Navigating Deep Time: Landmarks for Time From the Big Bang to the Present

Cesar Delgado[1,a]

ABSTRACT

People make sense of the world by comparing and relating new information to their existing landmarks. Each individual may have different landmarks, developed through idiosyncratic experiences. Identifying specific events that constitute landmarks for a group of learners may help instructors in gauging students' prior knowledge and in planning instruction that helps students build additional landmarks events. This paper proposes an operationalized definition for collective landmarks based on importance, accuracy, and precision. Including precision in the definition allows landmarks to be characterized for a group rather than an individual. This study evaluated the ability of undergraduate students in an interdisciplinary course to estimate scales of time related to major cosmological, geological, and historical events. Individual students responded to replicate questions in different formats with the same answers, indicating the testing format was valid. The students' estimates were then used to determine collective landmarks. The number of collective landmarks increased between the pretest and posttest. Collective landmarks included extremely ancient events (Big Bang, formation of the solar system) or relatively recent ones (Cold War, the age of empires, emergence of nation states). Intermediate events had low accuracy, low precision, or both. These data indicate that lecture courses can teach students collective landmarks for time. Because landmarks can be learned, geoscience programs might consider coordinated planning of key landmarks to be introduced at different stages of their academic programs. © *2013 National Association of Geoscience Teachers*. [DOI: 10.5408/12-300.1]

Keywords: deep time, landmarks, science education, crosscutting concepts

INTRODUCTION

Understanding deep time is a central problem in the study of many scientific disciplines including geoscience (Trend, 1998; Zen, 2001), evolution (Catley and Novick, 2009), paleontology, and cosmology (Dodick and Orion, 2006). Many science education standards in the K–12 sector propose that stability (or constancy) and change are crosscutting ideas that pervade science and can help students connect their knowledge across disciplines (American Association for the Advancement of Science, 1993; Bransford et al., 1999; National Research Council, 2012). University science standards also consider that scale—including scales of time—is a unifying concept (College Board, 2009). Understanding deep time is essential in grasping constancy and change, because a phenomenon seems to be constant or changing depending on the scale of time employed (e.g., the continents seem unchanging on a human timescale, but collide, merge, and break up on a geological timescale). A recent study found that not holding the belief that the Earth is at least four billion years old poses a challenge for students to be able to accept evolution (Cotner et al., 2010), yet an earlier study showed that fewer than half of the undergraduate participants believed that the Earth is between four and five billion years old (Libarkin et al., 2005).

There are relatively few research reports on student understanding of deep time (Dodick and Orion, 2006; for recent reviews, see Libarkin et al., 2007; and Teed and Slattery, 2011). However, studies on the understanding of size may shed light on the understanding of time, as the mental mechanisms for thinking about magnitudes of space and time seem to be related (Walsh, 2003; Casasanto and Boroditsky, 2008). Prior studies have found that ordering is easier for students than ascribing magnitudes to the items or events (e.g., Brown and Siegler, 2001; Trend, 2001; Tretter et al., 2006; Libarkin et al., 2007). Yet ordering is not enough; knowing the actual ages of some pivotal events is essential in developing a "deep time framework" (Trend, 2001). These important events of known ages can function as "landmarks" for students when thinking about deep time.

Given that various studies (reviewed below) have found a lack of knowledge of the time frame of most geological and biological events, it is important to define and operationalize landmarks for time, identify what events constitute landmarks for students at various grade levels, and examine whether students construct additional landmarks when taking courses in geoscience, history, or other disciplines where time is prominent. This paper addresses some of these research gaps. In particular, the novel concept of "collective landmark" is defined and operationalized with a methodology to assess whether specific events constitute landmarks for a group of learners. This definition and methodology may help make more systematic the study of the knowledge about age, size, or other magnitudes held by groups of learners. The identification of collective landmarks should be useful for instructors and curriculum designers, allowing them to gain a clearer picture of students' prior knowledge and to plan instruction that helps students build additional landmarks events. The study of how students' landmarks change when taking an undergraduate geoscience or history course also provides a baseline useful in planning and evaluating curricular or pedagogical innovations in such courses.

Received 2 February 2012; revised 10 September 2012; accepted 1 October 2012; published online 21 February 2013.

[1]Department of Curriculum and Instruction, University of Texas at Austin, 1912 Speedway Stop D5500, Austin, Texas 78712, USA
[a]Author to whom correspondence should be addressed: Electronic mail: cesar_delgado@austin.utexas.edu

THEORETICAL FRAMEWORK

From a constructivist perspective (e.g., Piaget, 1983; Steffe and Gale, 1995), learners do not simply accumulate information but actively construct their understanding. In this view, making connections across pieces of knowledge is fundamental to learning. As learners encounter new phenomena, they make sense of them by comparing and connecting to prior knowledge. Landmarks may allow students to learn about new events or objects by providing a familiar reference against which to compare and contrast new events or objects.

