JOURNAL OF GEOSCIENCE EDUCATION **60**, 393–407 (2012)

Utilizing Xenarthra (Tree Sloth, Anteater, Armadillo, Ground Sloth, Glyptodont, and Pampathere) Cranial Material to Evaluate Students' Understanding of This Thing Called Science

Barbara J. Shaw[1,2,a] and Luis A. Ruedas[2,3]

ABSTRACT

Two-thirds of U.S. citizens do not understand the scientific process. There is a clear misunderstanding about what science is—and is not—both in our society and in the classroom. Furthermore, students below basic proficiency are locked into an achievement gap. In response, the No Child Left Behind Act was passed in 2001. Since then, there has been some progress in decreasing the achievement gap. However, according to The Nation's Report Card, 34% of fourth grade and 43% of eighth grade students sampled by the National Assessment for Educational Progress still fall below a basic level of proficiency in science. To evaluate what is misunderstood about the scientific process, third through eighth graders were guided to discern science from pseudoscience, and form testable questions by using 45 animal skulls and design experiments, and to then collect and analyze data to answer their questions based on the graphs they developed. They were given a pre-assessment at the beginning and a postassessment the end of a 12-h unit to determine changes in learning. These data were analyzed by a paired Student's *t*-test. The results show that students gained significantly in memorizing facts and making objective observations about xenarthrans. Students were not able, however, to transfer the skills gained to make objective observations about dinosaurs. In addition, they had difficulty differentiating between scientific questions (objectively testable) from nonscience questions. © *2012 National Association of Geoscience Teachers*. [DOI: 10.5408/10-211.1]

Key words: inquiry, science education, testable question, pseudoscience

INTRODUCTION

Voters and politicians both rate education among the top 10 issues in the current sociopolitical situation of this country (Polling Report, 2007). Youth education is mandatory in all states, requiring attendance from ages 4 or 5 to usually at least age 16 years (U.S. Department of Education, 2007). It is a national goal for all children to obtain a specific level of understanding—or standard—in English, math skills, social science, and science, as expressed in and passed by the No Child Left Behind (NCLB) Act of 2001 (U.S. Department of Education, 2001) as well as in the science, technology, engineering, and mathematics (STEM) education initiatives of the National Science Foundation (NSF). The mission statement of the U.S. Department of Education encapsulates the importance of a solid education: The department's mission is "to promote student achievement and preparation for global competitiveness by fostering educational excellence and ensuring equal access." Scientists and educators are, however, failing at the basics in science: More than two-thirds of Americans do not understand science or the scientific process (NSF, 2004).

If education is the key to remain competitive in the global arena, then the United States is not meeting its stated objectives. The Trends in Mathematics and Science Study (TIMSS) assessment was developed by the International Association for the Evaluation of Educational Achievement (IEA) to measure students' achievements in mathematics and science. The Institute of Education Sciences of the U.S. Department of Education has a series of directives including participating in and maintaining the statistics of the TIMSS assessments (IEA, 2007). TIMSS provides participating countries with an unprecedented opportunity to evaluate students' progress in mathematics and science achievement on a regular 4-year cycle, which began in 1995, with the most current results being from 2007 (IEA, 2007). Through participation in TIMSS, the United States has obtained reliable and timely data on the mathematics and science achievement of U.S. students compared with those of students in other countries (Martin et al., 2004). One trend observed in U.S. science education is that as students progress through U.S. schools, their science scores are relatively highest in fourth grade (compared with those of other countries) and relatively lowest in 12th grade (IEA, 1995, 2007).

In each of the previous TIMSS (1995, 1999, 2003, and 2007), U.S. students in the fourth grade also improved in each assessment and statistically were fourth, after Singapore, Taiwan, Hong Kong, and Japan. The Russian Federation, Latvia, and England had higher mean scores, but were not significantly different from those of U.S. fourth graders.

The eighth grade U.S. students likewise improved significantly from 1997 to 2007 in both mathematics and science scores. However, by the 2007 TIMSS, 9 of 47 countries assessed countries had statistically higher mean scores in science than U.S. eighth graders. The top countries were Singapore, Taiwan, Japan, the Republic of Korea, England, Hungary, the Czech Republic, Slovenia, and the

Received 20 November 2009; revised 11 December 2011; accepted 14 July 2012; published online 6 November 2012.

[1]Colorado State University Extension, 1001 N. 2nd Street, Montrose, Colorado 81401, USA
[2]Department of Biology, Portland State University, Portland, Oregon 97201, USA
[3]Museum of Vertebrate Biology, Portland State University, Portland, Oregon 97201, USA
[a]Author to whom correspondence should be addressed. Electronic mail: Barbara.Shaw@colostate.edu

Russian Federation. The United States' and Hong Kong's scores were not significantly different, although Hong Kong's score was higher than the United States' score.

The highest assessment, TIMSS-A[dvanced], is given in the final year of secondary school students. This corresponds to 12th grade in the United States, although other countries have different number of required years of schooling. The United States participated in 1995, and the results were dismal: Of 21 countries, the U.S. ranked 16th in science and 19th in math, with an overall mean significantly below average. In 2008, TIMSS-A was offered again. The Bush Administration decided not to have U.S. high school seniors participate, reasoning that other countries test students older than those tested in the United States, and that many of these countries begin specializing in high school in different core areas such as physics or math (Mervis, 2007). Educators disagree, feeling that much can learned from these scores. The educational reforms over the past 6 years (2002, when NCLB went into effect, until 2008, when the TIMSS-A was administered to participating countries) include high school students taking more math and more advanced-placement science classes.

