( $\chi^2(1, N = 25) = 29.31, p < .001, \phi = .77$ ). A Cramer’s V value over 0.5 is considered a large effect (Cohen 1988). This demonstrates that substantially more students shifted to the correct answer in the time between post-instruction and the end of the course. This concept appeared on the first exam and the final exam, and therefore, we suggest that this was due to repetition of this concept in the course material.

The data indicate that students maintain their improved scientific conception from post-instruction to the end of the course. Therefore, we classified this conceptual change as a *false belief* (Chi 2008). This question asked students to identify a structure and is a simple recall question. Consequently, the question did not require much spatial reasoning and we did not categorize this into a spatial ontological category from Table 1. This example demonstrates that COD questions can

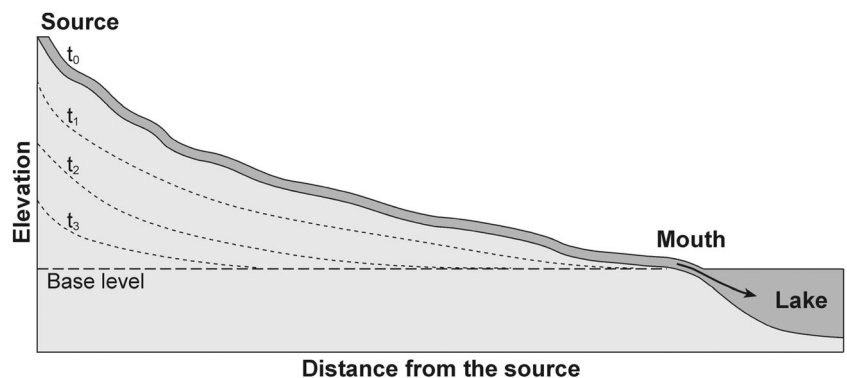
confirm misconceptions that are documented in the literature (Libarkin et al. 2005).

**Misconceptions: Flawed Mental Model or Robust Misconception**

**Base Level**

In geology, the term base level means “the lowest elevation a stream channel’s floor can reach at a given locality” (Marshak 2016, p.472). Base level may also be thought of as a “four-dimensional surface of equilibrium between erosion and deposition” (Catuneanu 2006, p. 84). Above base level, sediments will be eroded and below base level, sediments will be deposited. Simply put, a surface will be subject to erosion as long as it is above the elevation of the water body nearest to it. The profile (or side view) of a river from the headwaters upstream to the mouth, where the river empties into a body of water, typically shows a graded profile (i.e., concave and smooth) that erodes over time (Fig. 5). Because most rivers start as small tributaries with little discharge and end as large rivers with a high discharge, the rivers typically erode to base level more quickly toward the mouth. Once the river is eroded to base level, it does not continue eroding unless other geologic processes occur (i.e., regional tectonic uplift lowers relative sea level). Herrera and Riggs (2013) studied students’ conceptions of sedimentary systems. Of the concepts they

Fig. 5 A graded stream profile has the highest elevation and gradient toward the source of the stream and lowest elevation and gradient toward the mouth of the stream





studied, base level was the most difficult for students to understand (Herrera and Riggs 2013). In addition to requiring complex spatial and temporal reasoning, the concepts of base level and river profiles are typically only superficially covered in introductory courses.

To test the concept of base level, students were asked to predict the vertical elevation to which the river will erode, or downcut, at A and C (Fig. 5). During class, students observed the instructor sketching what the river profile would look like over time. The river should erode at a faster rate at A, near the headwaters, than at C, near the mouth of the river. This pattern is seen because the elevation of C is already at or near base level. Students with a scientific conception of base level would select locations for A and C that produce a flatter river profile, because as time progresses, location A erodes more than location C.

To evaluate whether students' conceptions of erosion to base level improved, we tested whether students' clicks for C were significantly different from base level (i.e., the level of the lake) and whether the elevation of points A and C significantly changed (Fig. 6). A single-sample  $t$  test revealed that students' elevations for C ( $M = 0.78$ ,  $SD = 0.05$ ) were significantly different from base-level elevation (0.70) pre-instruction ( $t_{\text{pre}}(24) = 7.91$ ,  $p < 0.0001$ ,  $d = 1.6$ ) and post-instruction ( $t_{\text{post}}(24) = 2.38$ ,  $p = 0.03$ ,  $d = 0.43$ ), but not at the end of the course ( $t_{\text{end}}(24) = 1.12$ ,  $p = 0.27$ ). A Cohen's  $d$  effect size of 0.5 is commonly regarded as a medium effect size while a 0.8 is a large effect size (Cohen 1988), indicating that there is a substantial difference between students' responses and the correct elevation of C pre-instruction and a moderate difference post-instruction. The insignificant  $t$  test at the end of the course demonstrates that students' clicks were correctly located at or near base level and that the river stopped eroding at the level of the lake (i.e., base level).