Importance of Landmarks

Various researchers have proposed that we make sense of the world by comparing and relating new information to landmarks (also called benchmarks, anchors, or reference points). Some of this research has focused on students' understanding of the size of objects, but is included here because the cognitive processes and neural bases for thinking of temporal and spatial magnitudes appear to be closely connected (Walsh, 2003; Casasanto and Boroditsky, 2008). Almost 40 years ago, Tversky and Kahneman identified "adjustment and anchoring," or making estimates starting from a known value for a related case and adjusting accordingly, as a fundamental heuristic that learners employ to predict values (1974). More recently, artificial intelligence studies have modeled how learners classify or identify objects based on proximity to anchors (Petrov and Anderson, 2005). Expert scientists have been found to envision size regimes or "worlds" characterized in part by landmarks (Tretter et al., 2006). Researchers have proposed that a sense of measurement is constructed by reference to benchmarks (Joram, 2003). Estimation strategies for temporal or spatial characteristics require knowledge of accurate magnitudes of "benchmarks" (Lee et al., 2011) or "reference points" (Joram et al., 1998). Estimating the date of events experienced personally becomes more accurate when landmarks are used (Loftus and Marburger, 1983). Similarly, estimation of data such as the population of countries or distance between cities improves after "seeding" with accurate information for related cases (Brown and Siegler, 2001), which again suggests the importance of landmarks in thinking about novel information. In the case of time, landmarks can help students establish a deep time framework by anchoring positions along a mental or externally represented timeline onto which other events can be placed.

Individual and Collective Landmarks

Each individual is likely to have a different set of landmarks for time and space, built from idiosyncratic experiences and interests. For instance, Jones and Taylor (2009) found that experts used anchor points when dealing with scale; these landmarks varied widely across scientists, including the diameter of red blood cells ($\sim$7 μm) and proteins ($\sim$100 nm) for a chemist, and the sizes of humans and elephants for a paleontologist.

Instructors in fields where deep time (or another magnitude) is crucial are likely to find it impractical to personalize instruction in order to leverage students' idiosyncratic *individual* landmarks. Thus, this study concerns the definition and identification of *collective* landmarks—important events or objects of which a *group* of students has accurate knowledge—that the instructor may use productively to help students build a better understanding of deep time. These collective landmarks are likely to vary across groups depending on the extent of their knowledge, and also within a single group over time, depending on instruction. Previously, two characteristics of landmarks have been proposed: *accuracy* of the estimates, and *importance* of the events (Loftus and Marburger, 1983). A third characteristic proposed in this paper for collective landmarks, is *precision*: for an event to constitute a collective landmark, it is necessary that most students estimate the same or similar age (i.e., that there is a narrow distribution of estimates). Determining collective landmarks for deep time might thus involve selecting events that are important in the disciplines of geoscience and biology and then measuring the accuracy and precision of the estimates of the ages of the events, across multiple students. The first two criteria (importance, accuracy) are straightforward—one could consult a textbook to determine what events are important and their scientifically accepted ages. However, the degree of precision that is required for an event to be considered a collective landmark still needs to be determined. Previous research studies that bear on this issue are discussed next.

The Precision Criterion for Collective Landmarks

Trend's research characterized the knowledge of geoscience events from the Big Bang (13.6 billion years ago [bya]) to the extinction of woolly mammoths ($\sim$10,000 years ago [ya]), for various age groups (1998, 2000, 2001). However, only the mean and standard deviation for the *rank* of the events was reported, not measures of central tendency and spread in the estimates themselves. Libarkin and colleagues (Libarkin et al., 2007) also investigated undergraduates' knowledge of geologic time. They used a timeline-construction task that involved estimating the ages of the following events: the age of Earth, the time required for the origin of the first life forms, the first appearance of dinosaurs, and the first appearance of humans. This study found an enormous range of estimates. Around 10% of students estimated 500,000 years or less for the age of the Earth and one estimated trillions of years; around 5% did not assign an age but indicated a creationist viewpoint. Appropriate measures of central tendency and spread that would inform about accuracy and precision were again not reported; thus, these two studies do not provide information on the precision required for an event to constitute a collective landmark. Catley and Novick (2009) investigated undergraduates' estimates of the age of seven evolutionarily significant events from the formation of the Earth (4–4.6 bya) to the establishment of the hominid lineage (2 million years ago [mya]). The undergraduates' answers varied widely for each of the seven events, and ranged over all events from 800 ya to 600 bya (even after excluding students with a creationist viewpoint). They concluded that students were unable to connect evolutionary events to historical happenings. Catley and Novick's study reported the estimates at the 25th, 50th, and 75th percentile, and it thus provides information about the measure of spread of student estimates. The narrowest distribution of estimates centered on an accurate value was for the age of the Earth, with estimates varying by a factor of three between the 25th and 75th percentiles. All other events had at least a thirtyfold factor.