At this time, individual states have the responsibility to develop standards for measuring student learning in mathematics and science. A direct comparison among students from different states is therefore impossible. In order to evaluate state efforts, the National Assessment of Educational Progress (NAEP) was founded in 1969 as a part of the U.S. Department of Education, under the National Center for Education Statistics. The NAEP sampled students from across the nation, and correlated demographics together with these scores to formulate The Nation's Report Card, with 2005's being the most current data available. The findings are shocking: 34% of fourth and 43% of eighth graders were not meeting proficiency in basic science as of that date (NAEP, 2005a).

Furthermore, the goal of ensuring equal access to education to all U.S. citizens is not being met. In particular, there exists an achievement gap among particular groups, as measured by standardized test scores. Those students who are scoring below the basic level can be broadly identified as being from low-income families, from within certain ethnic or racial groups, and/or students with limited English-language proficiency (NAEP, 2005b). According to the NAEP, a much higher percentage of African-American, Latino, and Native American students do not attain the minimum standards in reading, math, and science as do Anglo-European Americans or Asian Americans (NAEP, 2005a).

It is one thing if this achievement gap were only an artifact of standardized tests. Certainly, part of the gap is built upon test author bias; however, measures are attempted to control for that factor (Secada, 1992). Two reasons are hypothesized for students falling into the achievement gap, a gap already apparent in early childhood (Chapin, 2006). First, students could hold a general disinterest in the formal structure of information being distributed in classroom settings (Conchas, 2001). A potential remedy would be for educators to find the means for students to discover the power and joy of learning; the scientific process lends itself to this particular paradigm of learning (Lowery and Mattaini, 1999; Somnath and Fraizier, 2008). Secondly, students might be unable to translate—or transform—what they learn in an unstructured classroom setting into the standardized test format (Jordan et al., 2000). A potential remedy for this failure might be to help students develop metacognitive and critical thinking skills, such that they are able to apply information from one area to distinct scenarios (Jordan et al., 2000), and for this, teachers need consistent and repetitive training in these areas (Abd-El-Khalick and Akerson, 2009).

Thus, it would appear that there is more to the failure of meeting the standards than merely students not doing well on standardized tests. U.S. students simply are not engaged in science because of a multitude of reasons—from language nuances or English proficiency, through perceiving science learning as "white," to even having teachers unfamiliar with science (Secada, 1992; Koba, 1996; Poliquin, 1997; Visone, 2010). This achievement gap is therefore clearly not just an artifact of standardized testing, a true gap exists (Olszewski-Kubilius, 2006). Not only is there a gap in achievement based on ethnic traits, but also a similar achievement gap occurs based on socioeconomic status (SES). Students from low-SES households score much lower than students from high-SES households on the NAEP assessment (average scale scores were 142 and 159, respectively).

To improve standardized scores, we must start closing the gap starting as early as preschool and kindergarten (Chapin, 2007; Johnston, 2009; Akerson and Donnelly, 2010). Our students need to engage in science activities in order to discern what science is and appreciate it in a nonjudgmental environment. However, as in any endeavor, if the objectives are not clear, then the outcome will not be clear (Chapin, 2006; Brown and Abell, 2007; Sarkar and Frazier, 2008). The objectives of the learning exercises must therefore be made clear to the students in order that learning be achieved, in explicit instruction (Khishfe and Lederman, 2007).

Regardless of the scientific discipline (physical, Earth/space, or life science) the commonality among disciplines is that science is trying to make sense of the natural world and natural processes by testing hypotheses (Chalmers, 2003). Scientific methods are dynamic processes used by all scientists in all sciences, wherein testable questions form the foundation of the work. Scientists and educators have, however, failed to teach the public that science is a process by which testable questions are answered through observations in order to elucidate the underlying natural mechanisms of the observation, and explanation of the results is interpreted by individuals and upheld by consensus (Schwartz and Lederman, 2008).

In light of the foregoing, the present study aimed to evaluate the hypotheses that students are able to discern between science and nonscience questions, and additionally are able to take a novel situation and apply prior experience and knowledge to accurately predict outcomes, that is to say, apply critical thinking. These two factors are the foundation of scientific inquiry, and if students are unable to discern a scientifically testable question from questions that employ opinion, evoke supernatural questions, or are anthropomorphic in nature, then U.S. students are failing to grasp the nature of science. By employing cultural myths and stories about anteaters, sloths, and armadillos (collectively known as xenarthrans in mammalian taxonomy) in a sensitive and respectful presentation, students critically examined the differences between science and nonscience as different