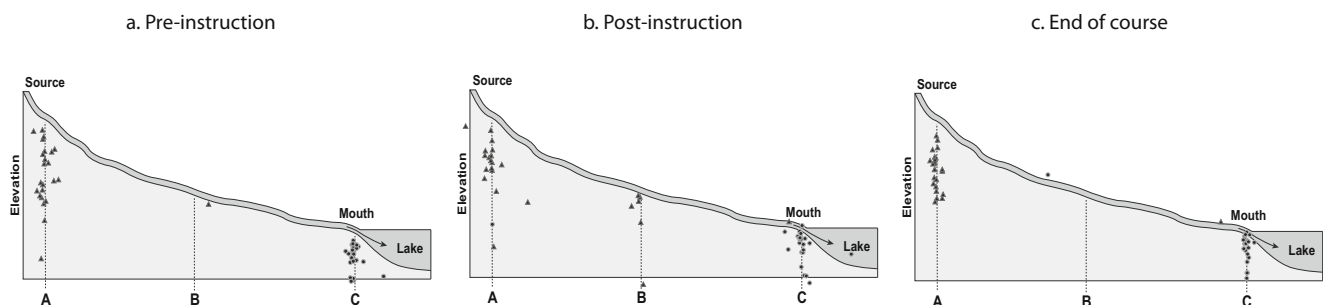
When we tested whether the slope of the line connecting students' clicks for points A and C became flatter (i.e., lower slope), paired-sample  $t$  tests indicated that the slope of the line connecting students' clicks did not change significantly from pre- to post-instruction ( $t_{\text{pre-post}}(24) = 1.43$ ,  $p = 0.17$ ), from post-instruction to end of the course ( $t_{\text{post-end}}(24) = 0.73$ ,  $p = 0.47$ ), or from pre-instruction to end of the course ( $t_{\text{pre-}}$

$t_{\text{end}}(24) = 0.91$ ,  $p = 0.37$ ). This indicates that students did not fully grasp that the slope of the river profile would become flatter with time, because C was already at base level.

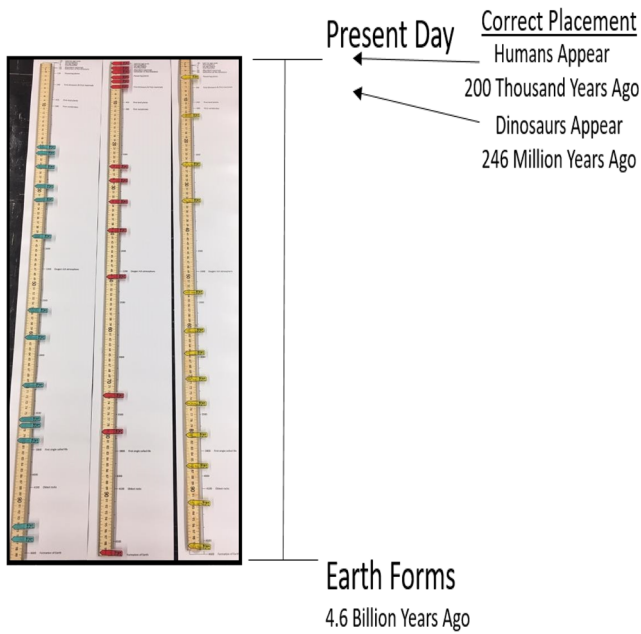
The COD data show some improvement in one aspect of students' mental model of base level, erosion at C, while other aspects of erosion are not correctly represented (i.e., decreasing slope of the river profile). This incremental improvement in students' answers may suggest this is an example of a *flawed mental model*. Understanding base level requires the integration of several spatial concepts (Table 1). Students must recognize that there are both vertical and lateral patterns of erosion. Vertically, there will be a faster rate of erosion at A than at C. Laterally, there will be decreasing slope over time as a consequence of the vertical pattern of erosion. Students may initially understand erosion as a local removal of sediment without understanding the system, with its spatial variation in erosion due to the variation in water flow. The integration may require a combination of spatial processes described by Newcombe and Shipley (2015): spatial visualization of multiple parts and how they change over time. Spatial visualization involves putting together spatial relations across multiple locations to, in this case, integrate the vertical and lateral patterns, and how relationships among the parts change over time. We propose this is an ontological shift that combines content knowledge (about how the process of water flowing varies and differentially effects different locations) and spatial thinking skills (Table 1). Testing this conception with an in-depth conceptual change, classroom intervention and documenting any systematic errors in students' answers would clarify whether this should be categorized as a *flawed mental model* or *robust misconception*.

## Geologic Time

Geologic time is a challenging conceptual hurdle for students because the magnitude of geologic events occurred at a scale beyond typical human experience (Cheek et al. 2017; Jones et al. 2009; Tretter et al. 2006); therefore, students often struggle with the relative spacing of geologic events (Libarkin et al. 2007). Understanding deep time in geology also impacts students' understanding of other scientific phenomena, such as



**Fig. 6** Student responses to “click along line A (or C) where you expect to find the elevation of the river after it has eroded for 10 years” for **a** pre-instruction, **b** post-instruction, and **c** at the end of the course ( $N = 25$ )



**Fig. 7** Photo of three students’ geologic timelines, with flags indicating their placement of events, and the correct placement for the COD question on geologic time

climate change (Lombardi and Sinatra 2012). There are several factors contributing to the misunderstanding of geologic time, including the non-linear nature of the geologic time scale as it is depicted in many textbooks (Resnick et al. 2017). Research on temporal reasoning indicates a cognitive relationship between space and time both in language (Lakoff and Johnson 2008), generally, and with respect to biology (Catley and Novick 2009) and geologic time (Dodick and Orion 2003; Trend 2001), specifically. To combat this conceptual challenge, geologists have implemented classroom activities (Richardson 2005) and informal education opportunities, such as the Trail of Time Exhibit at the Grand Canyon (Karlstrom et al. 2008), to build students’ understanding of the distribution of geologic events. These activities require the analogical mapping of time onto a spatial representation

(i.e., a rope or location) (Gentner 1989). Chi (2008) asserts that adjusting a *robust misconception* requires a shift between ontological categories. We chose an intervention where students relate space and time to improve students’ mental models. This is the type of intervention that could help counteract a *robust misconception* because students are confronted with a physical, scalar model of geologic time (Cheek et al. 2017). Here, the ontological adjustment is from a human-centric time frame to the vastness of geologic time using a spatially represented model.