Defining Collective Landmarks

Based on these studies, precision can now be operationalized. Since previous studies have found strong outliers, the 25th and 75th percentiles are used to assess the spread of the data rather than standard deviation, as the percentiles are more resistant to the influence of outliers. Taking the geometric midpoint between the first and second narrowest distributions in the study using percentiles (Catley and Novick, 2009), a tenfold difference between the 25th and 75th percentiles is proposed as the cutoff point for an event's distribution of students' estimates to be considered precise enough to serve as a useful collective landmark. The following operationalized definition is thus proposed:

> *A collective landmark for deep time is an important event in a discipline of science for which the 25th and 75th percentiles of the estimates of a group of learners are within an order of magnitude of each other, and where the accepted scientific value falls within the 25th and 75th percentile estimates.*

Applications of Collective Landmarks

Collective landmarks could be leveraged profitably both in instruction and in curriculum development. Instructors can identify and use the existing collective landmarks to place new or lesser-known events in context, in addition to providing the absolute age of the new events. This strategy would take advantage of the relative ease with which learners can order (Brown and Siegler, 2001; Trend, 2001; Tretter et al., 2006; Libarkin et al., 2007), and more importantly, would help students relate the magnitudes of events (or objects) to each other. Establishing these mental connections is hypothesized to lead to more robust understanding (Hiebert and Lefevre, 1986), knowledge that is easier to retrieve and recall. Collective landmarks can also be of use in curriculum design: instructors can determine the important landmarks they wish students to acquire during a course, and then include activities aimed at helping students develop those landmarks. This approach of including landmark development as a goal of instruction can be extended to a departmental program as well.

Analyzing the characteristics of events that are *not* collective landmarks can also be useful instructionally. Events that feature *precision but not accuracy* may signal a widely held alternative idea that instruction can be designed to explicitly confront. For instance, if most students believe that early human ancestors and dinosaurs coexisted, the estimates for the age of the extinction of dinosaurs might cluster around 1 mya (operationalized as the 25th through 75th percentiles of student estimates being within an order of magnitude, and including 1 mya), rather than the accepted value of around 65 mya. Events that feature *accuracy but not precision* in student estimates of age instead indicate a lack of certainty in the student group, but no single, commonly held alternative conception. Instruction that builds familiarity with such events (especially in relation to landmarks events) is likely to result in improved estimates by many students, increasing precision. Thus, different instructional approaches are suitable for different events that are not collective landmarks, depending on the accuracy and precision of the event.

Research Goals

The following research questions guide this study, which utilizes the novel concept of collective landmarks introduced above:

1. What events constitute collective landmarks for time for university students?
2. What impact do typical instructional activities have on student collective landmarks?

METHODS

Context

This study took place in an interdisciplinary undergraduate course crosslisted in four departments, including Geoscience and History, at a public research university in the midwestern U.S. This course covered major events from the Big Bang to the present from the disciplinary perspectives of various natural and social sciences. The course was an elective for students, but was one option to fulfill a requirement for the History BA degree program. The two key themes in the course were scale, and complexity and connection. The course focused on scale through the logarithmic organization of the syllabus (e.g., the week covering the Big Bang and early stages of the universe was labeled log[time] = 10, and the next week, focusing on geoscience, was labeled log[time] = 9). The course started with astrophysics, then geosciences, followed by chemistry, biology, then archaeology, and human history; each discipline was nested within the previous one. The focus on complexity and connections traced the emergence of ever more complex aggregations, with growing use of energy and increased instability. Instruction was in traditional lecture style, with smaller discussion sections led by graduate teaching assistants. Lecturers from the various departments led lectures on different topics, with the instructor of record leading a total of eight classes including the first two and the last class. A total of 26 additional instructors taught from one to three lectures each. The author was not involved in designing or teaching the course, and there was no specific attempt to guide students in developing landmark events. While this course provided rich opportunities for students to develop individual landmarks, there was no use of the concept of collective landmarks in the course.

Participants

The participants were recruited from among the 66 undergraduates initially enrolled in the course. Of 64 students who consented to participate, 48 matched pairs of pre- and posttests were obtained (the reduction is due mainly to students moving in and out of the class) and are analyzed here. Around 11% of the students were freshmen, 30% sophomores, 24% juniors, and 35% seniors. The majors of the participants varied widely, including 15 history, American culture, or classical civilization majors; five engineering majors; six math, Earth science, astronomy, neuroscience, and/or computer science majors; 10 economics, sociology, or political science majors; six undeclared; and others (double majors counted in each major).