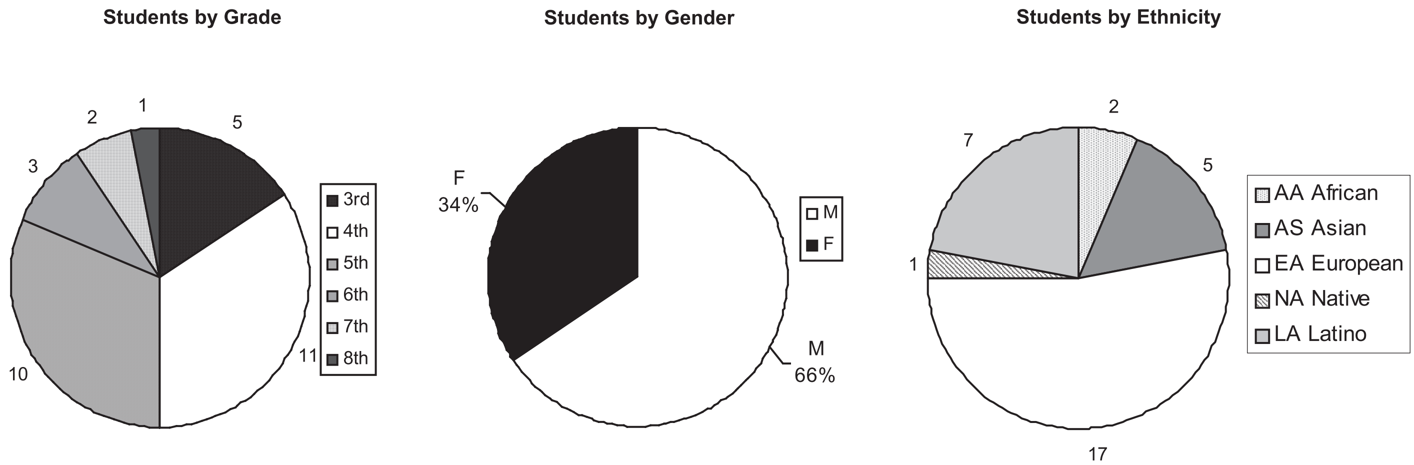


FIGURE 1: Demographics of students participating in the study. The students who participated in the course, but not in the study, are not included.

ways of learning and explaining in addition to the hands-on inquiry (Oliveira et al., 2012). The present study attempted to elucidate the core of the problem, Do students understand the foundation of science? If not, until that basic misconception is corrected, the United States will continue to fail children in their science education.

MATERIALS AND METHODS

The first author instructed all students. At the time of instruction, she was a Ph.D. candidate in biology, with her primary research in the systematic relationship xenarthrans through morphological and biomechanical analyses. In addition, she was a National Science Foundation fellow in the Center for Teaching and Learning in the West. During her training, she studied the philosophy and history of science as well as best practices for teaching science to diverse audiences. This experience and preparation in science circumvents the concerns of an inadequate foundation in the nature of science (Abd-El-Kahlick, 2000).

The mammalian superorder Xenarthra (extant species, i.e., anteaters, armadillos, tree sloths; and extinct species, i.e., glypdotons, pampatheres, giant ground sloths) is an excellent model for teaching the scientific process to upper elementary and middle school students. The species contained in the group show three broad distinct types of morphology contained within two taxonomic orders (taxonomic levels equivalent to, e.g., Primates or Rodents), together with a definite trend of increase then decrease in skull size through time. A total of 45 skulls and skull casts of 16 extinct and 9 extant xenarthran species, 1 species of monotreme, and 1 species of marsupial skulls were obtained (complete list of species are supplied in Appendix 1).

Fifteen 150-mm calipers and nine 600-mm calipers were provided. During the course of this study, minimal additional materials were purchased or supplied for student experimental design, depending on their questions. A student handbook (see http://www.sciencea2z.com/z_etomite/index.php?id=112 for supporting materials) was developed containing myths, legends, and stories about xenarthrans from South and Central America. Facts about the xenarthrans skulls contained in this unit, facts about the skulls of species outside the group, phylogenetic (evolutionary) trees, geological stratigraphy, directions for using calipers, and directions on how to produce histograms and scatter plots by hand and with Microsoft Excel. These notebooks were available to the students during the length of their classroom experience. The entire study was carried out with approval of Human Subjects Review no. 06,002 from Portland State University.

Students self-identified their gender, ethnicity, and grade level (Fig. 1). Additional demographics were collected on the students, specifically age and English proficiency (Fig. 1).

To evaluate the change in student learning, a pre-/postassessment was developed (Appendix 2). This assessment contained questions addressing science and nonscience questions, facts about the similarities and differences among the animals examined, observations about these animals, and observations about a dinosaur. Each of the questions was read aloud, with definitions given for any unfamiliar or unsure words in order to make sure that every student understood exactly what was being asked. This was particularly important for students correctly identifying whether a question was scientifically testable rather than trying to answer the question. Each one of these assessment sections was developed to anticipate directions students would take when engaged with the skulls, and to test the hypotheses about student conceptualization of how science works.

Three elementary schools and three middle schools in the Portland Public School District contracted with Saturday Academy, a nonprofit educational organization bringing professionals and students together, to present a Learning Enrichment Accelerated Program (LEAP) during the winter and spring terms of 2006. The school administrators selected the students to participate in this program. Each school had between 8 and 18 students in the before- or after-school program or during school in a pullout program, wherein the selected students left their classroom during regular school hours to participate in this class. In addition, 6 students attended a class at Portland State University. Altogether, 72 students participated in the program, 32 in grades three to five, and 40 in grades six to eight. Each program ran for a total of 12 h, (the standard Saturday Academy's LEAP program time). The administration for each school set the meeting parameters, and students met for 12 60-min sessions, 8 90-min sessions, or 6 120–min sessions. Financial scholarships were available; therefore, parents' income was not a limiting factor.

Of the 78 students, a total of 59 permission slips were returned from both parents and students. In order to include the student pre- and postassessment in this analysis, both parent and student participants had to provide signed permission slips. Of those 59 signed forms, 32 students completed both pre- and postassessment. These 32 students who completed both the pre- and postassessment, and returned a signed permission slip constitute the sample of the present study.

A free and reduced meal program (FRMP) is offered to students from low-income families, and therefore is one way to assess the SES of students at particular schools. Except for one school located in southeast Portland, all remaining Portland Public Schools were among those with the highest percentages of students participating in the FRMP, hence an indirect measure of the SES of schools participating in this study. The class held at Portland Public School District had enrollment open to all students in the Portland, OR, and Vancouver, WA, area. Saturday Academy freely offers Intel Scholarships for science classes, but scholarship information is not shared with the instructor; there is therefore no way to estimate SES for the six students who signed up for the Saturday class (Fig. 1).