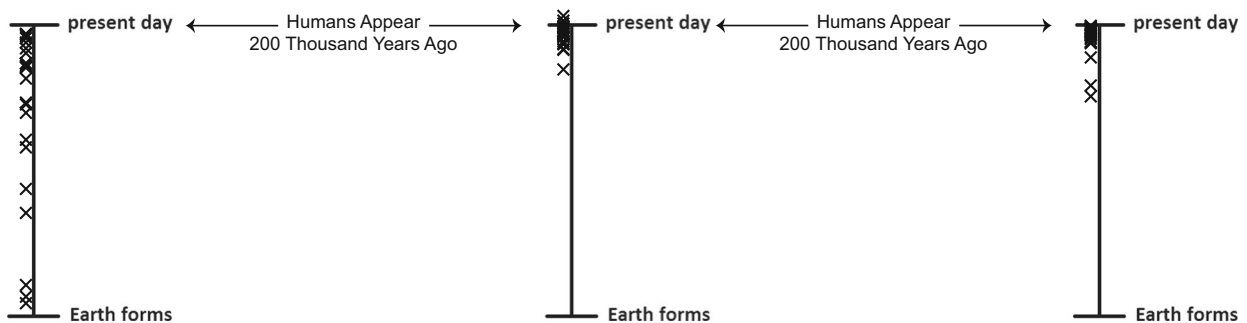
Prior to instruction, students were asked to click on a geologic timeline where they thought humans and dinosaurs appear (Fig. 7). Next, we implemented an activity requiring pairs of students to predict the order and spacing of a set of geologic events on a meter stick. After students committed to their prediction, they received a laminated strip of poster paper with the geologically accurate scaled model of the geologic event. Students attached the scientific model to their meter stick and were asked to compare their model to the scientific model (Fig. 7). We did not analyze students’ models from the class activity. Anecdotally, the instructor reported observing that students’ errors were consistent with the extant literature (e.g., overestimation of the time since dinosaurs appeared) (Libarkin et al. 2007; Resnick et al. 2017).

Click locations on the time line were based on the y-coordinate alone because the timeline is a vertical line. A paired-sample *t* test shows a significant improvement from pre-instruction ( $M = 0.39, SD = 0.24$ ) to both post-instruction ( $M = 0.17, SD = 0.03$ ) and end of the course ( $M = 0.19, SD = 0.05$ ) responses for the appearance of humans on the geologic time scale ( $t_{pre-post}(23) = 4.27, p < 0.001, d = 1.25$ ;  $t_{pre-end}(23) = 4.10, p < 0.001, d = 1.16$ ) (Fig. 8). A Cohen’s *d* of 1.0 represents a one standard deviation improvement in students’ answers, indicating that there was a substantial and stable change after instruction. There was also a significant improvement from pre-instruction ( $M = 0.54, SD = 0.24$ ) to both post-instruction ( $M = 0.29, SD = 0.13$ ) and end of the course ( $M = 0.31, SD = 0.15$ ) responses for the appearance of dinosaurs on

**a. Pre-instruction**

**b. Post-instruction**

**c. End of course**



**Fig. 8** Student responses to “If the line represents the entire timespan since Earth formed, click where you expect humans appeared on Earth” for **a** pre-instruction, **b** post-instruction, and **c** end of the course show improvement in students’ mental model ( $N = 25$ )

the geologic time scale ( $t_{\text{pre-post}}(23) = 4.69, p < 0.001, d = 1.32$ ,  $t_{\text{pre-end}}(23) = 4.76, p < 0.001, d = 1.15$ ) (Fig. 9). The large Cohen's  $d$  indicates a substantial and stable change after instruction. Students answers did not significantly change from post-instruction to the end of the course ( $t_{\text{human}}(23) = 1.13, p = 0.135$ ;  $t_{\text{dino}}(23) = 0.77, p = 0.224$ ), suggesting the change was stable over the length of the course. Students responses post-instruction converged on a more accurate and precise answer, with a standard deviation that was only 20% of the standard deviation of pre-instruction responses. Despite the students' clicks being significantly closer to the correct response for humans (the scientific consensus location is 0.00005) and dinosaurs (the scientific consensus location is 0.05), the students' responses overestimated the age of the appearance by several orders of magnitude at the end of the course.

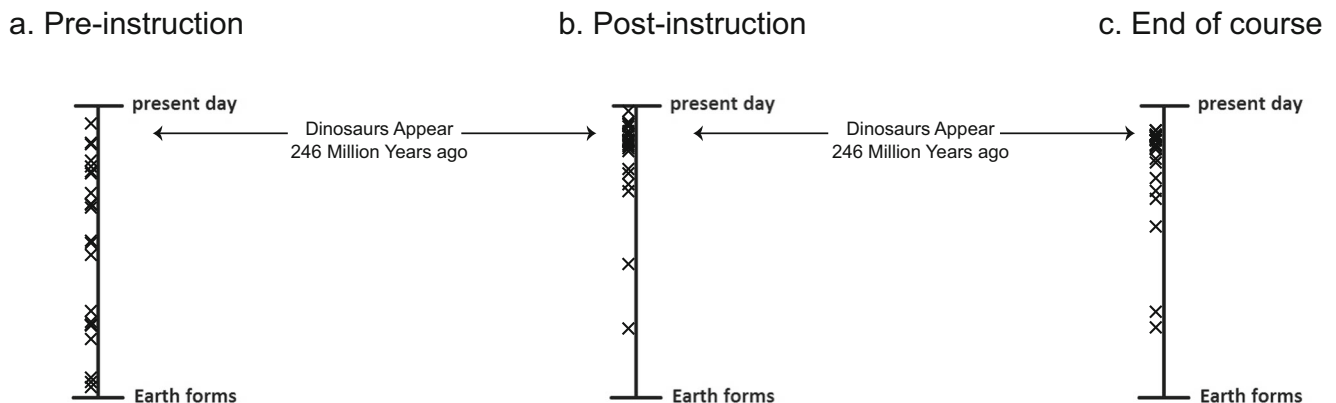
Geologic time is an example of a *robust misconception* because persistent systematic errors in students' conceptions of deep time are well documented (Dodick and Orion 2003; Libarkin et al. 2007; Trend 2001). A conceptual change intervention (LaDue 2018), in which students generate their own model of geologic time and compare that to a spatially accurate model, led to substantial improvement in their mental model of the appearance of humans and dinosaurs since the Earth was formed. Anecdotally, several students made verbal comments of surprise at the scientific spacing of the geologic events. Students experienced a shift from human-centric to geologic temporal scale, indicative of a scalar relationship ontological category shift (Table 1).