Data Sources

Students completed identical paper-and-pencil tests with five questions of various formats during the first and next to last discussion sections of the course. The question analyzed here asked students to date the following 10 events: Big Bang, formation of the solar system, appearance of first mammals, extinction of dinosaurs, start of the Stone Age, first controlled use of fire by hominids, invention of agriculture, advent of earliest empires, appearance of first modern nation-states, and the end of the Cold War (see Appendix A). The more recent historical events were included because they permit covering the entire range of time, and including these is relevant because prior research reported that undergraduate students had difficulty connecting more ancient evolutionary events with more recent, historical events (Catley and Novick, 2009). The older events have been included in various previous studies and allow comparisons to these studies. All of the events were important in class; the Big Bang, formation of the solar system, hominids, empires, and states were the topics of individual lectures or sections, while the others were key events in the topics of biology and archaeology as covered in the class. Ages of the event were estimated by students by classifying each event as belonging to one of 11 age ranges, from <10 ya to >10 bya. The format of this item is based on the Scale of Objects Questionnaire (Tretter et al., 2006) and Trend's Question 3 (Trend, 1998), but uses ranges that are consistently tenfold: 10–100 ya, 100–1,000 ya, etc. (Trend used ranges that varied from fivefold to 2,000-fold; Tretter and colleagues used ranges of either tenfold or 1,000-fold). Alternatively, students could write the age of the event directly. This format provides possible answers for students to choose from, which is desirable because previous studies have shown that some students are unable to produce an estimate (Libarkin et al., 2005); however, it also provides students with the opportunity to write down a specific date if they so prefer. For an event to be a landmark, its age must be known *by science* within a reasonably small range; for this study, each event could unambiguously be assigned to a range.

Semistructured interviews were conducted with 10% of the participants and provided a means of testing the validity of the assessment instrument.

Analysis

Identification of Collective Landmarks

Individual responses were transformed into a single number by taking the base 10 logarithm of the high end of the range selected (e.g., for an estimate of 10–100 ya for Cold War, the log of the value of the high end [100] was taken, resulting in a value of 2). Given that consistent tenfold ranges were used, using the geometric mean of each range would be essentially equivalent, but using an endpoint results in a round number that is easier to work with. Written answers were treated as if the student had marked the corresponding size range. For the highest age range (>10 bya), which is unbounded, 11 was used—as if the range were 10–100 bya. Using SPSS 19 for Mac, the percentiles were calculated for each multiple of five (i.e., 5th, 10th, 15th . . . 95th percentiles) for each event. Finally, collective landmarks were identified as those events that had the same value for the 25th and 75th percentiles, where that value also corresponded to the scientifically accepted value. Recall that these values originated with the selection of a tenfold age range, so having the same value for the 25th and 75th percentiles conforms to the precision criterion in the definition for collective landmarks presented earlier, of having the 25th and 75th percentiles within an order of magnitude.

This analysis did not distinguish between STEM and non-STEM majors, but in prior research, order of magnitude estimates did not differ in accuracy between students with stronger and weaker biology backgrounds (Catley and Novick, 2009).

Effect of Instruction on Collective Landmarks

This analysis consisted of two parts. The first simply involves comparing the events that met the criterion for collective landmarks at the beginning and the end of the class. The second part of the analysis stems from the recognition that the accuracy and/or precision of the estimates by the students in a class might conceivably improve, yet not meet the criteria for collective landmark status. Each event that was not a collective landmark by the end of class was analyzed for changes in accuracy and precision. For accuracy, the criterion that the accepted value fall between the 25th and 75th percentile of the student answers was used. For precision, the difference in the values corresponding to the 25th and 75th percentiles was compared, to see if the difference decreased. With the task format that this study employed, the precision criterion for a collective landmark is a difference in values of the 25th and 75th percentiles of zero, meaning that the same tenfold time range was selected by students at the 25th and 75th percentiles. In a free-response format, the precision criterion would instead be for the 25th and 75th percentiles to be, at most, one order of magnitude apart.

RESULTS AND DISCUSSION

The findings presented and discussed below include identification and comparison of the collective landmarks at the beginning and end of the course, and the analysis of the effects of instruction on precision and accuracy for non-landmark events, from beginning to end of course. The values and patterns of student estimates are also briefly compared to those reported in previous research.

Identification of Collective Landmarks at Beginning of Course

Figure 1 shows the value of student estimates from the 5th to the 95th percentiles, at the beginning of the course. Correct values for each event are bolded, the region comprising 25th through 75th percentile values is offset by a blank line, and the intersection of these two sets is shaded gray (i.e., the percentiles between 25th and 75th with accurate estimates).

Events that met the criteria for both accuracy and precision are the Big Bang, the age of empires, and Cold War. These three events qualified as collective landmarks and are outlined with a black border in Fig. 1.

Events that met the criterion for accuracy but not for precision include the formation of the Solar System, the extinction of the dinosaurs, the invention of agriculture, and the emergence of the first modern nation-states. There is a lack of knowledge about the age of these events among the students in the class, but no indication of a single alternative idea. While over half of the students estimated the correct age range for nation-states (60.4%), there were also many (29.2%) who selected the next greater age range. The 25th and 75th percentile criterion for precision detected this bimodal distribution, whereas a precision criterion focusing only on the proportion of students with accurate responses would not allow the detection of a bimodal distribution.