Individual school administrators selected the time for this class (before, during, or after school). They also determined which students would be eligible for this program. Most of the students were selected from the Talented and Gifted (TAG) program at their school. The TAG status of students who took the class as traditional Saturday Academy program is unknown; however, these students had an interest in science, as they voluntarily attended on their weekend.

Students were given the pre-assessment on the first day of class. The postassessment was given during the last half hour of class on the last day. The difference between the beginning and ending scores was analyzed with a paired Student's t-test (Table I).

The first part of this course was guided inquiry, working through the complete scientific process (Table II). To introduce the skulls and allow students time to make observations, the first activity with them was simply determining some sort of criteria, and then to classify the skulls based on those criteria. The students were instructed to repeat the activity two more times with a different set of criteria, and then discuss the differences and similarities between how the skulls were assigned into groups. Students were taught to use Vernier calipers, and the difference between accurate and precise measurements. Students were instructed to measure the depth of their desktop, and results were compared for consistency. Each subsequent activity expanded on the activity just before it. When completed, the students were taking precise and accurate measurements, recording results, and building and reading graphs for answers to their questions. After the first series of exercises were completed, students were asked if they had any questions about the skulls, and what they would like to explore, based on the skulls. The younger students generated more questions, both testable and non-testable. The middle school students were more likely to wait for prompting. In the time remaining (about 5.5 h after the guided lessons), students developed their own testable questions, and designed and conducted experiments to answer their questions based on the data collected. Non-testable questions were discussed and eliminated. Graphs were built with Excel and/or graph paper, colored pencils, and rulers.

After all classes were completed, the pre- and post-assessments were matched by student and randomly assigned a number between 1 and 32. Both the pre- and postassessments were identified with the assigned student's number, and then the student's name was removed from the assessments to ensure the students' protection and privacy.

Assessments were shuffled and the first page with the identifying information (without a name) placed at the end. A rubric was used to score each assessment. The rubric was "0" for an incorrect answer and "1" for a correct answer, except for observations and facts assessment questions. To score these sections, if the student correctly stated a fact or facts, s/he received 1 point, regardless of the number of statements made. If the student made multiple statements, some true and some false, s/he received ½ point. If the student made incorrect statements (or did not answer), s/he received 0 points. Therefore, the student could receive a 0, ½, or 1, regardless of the number of facts they might have given in these sections.

RESULTS

Individual student net change in scores varied from −6.00 to +6.50. There were no significant trends in variation among students (age, gender, ethnicity, or race) who had a net loss or net gain from their pre- to postassessment (Figs. 2 and 3). Although not significant, the more hours students attended, the higher the results between pre- and post-assessment scores (average of 3.85 for 12 h and average of 2.79 for 11 h or less). To determine if the data were equal or unequal variance, the total data and the 20 testable pre- and postquestions were assessed with an F test [$F(31) = 0.7980$, $p = 0.2669$; $F(10)0.0825$, $p = 0.0003$; and $F(10) = 0.2034$, $p = 0.0110$].

Overall scores between the pre-assessment ($M = 30.79$) and postassessment ($M = 27.50$) showed a significant gain [$t(23) = 2.46$, $p = 0.0219$] (Figs. 4 and 5). However, not all questions showed a significant gain, and many questions did not show a positive change (Fig. 6). The total science/nonscience questions 1–20 (pre-$M = 6.05$, post-$M = 7.15$) and the "dinosaur" question 22 (pre-$M = 1.28$, post-$M = 1.41$) were not significantly different [$t(37) = -0.5013$, $p = 0.6191$ and $t(31) = -0.8916$, $p = 0.3795$, respectively] between the pre- and postassessment score. Questions that showed a significant gain between the pre- and postassessment scores were no. 21 ("As a scientist, describe the picture of this mammal, a giant armadillo." Pre-$M = 1.28$, post-$M = 1.75$, $t(31) = -3.1499$, $p = 0.0036$], no. 23 ["Please tell me something you know about tree sloths, anteaters, armadillos, ground sloths, glyptodonts, and/or pampatheres." Pre-$M = 1.16$, post-$M = 1.72$, $t(31) = -3.0440$, $p = 0.0047$], no. 24 ["What have you noticed about tree sloths, anteaters, armadillos that are *the same* as ground sloths, glyptodonts, and pampatheres?" Pre-$M = 1.28$, post-$M = 1.75$, $t(31) = -3.1499$, $p = 0.0036$], and no. 25 ["What have you noticed about tree sloths, anteaters, armadillos that are *different* than

TABLE I: The total raw pre- and postassessment scores and percentage out of the total points possible, the difference of the preassessment subtracted from the postassessment, the percentage of each of these score, and the significance.[1]