### Hot Spot Formation

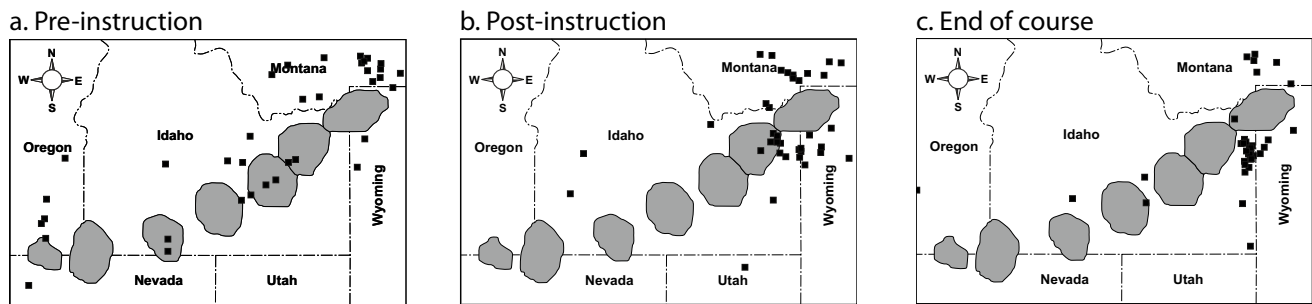
Several aspects of volcanic processes have been included in past research on conceptions (Parham et al. 2010), including the mechanisms involved in melting within the asthenosphere (Clark et al. 2011). We were unable to identify any prior studies on students' understanding of spatial reference frame with respect to hot spots and tectonic plate motion. As described in the introduction, a geologic hot spot results from a rising

plume of hot mantle rock that melts near the surface (Marshak 2016). The lithospheric tectonic plate moves over the hot spot causing a chain of successively younger volcanic calderas to form. Most people are familiar with the Hawaiian volcanic island chain. Yellowstone National Park is another, if less well-known, example of a currently active hot spot, with a chain of extinct volcanic calderas located in the Snake River Plain of Idaho. The concept of hot spots is spatially complex because from the human-centric perspective of standing on the surface, hot spots move across the fixed landscape. From a geologic reference frame, the landscape (i.e., lithospheric plate) is moving relative to the hot spot, which has a fixed location in the mantle below. In fact, the progressive age of volcanic calderas at hot spots was the original evidence used to deduce the rate and direction of past tectonic plate motion (Jarrard and Clague 1977). Similar to the ontological shift from a human to a geologic temporal scale, understanding hot spot formation represents a shift from a human to geologic spatial frame of reference. Therefore, we expect to find a *robust misconception* (Chi 2008) about hot spot formation.

We developed a COD question to evaluate whether students are able to predict where the next hot spot will form if they are given the direction of plate motion. The COD question indicated that X is the currently active volcanic caldera and that the tectonic plate had moved southwest over the hot spot (Fig. 10). Students were asked to click where the next caldera would form if the plate started moving north. The correct location of the next caldera is to the south of the caldera labeled with the X. Pre-instruction, many students made the incorrect prediction that the next hot spot will form north of the currently active hot spot. Other students selected points generally north of various points on the hot spot track. During class, students were asked to discuss their predictions with their group partners. Subsequently, all students viewed a video that included map-view diagrams of hot spots and a side view animation about how hot spots form (Incorporated Research Institutions for Seismology [IRIS], no date). Following the video, students were asked again to discuss



**Fig. 9** Student responses to “If the line represents the entire timespan since Earth formed, click where you expect dinosaurs appeared on Earth” for **a** pre-instruction, **b** post-instruction, and **c** end of the course show improvement in students' mental model ( $N = 25$ )



**Fig. 10** Student responses to “X represents the currently active volcanic caldera. The tectonic plate has moved southwest over the hot spot. If the plate started moving north, click where you expect the next caldera will form.” for **a** pre-instruction, **b** post-instruction, and **c** end of the course ( $N = 35$ )

with their partners where the next volcano would be likely to form. Results of the post-instruction student responses converged on one of two possible locations, north or south of the most recent hot spot (Fig. 10b). At the end of the course, the misconception clearly had persisted (Fig. 10c).

Students’ responses were plotted in ArcGIS and we generated a polygon of the region containing the correct answer (i.e., south of the current hot spot X). Using the “select by location” feature, ArcGIS generated a list of students with correct and incorrect answers. There was a significant difference between students’ answers pre-instruction to post-instruction ( $\chi^2(1, n = 35) = 19.66, p < 0.00001, \phi = 0.53$ ) and there was not a significant change from post-instruction to end of the course ( $\chi^2(1, N = 35) = 0.952, p < 0.329$ ). This suggests that there was a substantial shift from pre- to post-instruction but that changes between post-instruction and the end-of-course responses were modest. Nevertheless, we observed a pattern of switching between North (the incorrect answers) and South (the correct answer) tended toward correction and warrant further investigation. From post-instruction to end of the course, seven students switched from incorrect (North) to correct (South), while three students persisted in the incorrect answer (North) and another four students reverted from the correct (South) to the incorrect (North) answer.