PERCENTILES	BIG BANG	SOLAR SYSTEM	MAMMALS	DINOSAURS	STONE AGE	FIRE	AGRICULTURE	EMPIRES	NATION-STATE	COLD WAR
5	10.00	7.00	4.00	5.00	4.00	3.35	3.45	3.45	2.40	**2.00**
10	10.00	7.80	4.00	5.00	4.00	4.00	4.00	**4.00**	**3.00**	**2.00**
15	**11.00**	9.00	5.00	6.00	4.00	4.00	4.00	**4.00**	**3.00**	**2.00**
20	**11.00**	9.60	5.00	6.00	4.00	4.00	4.00	**4.00**	**3.00**	**2.00**
25	**11.00**	**10.00**	6.00	7.00	4.00	4.00	4.00	**4.00**	**3.00**	**2.00**
30	**11.00**	**10.00**	6.00	7.00	5.00	4.00	4.00	**4.00**	**3.00**	**2.00**
35	**11.00**	**10.00**	6.00	7.00	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
40	**11.00**	**10.00**	6.60	7.00	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
45	**11.00**	**10.00**	7.00	**8.00**	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
50	**11.00**	**10.00**	7.00	**8.00**	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
55	**11.00**	**10.00**	7.95	**8.00**	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
60	**11.00**	**10.00**	8.00	**8.00**	6.00	5.00	4.00	**4.00**	**3.00**	**2.00**
65	**11.00**	**10.00**	8.00	**8.00**	6.00	5.00	4.00	**4.00**	3.20	**2.00**
70	**11.00**	**10.00**	8.00	**8.00**	6.00	5.00	**5.00**	**4.00**	4.00	**2.00**
75	**11.00**	11.00	8.00	**8.00**	6.75	5.00	**5.00**	**4.00**	4.00	**2.00**
80	**11.00**	11.00	8.20	**8.00**	**7.00**	**6.00**	**5.00**	**4.00**	4.00	**2.00**
85	**11.00**	11.00	**9.00**	**8.00**	7.65	7.00	**5.00**	5.00	4.00	**2.00**
90	**11.00**	11.00	**9.00**	9.00	9.00	7.30	6.00	5.10	4.00	**2.00**
95	**11.00**	11.00	**9.00**	9.60	9.55	8.00	7.00	6.00	4.60	**2.00**

FIGURE 1: Pretest value of student estimates from 5th through 95th percentiles. Correct values for each event are bolded, and correct values within the 25th through 75th percentiles are shaded gray.

There were no events that met the criterion for precision but not for accuracy. Such events would indicate the existence of a single, widely held alternative idea. However, the invention of agriculture and the first controlled use of fire were close to meeting the precision criterion (25th through 65th and 35th through 75th percentiles, respectively). In both cases, student estimates tended to be one order of magnitude smaller than the accepted value.

The appearance of mammals, the Stone Age, and the controlled use of fire failed to meet the criteria for both accuracy and the precision. The accepted value is not included in the range between the 25th and 75th percentiles (accuracy criterion not met), and the 25th and 75th percentiles are not within an order of magnitude. That is, they do not correspond to the same tenfold age range (precision criterion not met).

These findings indicate that these students have few landmarks to leverage in building a "deep time framework" (Trend, 2001). Thus, instructors should aim to develop some key events into collective landmarks with which to contextualize other events. The age of most non-landmark events was underestimated. This is consistent with the "forward telescoping" (Loftus and Marburger, 1983) reported by prior research (Catley and Novick, 2009; Lee et al., 2011). The greater accuracy and precision of events at the extremes of the range is also consistent with prior research, which has suggested that people may use the endpoints of a scale to inform their estimates (Berger et al., 1987).

Effect of Instruction on Collective Landmarks

Figure 2 shows the 5th through 95th percentiles of student estimates for the same 10 events at the end of the course. The Big Bang, age of empires, and Cold War continue to be collective landmarks, and the formation of the solar system and the emergence of modern nation-states became collective landmarks. The extinction of dinosaurs, invention of agriculture, and appearance of mammals showed no change in accuracy (dinosaurs and agriculture still meeting, and mammals still not meeting the criterion for accuracy) or precision (with the same values for the 25th and 75th percentiles as on the pretest). The Stone Age increased in precision (from 4–6.75 on the pretest to 5–6 on the posttest) but still did not meet the accuracy criterion. Finally, fire met the accuracy criterion on the posttest, after failing to meet it on the pretest. In sum, nation-states and solar system became landmarks, Stone Age improved in precision, and fire improved in accuracy; the other events did not change. Presumably, these changes took place due to the course; however, other factors cannot be discounted and thus causality cannot be claimed.

It is worth noting that this group has no collective landmarks for the entire history of the planet, up until historical times. Around one-quarter of the lectures in the course dealt with geosciences, chemistry, biology, and archaeology, and were situated in the time period after the formation of the Earth and before recorded history; yet, events from this time period (the emergence of mammals, extinction of the dinosaurs, the Stone Age, and controlled use of fire) did not become collective landmarks. Once again, estimates of ages of events nearer the extremes were more accurate and precise than those in the middle (Berger et al., 1987), and the age of non-landmark events was underestimated, consistent with forward telescoping (Loftus and Marburger, 1983).