Question	Possible	Pre-assessment	Postassessment	Difference	Pre-assessment (%)	Postassessment (%)	Difference (%)	p value
1	32	29	31	2	90.62	96.88	6.26	0.3251
2	32	32	29	−3	100	90.62	−9.38	0.0831
3	32	17	15	−2	53.12	46.88	−6.24	0.4882
4	32	19	24	5	59.38	75	15.62	0.096
5	32	13	16	3	40.62	50	9.38	0.414
6	32	26	31	5	81.25	96.88	15.63	0.0574
7	32	31	32	1	96.88	100	3.12	0.3251
8	32	4	9	5	12.5	28.13	15.63	0.0228*[2]
9	32	21	20	−1	65.63	62.5	−3.13	0.7864
10	32	29	30	1	90.62	93.75	3.13	0.572
11	32	24	22	−2	75	68.75	−6.25	0.4882
12	32	29	31	2	90.62	96.88	6.26	0.3251
13	32	27	27	0	84.37	84.37	0	1
14	32	29	31	2	90.62	96.88	6.26	0.3251
15	32	30	30	0	93.75	93.75	0	1
16	32	29	29	0	90.62	90.62	0	1
17	32	31	30	−1	96.88	93.75	−3.13	0.572
18	32	18	23	5	56.25	71.87	15.62	0.1338
19	32	29	28	−1	90.62	87.5	−3.12	0.572
20	32	30	31	1	93.75	96.88	3.13	0.572
1–20	640	497	519	22	77.66	81.09	3.44	
21	32	20.5	28	7.5	64.06	87.5	23.44	0.0036*
22	32	20.5	22.5	2	64.06	70.31	6.25	0.3795
23	32	19	27.5	8.5	59.38	85.97	26.59	0.0047*
24	32	14	20	6	43.75	62.5	18.75	0.0499*
25	32	10.5	20.5	10	32.81	64.06	31.25	0.0046*
Total	800	581.5	637.5	56	37.97	41.65	3.68	

[1]Note that the science/nonscience questions are subtotaled.
[2]*, significant difference.

ground sloths, glyptodonts, and pampatheres?'' Pre-M = 0.88, post-M = 1.25, $t(31) = -2.0406$, $p = 0.0499$]. Question 8 ("Can tarot cards tell the future?'') incorporated a supernatural topic; however, the question itself is testable. This question, therefore, afforded students an opportunity to examine what makes a testable question. They correctly identified it as scientifically testable [pre-M = 0.13, post-M = 0.29, $t(30) = -2.4019$, $P = 0.0227$]. The students, who correctly answered the question, ranged in ages from 9 to 12 years and came from different schools. Every student who did answer this question correctly attended at least 10.5 h or more of the course; however, not all students who attended at least 10.5 h responded correctly.

The questions can be grouped into four subcategories: five science questions about xenarthran animals, six science questions on other science topics, five nonscience questions about xenarthrans, and four nonscience questions on other topics (Fig. 7). The individual questions were not significant, with the exception of no. 8; however, each of the grouped subcategories showed significant gains [pre-M = 1.6, post-M = 30.4, $t(8) = -50.91$, $p = 2.45 \times 10^{-11}$; pre-M = 5.2, post-M = 26.8, $t(10) = -4.2665$, $p = 0.0017$; pre-M = 11.6, post-M = 20.4, $t(8) = -2.6155$, $p = 0.0309$, and pre-M = 6, post-M = 26, $t(6) = 0.00018$, $p = 0.0004$, respectively] (Fig. 8).

SES is protected information, and accordingly, the only means to evaluate the impact of this study on students living in poverty is to estimate based on the percentage of students at the school participating in the FRMP. Administrators selected students primarily on account of their prior participation in the TAG program, rather than in proportion to the demographics of the school. If students who do better in school come from households with a higher SES than do students from households with a lower SES, then the students in this study do not reflect the school proportion of students in the FRMP. However, the schools were home to some of the largest percentages of students enrolled in the FRMP in Portland Public School District. Students from schools with more than 70% of the student body partici-

TABLE II: Schedule of course and approximate duration for each activity.

Activity	Description	Approx time (h)
Pre-assessment	Baseline answers	0.50
Xenarthran myths	Read myths, legends, and stories about xenarthrans, and discuss the difference between natural and supernatural using these stories as examples	0.50
Skull sorting	As an introduction to the skulls, students determined their own criteria for sorting the skulls into groups, and then asked to use a different criteria, and resort	1.00
Caliper instruction	Students are taught to take accurate and precise measurements using Vernier calipers and practice measuring the depth of the tabletop and verify for consistency with classmate measurements	0.50
Variation in a population	Using the 15 nine-banded armadillo skulls (*Dasypus novemcinctus*), students measure the length of the skull with the Vernier calipers and analyze by building an histogram to evaluate the size distribution in a bell curve, as well as establishing a control	1.00
How big was the skull?	Students are guided through the scientific process when given this scenario: "This tooth belongs to an extinct ground sloth. I found it in Patagonia, Argentina. I looked, but I didn't find any other part of this fossil, only the tooth. I measured this tooth, and it 23.70 mm long. Can we figure out how big its skull was?"	2.00
Scientific testable questions	Throughout the first 5 h of instruction, any questions the students asked were recorded. At this point, we discussed what questions are scientifically testable, what questions are not scientifically testable.	0.50
Student-designed experiments	The remaining time give for this project was devoted to helping the students design, collect data, and evaluate their own questions. The younger students took longer working through the guided inquiry above, and therefore were able to complete only one or two additional experiments. Middle school students completed three or four different inquiries. Here are some of the inquiries (and generally, most fell into these kinds of questions: "Were any of these animals big enough to squish a cantaloupe if they stood on it?" (This was a 3rd grader question, and when the middle school students struggled with their first experimental design, I told them about this particular question. It was replicated by almost every group.) "Which animal had the strongest bite?" "How does my bite compare to the xenarthrans?" "Are anteaters more closely related to the armored forms (armadillo, glyptodont, and pampathere) or hairy forms (tree sloth and ground sloth)?"	~ 5.5
Postassessment	Final answers to compare with baseline	0.50

pating in the FRMP showed a positive change between their pre- and postassessment scores (average of 4.29) greater than those students participating from schools with lower FRMP participation (average 2.93).