One line of evidence that distinguishes a *robust misconception* from a *flawed mental models* is its persistence in the presence of conflicting information. After discovering the cluster of incorrect responses north of the hot spot post-instruction, the instructor engaged the students in modeling the plate motion using a piece of paper with a pencil held vertically beneath the paper (i.e., gesturing the relative plate motion). When students moved the paper north, modeling the plate moving north, some responded with immediate surprise that the pencil, modeling the mantle plume, was south of the most recent hot spot. This was a post hoc, informal attempt at physically modeling plate motion, which may be a profitable component of a comprehensive activity to address conceptual understanding of hot spot formation. Despite multiple encounters with the content (i.e., worksheet, animation, physical modeling), 20% of the students ( $n = 7$  of 35) still chose

an incorrect location (North) for the next hot spot at the end of the course. We propose this persistent error is a result of the need to change spatial reference frames (Table 1). The human-centric reference frame may be hard to abandon because our experience is that the lithosphere, on which we live, is not moving. In the Earth-centric reference frame, the mantle plume is stationary and the lithospheric tectonic plate is moving relative to the mantle plume. Despite several rounds of refutation, the persistence of a systematic incorrect answer is consistent with the definition of a *robust misconception* (Chi 2008).

## Discussion

The foci of this study were to test the value of COD questions as a tool for identifying students’ conceptions and to probe for as yet undiscovered robust misconceptions in geology. We demonstrated that COD questions can elicit and confirm known misconceptions related to the Earth’s interior, base level, and geologic time. We uncovered a previously unreported misconception using a COD question: conceptual difficulty predicting the location of a new hot spot. Conceptual change research typically utilizes multiple-choice text and image questions, open-ended text questions, drawing tasks, and interviews to reveal students’ concepts. COD questions supplement the existing toolbox in a unique way. COD questions allow large numbers of students to provide spatial information about their mental models in the form of an open-ended spatial response. A critical next step is to confirm the spatial ontological categories we propose with semi-structured interviews about the COD conceptions tested in this study. This approach would clarify to what extent the spatial information afforded by CODs relates to various types of mental models and, consequently, misconceptions.

The unique benefit of COD questions is the ability to gather conceptual challenges associated with spatial information with large groups of students, immediately. Although COD questions are useful for common false beliefs, like the Earth’s interior, students’ conceptual challenges appear to be more robust and persistent for questions involving spatial

processes. Three spatial categories we tested were spatial integration, scalar relationships, and spatial reference frame. The base level example demonstrated how students' integrated patterns of erosion in a river to understand how base level changes over time. Integration of pieces of information into a whole structure is one of the critical spatial skills for geologic thinking identified in Shipley et al. (2013). The geologic time questions demonstrated that students improved their sense of scalar relationships by shifting from human-centric to a deep time framework. This is a well-studied conceptual challenge and our study demonstrates that it is easily diagnosed using a COD question (Cheek et al. 2017; Tretter et al. 2006; Jones et al. 2008; Libarkin et al. 2007). The hot spot example supports spatial reference frame as a necessary category. Students must reconcile the human-centric model of a stationary lithosphere with the geo-centric model of a stationary mantle beneath the moving lithosphere. Spatial reference frame may be particularly challenging (and perhaps the most complex of those considered here), involving spatial skills such as mental animation (Newcombe and Shipley 2015), perspective-taking (Hegarty and Waller 2004), and/or retrodiction (Ault 1998; Trend 2000). COD questions offer an elegant research tool to identify, analyze, and respond to students' spatial conceptions, particularly for a spatially rich domain like geology.

The sampling of statistical and analytical techniques presented here illustrates the range of analysis options for COD data. For questions where slope of a line connecting two points is critical to the conception, the data may be plotted in a graphical software so that the slopes can be compared pre- and post-instruction within one program. Means and standard deviation of points on a line are easily calculated using Excel. The most powerful software for analyzing COD data is ArcGIS because the software is designed for spatial statistics. Polygon analysis combined with the chi-square statistic is useful for comparing the location of students' conceptions pre- and post-instruction when a region of the diagram contains the correct answer rather than a single point. The potential for spatial statistics using ArcGIS was not fully exploited in this study, and opportunities for other types of analyses should be explored further. ArcGIS has additional spatial analysis tools, such as cluster analysis and regression, that could profitably be applied to future studies that involve COD data.

The examples presented above demonstrate that COD questions are a useful tool in service to a program of research investigating students spatial and temporal reasoning. This is an opportunity to use smart technology to fill a gap in the literature (Cheek 2010; Herrera and Riggs 2013; NRC 2012). The technique couples geologic reasoning and diagrams, both of which can involve challenging spatial concepts. COD offer a way to improve students' mental model and thus support learning in the classroom. Administering COD question in a classroom requires students to commit to a prediction, by clicking on a diagram. The instructor provides feedback on the spatial location of the correct answer by showing the resulting heat map generated by the CRS

software and noting the correct location or regions. Recent work has tested prediction and feedback as a pedagogy to promote spatial and temporal reasoning in geology. Gagnier et al. (2017) found that asking students to make predictions of the interior of a geologic structure led to significant improvement on a geologic penetrative-thinking test, compared with visualization alone or copying the sketch of a structure. The authors suggest that sketching facilitates spatial reasoning about three-dimensional structures because students are asked to commit to their mental models through externalization as a sketch. Engaging students in predictions using COD questions, which are also open-ended and spatial responses, could have a similar effect. Resnick et al. (2017) compared multiple-choice CRS questions about the geologic time scale to a hands-on activity with a physical model of the geologic time scale on a meter stick. In both cases, students made predictions and received feedback about their errors; however, the amount of class time required for the clickers was much less than the hands-on activity. The authors found that both conditions promoted accurate representations of the linear spacing of large-scale geologic events, indicating that CRS questions may be a sufficient mechanism for spatial feedback to improve students' mental models. Although Resnick et al. (2017) used multiple-choice response options, the alternative choices were placed at points on the geologic time scale based on previous work on likely errors. The open-ended nature of COD questions eliminates the need for previous research or practitioners' intuition in choosing response options. The possibility of spatial prediction and feedback as a pedagogical intervention to generally improve spatial and temporal reasoning (Uttal et al. 2013) warrants further study.