Comparison of Student Estimates to Previous Research

Most previous studies on students' ideas about the age of events have reported the proportion of accurate answers. Table I shows the percentage of student estimates that were correct for each event, on pre- and posttest. These range from 100% for Cold War (posttest) to less than could be expected from random guessing for fire on the pretest and Stone Age on the posttest (chance is 1/11, or 9%). Libarkin and colleagues suggest that accuracy below chance may indicate the existence of a widespread alternative conception (2005); however, this criterion is insufficient, as students may hold a variety of inaccurate beliefs. For instance, on the posttest, student responses for Stone Age included 33.3% for one order of magnitude less and 35.4% for two orders of magnitude less than the accepted value. The criterion to diagnose a widely held alternative idea proposed in this paper, of precision without accuracy (where precision is operationalized as 25th and 75th percentiles within an order of magnitude), is more complete than simply using accuracy below chance.

Some events overlap between this study and the Libarkin et al. (2007), Catley and Novick (2009), and Lee et al. (2011) studies—all conducted with undergraduates (see Table I). All studies had similar rates of accurate estimates, despite involving undergraduates from different institutions and using different task formats. For instance, around one-half to two-thirds of students correctly estimated the age of the solar system (or Earth) in the pretest of this study and the other three studies. Most undergraduates are somewhat familiar with the ages of significant events, and the ages of some events are more likely than others to be estimated accurately. The Big Bang was the event most commonly estimated with accuracy, followed by the age of the solar system or dinosaur extinction, and the origin of mammals had the least accurate estimates. Further research is needed to elucidate why some events have lower percentages of accurate estimates than others; the presence of alternative conceptions presented in popular media (e.g., dinosaurs coexisting with humans) as well as religious teachings may be involved (Libarkin et al., 2005).

Reliability and Validity

The tasks analyzed here are very similar to tasks used in papers published previously in peer-reviewed literature (Trend, 1998; Tretter et al., 2006), although the analysis and conceptualization are novel. Validity was investigated by asking several students to explain their answers on the grid of events and age ranges (see Appendix A), and they confirmed that they were answering as intended. Around 12.5% of the tests included one or more estimates of the age of events expressed as both a written, open-response estimate and as a mark on a cell corresponding to age ranges; over 99% of these were consistent, again showing that students were recording their estimates using the grid as intended. In sum, there is evidence from various sources that the instrument was gathering the intended responses from students. In a few cases, students marked two responses for a single event, usually skipping the adjacent event, probably due to a visual error. Such responses were treated as missing data for those events. The sensitivity to instruction of the instrument, with landmarks developing for events that were mentioned during the class, provides evidence for content validity.

A limitation of this study stems from the structured format, in which possible answers are provided to students. The answers precluded students' providing extreme answers smaller than 10 ya or larger than 10 billion ya, and scaffolded students in providing possible answers. Additionally, the use of tenfold ranges for the estimates of age creates some issues. For instance, a student who believes that agriculture originated only 6,000 years ago (rather than the accepted value of around 12,000 ya) would mark the column for 1,000–10,000 ya and would contribute to the spread of the distribution; whereas a student who believes agriculture

PERCENTILES	BIG BANG	SOLAR SYSTEM	MAMMALS	DINOSAURS	STONE AGE	FIRE	AGRICULTURE	EMPIRES	NATION-STATE	COLD WAR
5	10.00	8.45	4.45	4.00	4.00	4.00	3.00	3.00	**3.00**	**2.00**
10	10.00	9.00	5.00	4.90	4.00	4.00	3.90	3.90	**3.00**	**2.00**
15	10.00	**10.00**	5.35	6.00	4.00	4.00	4.00	**4.00**	**3.00**	**2.00**
20	10.80	**10.00**	6.00	6.00	4.00	4.00	4.00	**4.00**	**3.00**	**2.00**
25	**11.00**	**10.00**	6.00	7.00	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
30	**11.00**	**10.00**	6.00	7.00	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
35	**11.00**	**10.00**	7.00	7.00	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
40	**11.00**	**10.00**	7.00	7.00	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
45	**11.00**	**10.00**	7.00	7.05	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
50	**11.00**	**10.00**	7.50	**8.00**	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
55	**11.00**	**10.00**	8.00	**8.00**	5.00	5.00	4.00	**4.00**	**3.00**	**2.00**
60	**11.00**	**10.00**	8.00	**8.00**	6.00	5.00	**5.00**	**4.00**	**3.00**	**2.00**
65	**11.00**	**10.00**	8.00	**8.00**	6.00	5.00	**5.00**	**4.00**	**3.00**	**2.00**
70	**11.00**	**10.00**	8.00	**8.00**	6.00	5.00	**5.00**	**4.00**	**3.00**	**2.00**
75	**11.00**	**10.00**	8.00	**8.00**	6.00	**6.00**	**5.00**	**4.00**	**3.00**	**2.00**
80	**11.00**	**10.00**	8.20	**8.00**	6.00	**6.00**	**5.00**	**4.00**	4.00	**2.00**
85	**11.00**	**10.00**	**9.00**	**8.00**	6.00	**6.00**	**5.00**	**4.00**	4.00	**2.00**
90	**11.00**	10.10	**9.00**	9.00	**7.00**	**6.00**	6.00	5.00	4.00	**2.00**
95	**11.00**	11.00	**9.00**	9.00	7.55	**6.60**	6.00	5.55	4.00	**2.00**