DISCUSSION

School administrators selected students for this class. Most were identified as TAG students, although one school allowed anyone interested to attend with the TAG students. It is noteworthy that administrators selected boys to girls in a 2:1 ratio, except at the school with open enrollment. The ratio at this school still had boys outnumbering girls by 5:4. By selecting students in the TAG program, the subjects in this study are more likely to be obtaining proficiency in their NAEP scores than students who are not enrolled in the TAG program, rather than a representative sampling of all students at a particular school.

The opening activity was reading myths and stories of xenarthrans, and discussing the difference between science and nonscience. In one story, an armadillo curls into a ball, rolls down a hill, strikes a tree, and breaks into nine bands. This particular story was used to discuss fact within fable. Of the 21 extant species of armadillos, only the 2 species of three-banded armadillos are capable of rolling into a ball; nine-banded armadillos cannot. Other stories attributed anthropomorphic characters to the animals, and we discussed how science cannot attribute human emotion. We dissected each story and identified the parts of the story that were outside the realm of science. This allowed time to discuss the narrow definition of what science is: the study of the natural world through a systematic approach to elucidate natural mechanisms behind our observations. After the discussion, students sorted the skulls based on characteristics that they chose. After the initial sort (usually on skull size), students sorted on a second set of criteria, and then one more time, on a third set of criteria. Next, students were taught how to use calipers, measured 15 nine-banded armadillo skulls, and assessed the measurable variation in the nine-banded armadillo crania by building a histogram, thus establishing a control for their future questions. The next project, estimating the length of an unknown skull based on the size of a tooth, was presented to introduce the scientific process of answering a testable question by a carefully designed experiment, analyzing and graphing the data, and interpreting the results from graphs to answer the question. From the beginning exercise of sorting the skulls, it was apparent that most of the students in the group did not understand how to read a graph of collected data, although they could. After measuring all the skulls to determine the approximate size of the "missing" skull, students guessed as to its length, even though we carefully discussed and built a scatter plot graph. Regardless of students' grade, they did not consider that the graph contained the answer. When shown how to find the answer within the graph, most students were surprised. Oregon Mathematics Standards

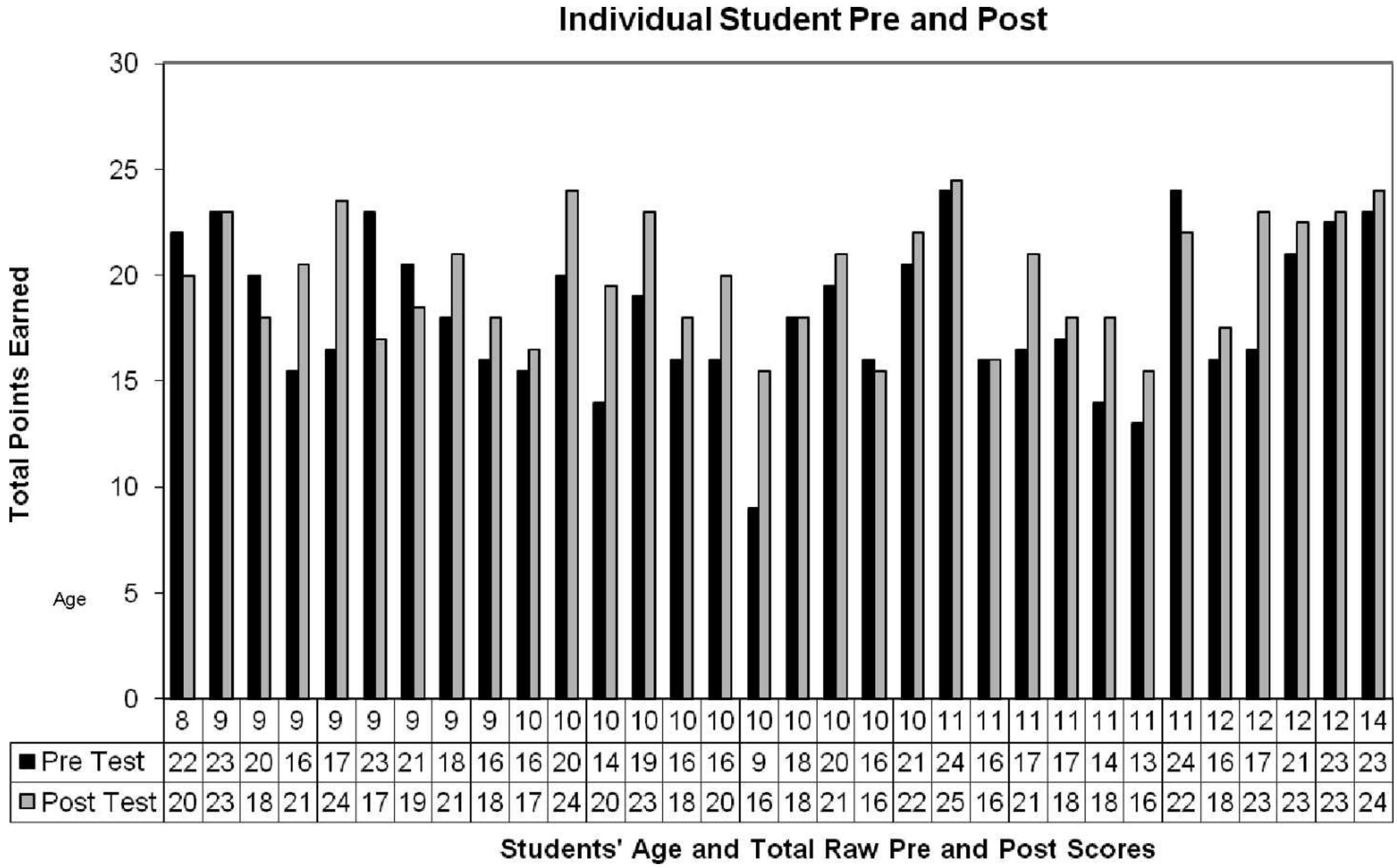


FIGURE 2: Students' individual pre- and postassessment scores organized by youngest to oldest.