In addition to being a useful tool for researchers and a means for providing individual feedback, COD questions offer significant benefits for instructors. COD questions can reveal students' conceptual understanding in real time to allow instructors to adapt their teaching. This formative assessment technique facilitates students' learning more efficiently than passively listening to a lecture. The principles of CRS-based pedagogy outlined in the TEFA framework (Beatty and Gerace 2009) have already been tested with astronomy photographs by Lee and Feldman (2015) with success. COD questions have the added benefit of an open-ended, spatial response, which may be more engaging than multiple-choice questions, and reduce the demand on instructor to generate alternative answers. Spatial prediction and feedback using COD questions extend the usefulness of the TEFA framework to spatially challenging concepts and diagrams. This is fertile ground for future research on students' conceptual models of spatially and temporally complex concepts.

## Limitations and Future Research

The present study is limited by several issues, including the proprietary online platform where x-y data may be gathered

and the need for a mixed-methods study to evaluate the convergence of COD data with interviews and drawings of conceptual understanding. There are currently several companies that offer COD-style question formats but only Top Hat provides the x-y coordinate data. For the researcher, Qualtrics, an online survey platform, does offer COD questions and x-y data capture. This may be a useful alternative for research purposes, but does not provide the resources of a CRS software platform for instructional use.

A limitation of the study is the possibility of test-retest effects resulting from administering the same question three times throughout a 15-week course (for more detail, see “Implementation” section). The authors acknowledge this potential confound to the effect size, but prioritized good classroom pedagogy for learning over the goal of the research study. Furthermore, the test-retest effect would also underestimate the robustness of the misconceptions we identified in the study.

The utility of COD questions to document misconceptions and conceptual change will require additional work with traditional methods of assessing conceptual change. For this initial work, we did not interview students or ask them to draw. A mixed-methods approach that purposefully samples a cross-section of students who respond to COD questions with the most common errors would facilitate deeper understanding and any potential limitations of this technique as a research method. In the case of Gagnier et al.’s (2017) study, the act of sketching was a robust externalization of a students’ mental model. A useful next step is to test sketching versus clicking on an open-ended COD question to evaluate how robust the externalization must be, or whether, in the case of Resnick et al. (2017), spatial prediction and feedback through clicker response can be sufficient to promote an improved, scientific model for a variety of changes to mental models.

## Conclusion

Diagrams are an important tool for communicating spatial and temporal models of Earth processes. Current methods to study students’ conceptual understanding are laborious for the researcher and prohibitive for the classroom instructor. COD questions offer a tool to supplement the existing methods available to researchers. COD questions enable the researcher to target conceptions associated with spatially complex structures and processes more efficiently than is possible using traditional multiple-choice questions and verbal reports. Further, COD questions are a promising pedagogical tool for implementing a TEFA framework to improve conceptual understanding of spatial structures and processes in the class. Future research should exploit COD questions to understand the mechanism of spatial feedback in building scientific models of geoscience processes.

**Acknowledgements** The authors appreciate assistance from Kerri Gefeke and Sheldon Turner for assisting with the ArcGIS analysis protocol, Mark Howland for re-drafting the graphics used for the CRS questions, Allison Jaeger for input on the COD questions, Mia Velazquez for generating the figures, and Doug Lombardi for thoughtful suggestions on this manuscript.

**Funding** This study was funded in part by the National Science Foundation (Grant 1640800). Preparation of this manuscript was supported in part by NIU’s PI Academy and National Science Foundation grant 1640800 to TFS & NDL.

## Compliance with Ethical Standards

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical Approval** All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

**Informed Consent** Informed consent was obtained from all individual participants included in the study.

## References

- Atit, K., Shipley, T. F., & Tikoff, B. (2013). Twisting space: are rigid and non-rigid mental transformations separate spatial skills? *Cognitive Processing, 14*(2), 163–173.
- Ault, C. R. (1998). Criteria of excellence for geological inquiry: the necessity of ambiguity. *Journal of Research in Science Teaching, 35*(2), 189–212.
- Beatty, I. D., & Gerace, W. J. (2009). Technology-enhanced formative assessment: a research-based pedagogy for teaching science with classroom response technology. *Journal of Science Education and Technology, 18*(2), 146–162.
- Brewer, W. F. (2008). Naïve theories of observational astronomy: review, analysis, and theoretical implications. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 155–204). New York: Routledge.
- Carlson, L. A. (1999). Selecting a reference frame. *Spatial Cognition and Computation, 1*(4), 365–379.
- Catley, K. M., & Novick, L. R. (2009). Digging deep: exploring college students’ knowledge of macroevolutionary time. *Journal of Research in Science Teaching, 46*(3), 311–332.
- Catuneanu, O. (2006). *Principles of sequence stratigraphy*. Amsterdam: Elsevier.
- Cheek, K. A. (2010). Commentary: A summary and analysis of twenty-seven years of geoscience conceptions research. *Journal of Geoscience Education, 58*(3), 122–134.
- Cheek, K. A. (2013). Exploring the relationship between students’ understanding of conventional time and deep (geologic) time. *International Journal of Science Education, 35*(11), 1925–1945.
- Cheek, K. A., LaDue, N. D., & Shipley, T. F. (2017). Learning about spatial and temporal scale: current research, psychological processes, and classroom implications. *Journal of Geoscience Education, 65*(4), 455–472.
- Chi, M. T. H. (1997). Creativity: shifting across ontological categories flexibly. In T. B. Ward, S. M. Smith, & J. Vaid (Eds.), *Creative thought: an investigation of conceptual structures and processes*