FIGURE 2: Posttest value of student estimates from 5th through 95th percentiles. Correct values for each event are bolded, and correct values within the 25th through 75th percentiles are shaded gray.

originated 24,000 years ago would mark the column including the correct value and would not contribute to the spread of the data, even though both students were off by a factor of two. The use of 10 specific events to test as possible landmarks clearly leaves many important events in geoscience and history unaddressed. Furthermore, logistical constraints made it impossible to conduct follow-up interviews after the end of the course to evaluate the stability of the changes in students' collective landmarks.

CONCLUSION AND IMPLICATIONS

Based on constructivist learning theory, an individual's landmarks for time can be instrumental in developing a deep

TABLE I: Comparison of accuracy of estimates in this and other studies (percentages).

Event	This Study, Pretest	This Study, Posttest	Catley and Novick (2009)	Libarkin et al. (2007)	Lee et al. (2011)[1]
Big Bang/Age of Universe	89.6	81.3			~80
Solar System/Earth	52.1	81.3	67.3	57	~50
Mammals	16.7	16.7	23.8		
Dinosaur extinction	45.8	41.7	35		~55

[1]Values are approximate as these were read from a figure.

time framework. Since instruction takes place in groups, however, it is important to think about how a group's landmarks can be defined and measured. This paper proposed a definition for collective landmarks that includes importance, accuracy, and precision, and that can be applied to free response or appropriate forced choice tests. The methodology was tested on undergraduates enrolled in a course that paid attention to scales of time and space, and which covered events from the Big Bang to the present; new collective landmarks emerged at the end of the course, and some events increased in accuracy or precision but did not reach landmark status.

Instructors of geoscience, history, or other fields where deep time is crucial can conveniently measure their classes' initial collective landmarks by administering a test much like the one employed in this study, but customized to include the events that they consider important for that course. Using such a test as formative assessment (Black and Wiliam, 1998) will allow the instructors to tailor their classes to their students, making it more responsive to their needs. Students' progress in establishing the desired landmarks can be monitored throughout the course. While this study used a paper-based instrument, the use of electronic survey systems could conceivably automate analysis of the results.

Faculty in departments that have well-defined sequences of courses can collaborate in establishing a plan for student development of landmarks across their entire program, with each class's final landmarks constituting the next course's starting point. The definition and methodology proposed and tested here allow for the instruction in a department to be coordinated more systematically around timescales and the crosscutting concepts of constancy and change.

The definition and operationalization of collective landmarks also provides a basis for more systematic research on students' ideas of the magnitude of phenomena such as the age of events or the size of objects. This study showed that traditional lecture-style instruction can build additional collective landmarks for events that are salient in the class. Future research is needed to determine whether alternate modes of instruction are more effective in helping students build landmarks.

This study identified three collective landmarks for the age of events that undergraduates had coming into the course (Big Bang, age of empires, Cold War), and two that developed during the course (nation-state emergence, formation of the solar system).

Since collective landmarks are sensitive to instruction, faculty in geology (and other) departments can collaboratively determine how to best develop these landmarks throughout their programs. Building a solid set of mental landmarks for time and for space during freshman-level courses would provide a scaffold for students to meaningfully learn about phenomena where scale is important in subsequent courses—and those courses can continue to support students to build additional collective landmarks. Collective landmarks can constitute a useful tool to help geology students make temporal and spatial connections throughout the courses of a degree program, in effect leveraging the potential of scale and constancy and change as crosscutting ideas in science.

Acknowledgments

This paper benefited from very helpful comments by Cathy Manduca, Jill Marshall, and an anonymous reviewer. A heartfelt thanks to the instructor, lecturers, teaching assistants, and students in the Geoscience/History class. This study was supported by the University of Michigan Rackham Merit Fellowship Program and the University of Michigan School of Education Scholar's Award fellowship.

REFERENCES

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

Berger, C.F., Pintrich, P.R., and Stemmer, P.M. 1987. Cognitive consequences of student estimation on linear and logarithmic scales. *Journal of Research in Science Teaching*, 24:437–450.

Black, P., and Wiliam, D. 1998. Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan*, 80:139–148.

Bransford, J.D., Brown, A.L., and Cocking, R.R. eds. 1999. How people learn: Brain, mind, experience, and school, 1st ed. Washington, DC: National Academies Press.

Brown, N.R., and Siegler, R.S. 2001. Seeds aren't anchors. *Memory & Cognition*, 29:405–412. doi: 10.3758/BF03196391.

Casasanto, D., and Boroditsky, L. 2008. Time in the mind: Using space to think about time. *Cognition*, 106:579–593.