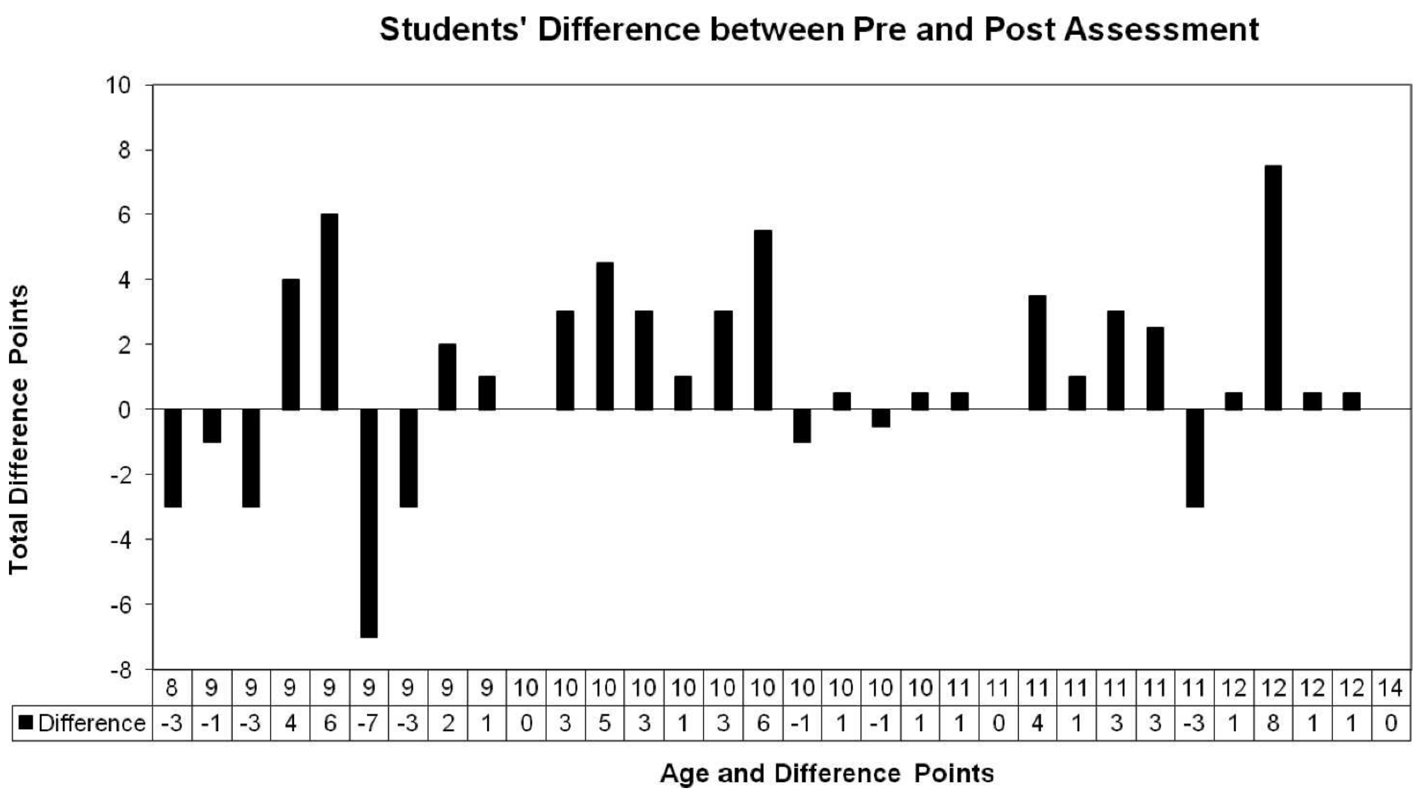


FIGURE 3: Students net change, the difference between pre- and postassessment, organized by youngest to oldest.