- (pp. 209–234). Washington, DC: American Psychological Association. <https://doi.org/10.1037/10227-009>.
- Chi, M. T. H. (2008). Three types of conceptual change: belief revision, mental model transformation, and categorical shift. In S. Vosniadou (Eds.), *International handbook of research on conceptual change* (pp. 61–82). New York: Routledge.
- Clark, S. K., & Libarkin, J. C. (2011). Designing a mixed-methods research instrument and scoring rubric to investigate individuals' conceptions of plate tectonics. *Geological Society of America Special Papers*, 474, 81–96.
- Clark, S. K., Libarkin, J. C., Kortz, K. M., & Jordan, S. C. (2011). Alternative conceptions of plate tectonics held by nonscience undergraduates. *Journal of Geoscience Education*, 59(4), 251–262.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd). New York: Routledge Academic.
- Dodick, J., & Orion, N. (2003). Cognitive factors affecting student understanding of geologic time. *Journal of Research in Science Teaching*, 40(4), 415–442.
- Dolphin, G., & Benoit, W. (2016). Students' mental model development during historically contextualized inquiry: how the 'tectonic plate' metaphor impeded the process. *International Journal of Science Education*, 38(2), 276–297.
- Dove, J. E. (1998). Students' alternative conceptions in Earth science: a review of research and implications for teaching and learning. *Research Papers in Education*, 13(2), 183–201.
- Emenike, M. E., & Holme, T. A. (2012). Classroom response systems have not "crossed the chasm": estimating numbers of chemistry faculty who use clickers. *Journal of Chemical Education*, 89(4), 465–469.
- Forbus, K. D., Ferguson, R. W., Lovett, A., & Gentner, D. (2017). Extending SME to handle large-scale cognitive modeling. *Cognitive Science*, 41(5), 1152–1201.
- Francek, M. (2013). A compilation and review of over 500 geoscience misconceptions. *International Journal of Science Education*, 35(1), 31–64.
- Gagnier, K. M., Atit, K., Ormand, C. J., & Shipley, T. F. (2017). Comprehending 3D diagrams: sketching to support spatial reasoning. *Topics in Cognitive Science*, 9(4), 883–901.
- Garvin-Doxas, K., & Klymkowsky, M. W. (2008). Understanding randomness and its impact on student learning: lessons learned from building the Biology Concept Inventory (BCI). *CBE Life Sciences Education*, 7(2), 227–233.
- Gentner, D. (1989). Analogical learning. In S. Vosniadou & A. Ortony (Eds.), *Similarity and analogical reasoning* (p. 199). Cambridge: Cambridge University Press.
- Gobert, J. D. (2000). A typology of causal models for plate tectonics: inferential power and barriers to understanding. *International Journal of Science Education*, 22(9), 937–977.
- Gurel, D. K., Eryılmaz, A., & McDermott, L. C. (2015). A review and comparison of diagnostic instruments to identify students' misconceptions in science. *Eurasia Journal of Mathematics, Science and Technology Education*, 11(5), 989–1008.
- Hegarty, M. (2011). The cognitive science of visual-spatial displays: implications for design. *Topics in Cognitive Science*, 3(3), 446–474.
- Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32(2), 175–191.
- Hegarty, M., Crookes, R. D., Dara-Abrams, D., & Shipley, T. F. (2010). Do all science disciplines rely on spatial abilities? Preliminary evidence from self-report questionnaires. In International Conference on Spatial Cognition (pp. 85–94). Berlin, Heidelberg: Springer.
- Herrera, J. S., & Riggs, E. M. (2013). Identifying students' conceptions of basic principles in sequence stratigraphy. *Journal of Geoscience Education*, 61(1), 89–102.
- Hestenes, D., Wells, M., & Swackhamer, G. (1992). Force concept inventory. *Physics Teacher*, 30(3), 141–158.
- Jarrard, R. D., & Clague, D. A. (1977). Implications of Pacific island and seamount ages for the origin of volcanic chains. *Reviews of Geophysics*, 15(1), 57–76.
- 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.
- Jones, G., Taylor, A., & Broadwell, B. (2009). Estimating linear size and scale: Body rulers. *International Journal of Science Education*, 31(11), 1495–1509.
- Karlstrom, K., Semken, S., Crossey, L., Perry, D., Gyllenhaal, E. D., Dodick, J., Williams, M., Hellmich-Bryan, J., Crow, R., Watts, N. B., & Ault, C. (2008). Informal geoscience education on a grand scale: the trail of time exhibition at Grand Canyon. *Journal of Geoscience Education*, 56(4), 354–361.
- Kastens, K. A., & Rivet, A. (2010). Using analogical mapping to assess the affordances of scale models used in earth and environmental science education. In International Conference on Spatial Cognition (pp. 112–124). Berlin, Heidelberg: Springer.
- Kastens, K. A., Pistolesi, L., & Passow, M. J. (2014). Analysis of spatial concepts, spatial skills and spatial representations in New York State regents Earth science examinations. *Journal of Geoscience Education*, 62(2), 278–289.
- King, C. (2008). The earth science misconceptions of some science writers: how wrong can they be. *Teaching Earth Sciences*, 33(2), 9–11.
- King, C. J. H. (2010). An analysis of misconceptions in science textbooks: Earth science in England and Wales. *International Journal of Science Education*, 32(5), 565–601.
- Konrad, K., Koppers, A. A., Steinberger, B., Finlayson, V. A., Konter, J. G., & Jackson, M. G. (2018). On the relative motions of long-lived Pacific mantle plumes. *Nature Communications*, 9(1), 854.
- LaDue, N. D. (2018). *Scaffolding temporal reasoning with geologic timelines*. Retrieved from: <https://serc.carleton.edu/181886>. Accessed 2 Apr 2018.
- LaDue, N. D., Libarkin, J. C., & Thomas, S. R. (2015). Visual representations on high school biology, chemistry, earth science, and physics assessments. *Journal of Science Education and Technology*, 24(6), 818–834.
- Lakoff, G., & Johnson, M. (2008). *Metaphors we live by*. Chicago: University of Chicago press.
- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65–100.
- Lee, H., & Feldman, A. (2015). Photographs and classroom response systems in middle school astronomy classes. *Journal of Science Education and Technology*, 24(4), 496–508.
- Lee, H., & Schneider, S. E. (2015). Using astronomical photographs to investigate misconceptions about galaxies and spectra: question development for clicker use. *Physical Review Special Topics - Physics Education Research*, 11(2), 020101.
- Libarkin, J. C. (2006). College student conceptions of geological phenomena and their importance in classroom instruction. *Planet*, 17(1), 6–9.
- Libarkin, J. C., & Anderson, S. W. (2005). Assessment of learning in entry-level geoscience courses: results from the Geoscience Concept Inventory. *Journal of Geoscience Education*, 53(4), 394–401.
- Libarkin, J. C., & Kurdziel, J. P. (2006). Ontology and the teaching of Earth system science. *Journal of Geoscience Education*, 54(3), 408–413.
- Libarkin, J. C., Anderson, S. W., 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(5), 413–422.