Catley, K.M., and Novick, L.R. 2009. Digging deep: Exploring college students' knowledge of macroevolutionary time. *Journal of Research in Science Teaching*, 46:311–332. doi: 10.1002/tea.20273.

College Board. 2009. Science college board standards for college success. New York: College Board.

Cotner, S., Brooks, D.C., and Moore, R. 2010. Is the age of the earth one of our "sorest troubles?" Students' perceptions about deep time affect their acceptance of evolutionary theory. *Evolution*, 64:858–864.

Dodick, J., and Orion, N. 2006. Building an understanding of geological time: A cognitive synthesis of the "macro" and "micro" scales of time. *In* Manduca, C.A. and Mogk, D.W. eds., Earth and mind: How geologists think and learn about the Earth, Special Paper. Boulder, CO: Geological Society of America, p. 77–93.

Hiebert, J., and Lefevre, P. 1986. Conceptual and procedural knowledge. *In* Hiebert, J. ed., Conceptual and procedural knowledge: The case of mathematics. Hillsdale, NJ: L. Erlbaum Associates, p. 1–27.

Jones, M.G., and Taylor, A.R. 2009. Developing a sense of scale: Looking backward. *Journal of Research in Science Teaching*, 46:460–475. doi: 10.1002/tea.20288.

Joram, E. 2003. Benchmarks as tools for developing measurement sense. *In* Clements, D.H. and Bright, G.W., eds., Learning and teaching measurement: 2003 yearbook. Reston, VA: NCTM, p. 57–67.

Joram, E., Subrahmanyam, K., and Gelman, R. 1998. Measurement estimation: Learning to map the route from number to quantity and back. *Review of Educational Research*, 68:413–449. doi: 10.3102/00346543068004413.

Lee, H., Liu, O.L., Price, C.A., and Kendall, A.L.M. 2011. College students' temporal-magnitude recognition ability associated with durations of scientific changes. *Journal of Research in Science Teaching*, 48:317–335. doi: 10.1002/tea.20401.

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

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

Loftus, E.F., and Marburger, W. 1983. Since the eruption of Mt. St. Helens, has anyone beaten you up? Improving the accuracy of retrospective reports with landmark events. *Memory & Cognition*, 11:114–120.

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

Petrov, A.A., and Anderson, J.R. 2005. The dynamics of scaling: A memory-based anchor model of category rating and absolute identification. *Psychological Review*, 112:383–416. doi: 10.1037/0033-295X.112.2.383.

Piaget, J. 1983. Piaget's theory. *In* Mussen, P., ed., Handbook of child psychology, 4th ed., Vol. 1. New York: Wiley.

Steffe, L.P., and Gale, J.E., eds. 1995. Constructivism in education. Hillsdale, NJ: Lawrence Erlbaum.

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

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

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

Trend, R.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*, 38:191–221. doi: 10.1002/1098–2736(200102)38:2<191::AID-TEA1003>3.0.CO;2-C.

Tretter, T.R., Jones, M.G., Andre, T., Negishi, A., and Minogue, J. 2006. Conceptual boundaries and distances: Students' and experts' concepts of the scale of scientific phenomena. *Journal of Research in Science Teaching*, 43:282–319. doi: 10.1002/tea.20123.

Tversky, A., and Kahneman, D. 1974. Judgment under uncertainty: Heuristics and biases. *Science*, 185:1124–1131.

Walsh, V. 2003. A theory of magnitude: Common cortical metrics of time, space, and quantity. *Trends in Cognitive Sciences*, 7:483–488. doi: 10.1016/j.tics.2003.09.002.

Zen, E.A. 2001. What is deep time and why should anyone care? *Journal of Geoscience Education*, 49:5–9.

APPENDIX A

Interview: NO YES (email: ________________________) Name_____________________________

Post-Course Assessment

You will have 15 to 20 minutes to answer five questions. This assessment will help us track student learning across the semester. Don't worry if you are not sure about the answers -- just provide your best estimate. This assessment ***WILL NOT*** affect your grade in the class. Raise your hand when you are done with the first section so we can give you the second section.

This question is about your perception of events in the past. Please consider each of the following events, and indicate HOW LONG AGO each one occurred. Please "X" the box that is closest to your own estimate. Please try to respond to each item. *If you prefer to give an exact time instead, write it in the first column.*

Event in the past	Estimated time in years since event occurred	A: <10 yrs ago	B: 10 to 100 yrs ago	C: 100 to 1000 yrs ago	D: 1000 to 10,000 yrs ago	E: 10,000 to 100,000 yrs ago	F: 100,000 to 1 million yrs ago	G: 1 million to 10 million yrs ago	H: 10 million to 100 million yrs ago	I: 100 million to 1 billion yrs ago	J: 1 billion to 10 billion yrs ago	K: >10 billion yrs ago
The Big Bang												
The Cold War (USA vs USSR)												
Beginning of Stone Age												
Controlled use of fire by hominids												
Formation of the solar system												
First modern nation-states appear												
Extinction of dinosaurs												
Invention of agriculture												
First mammals appear on earth												
Earliest empires appear												