- Lombardi, D., & Sinatra, G. M. (2012). College students' perceptions about the plausibility of human-induced climate change. *Research in Science Education, 42*(2), 201–217.
- Lombardi, D., Brandt, C. B., Bickel, E. S., & Burg, C. (2016). Students' evaluations about climate change. *International Journal of Science Education, 38*(8), 1392–1414.
- Manduca, C. A., & Kastens, K. A. (2012). Geoscience and geoscientists: uniquely equipped to study Earth. *Geological Society of America Special Papers, 486*, 1–12.
- Marshak, S. (2016). *Essentials of geology* (5th ed.). New York: W.W. Norton & Company, Inc..
- Mills, R., Tomas, L., & Lewthwaite, B. (2016). Learning in Earth and space science: a review of conceptual change instructional approaches. *International Journal of Science Education, 38*(5), 767–790.
- National Research Council. (2012). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. S. R. Singer, N. R. Nielsen, & H. A. Schweingruber (Eds.), Washington, D.C.: National Academies Press.
- New York State Education Department [NYSESED]. (2018). Science Regents Exams: Physical Setting/Earth Science. Retrieved from: <http://www.nysedregents.org/EarthScience/>.
- Newcombe, N. S., & Shipley, T. F. (2015). Thinking about spatial thinking: new typology, new assessments. In *Studying visual and spatial reasoning for design creativity* (pp. 179–192). Dordrecht: Springer.
- Parham, T. L., Cervato, C., Gallus, W. A., Larsen, M., Hobbs, J., Stelling, P., Greenbowe, T., Gupta, T., Knox, J.A., Gill, T. E. (2018). The InVEST volcanic concept survey: Exploring student understanding about volcanoes. *Journal of Geoscience Education, 58*(3), 177–187.
- Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education, 66*(2), 211–227.
- Resnick, I., & Shipley, T. F. (2013). Breaking new ground in the mind: an initial study of mental brittle transformation and mental rigid rotation in science experts. *Cognitive Processing, 14*(2), 143–152.
- Resnick, I., Davatzes, A., Newcombe, N. S., & Shipley, T. F. (2017). Using relational reasoning to learn about scientific phenomena at unfamiliar scales. *Educational Psychology Review, 29*(1), 11–25.
- Richardson, R. (2005). Teaching time in large enrollment intro classes: an active approach. *Geological Society of America Abstracts with Programs, 37*(7), 153.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction, 13*(2), 141–156.
- Shea, K. M. (2016). Beyond clickers, next generation classroom response systems for organic chemistry. *Journal of Chemical Education, 93*(5), 971–974.
- Shipley, T. F., Tikoff, B., Ormand, C., & Manduca, C. (2013). Structural geology practice and learning, from the perspective of cognitive science. *Journal of Structural Geology, 54*, 72–84.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction, 4*(4), 295–312.
- 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*(5), 539–555.
- 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., & Minogue, J. (2006). Accuracy of scale conceptions in science: mental maneuverings across many orders of spatial magnitude. *Journal of Research in Science Teaching, 43*(10), 1061–1085.
- United States Geological Survey [USGS]. (2017). Geology and the National Parks. Retrieved from: <https://geomaps.wr.usgs.gov/parks/pltec/>.
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: a meta-analysis of training studies. *Psychological Bulletin, 139*(2), 352–402.
- Vosniadou, S. (2002). On the nature of naive physics. In *Reconsidering conceptual change: issues in theory and practice* (pp. 61–76). Dordrecht: Springer.
- Vosniadou, S., & Brewer, W. F. (1987). Theories of knowledge restructuring in development. *Review of Educational Research, 57*(1), 51–67.
- Vosniadou, S. & Brewer, W. F. (1992) Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology, 24*(4), 535–585.
- Vosniadou, S., Vamvakoussi, X., & Skopeliti, I. (2008). The framework theory approach to the problem of conceptual change. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 3–34). New York: Routledge.
- Wade, N. J. (1996). Frames of reference in vision. *Minimally Invasive Therapy & Allied Technologies, 5*(5), 435–439.
- Wade, N. J., & Swanston, M. (2013). *Visual perception: An introduction*. London: Psychology Press.