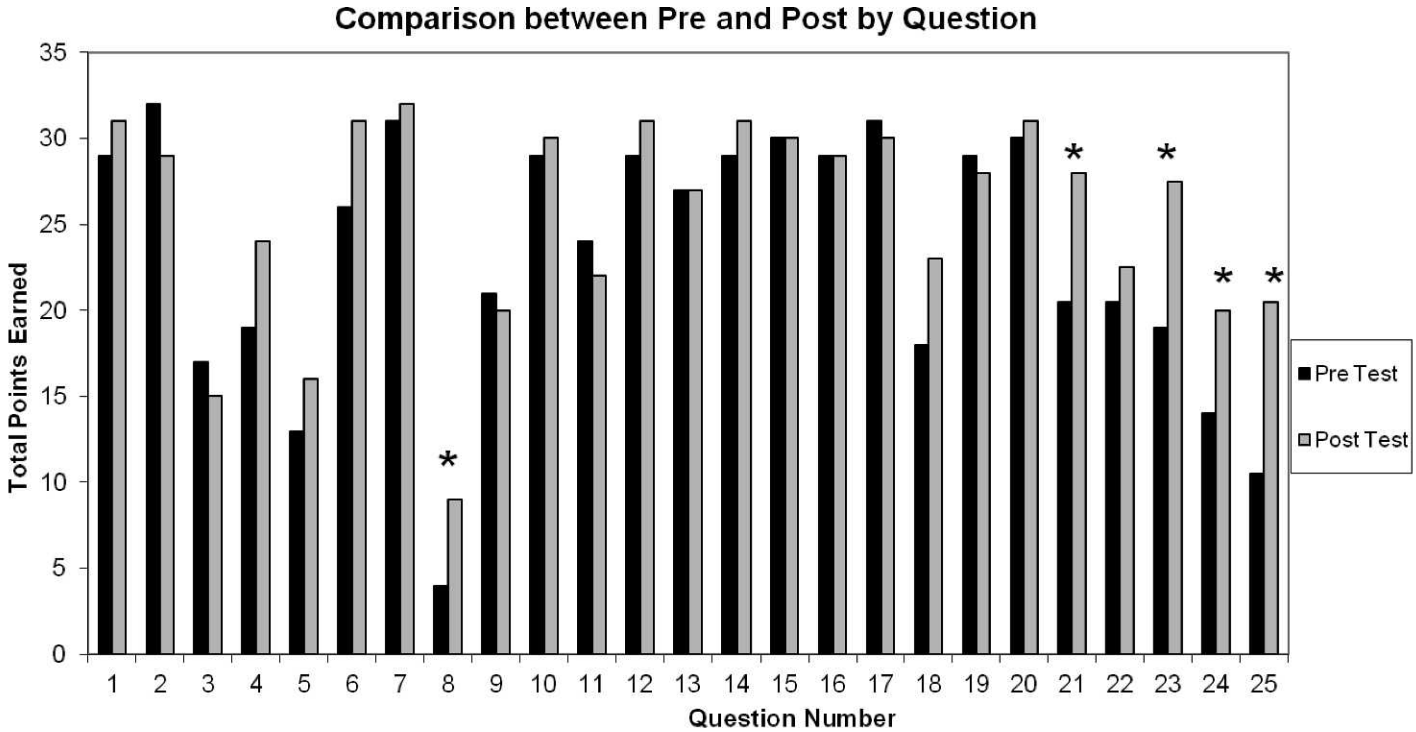
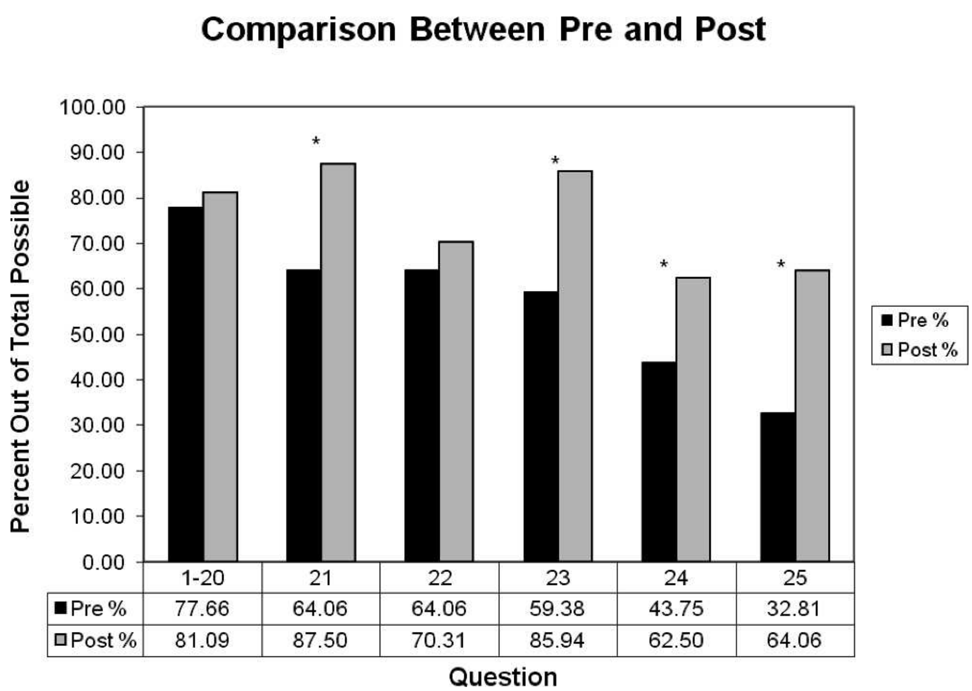


FIGURE 4: Pre- and postassessment scores for each of the questions. An * indicates the gain was significant: question 8, $p = 0.0228$; question 21, $p = 0.0036$; question 23, $p = 0.0047$; question 24, $p = 0.0499$; and question 25, $p = 0.0046$.

include building and reading scatter plot in the eighth grade; yet, according to Oregon's Common Core State Standards for Mathematics, a third grade mathematics standard includes representing and interpreting data. If students from third to eighth grade are working on representing and interpreting data, then it can be expected that they will look at the data for answers to questions rather than guessing.

In additional experiments, most students did not at first refer to their graphs when answering their questions—until prompted to do so. Eventually, students began to look at the



	1-20	21	22	23	24	25
Pre %	77.66	64.06	64.06	59.38	43.75	32.81
Post %	81.09	87.50	70.31	85.94	62.50	64.06

FIGURE 5: Pre- and postassessment scores for each of the questions. The science/nonscience questions (questions 1–20) were analyzed together. An * indicates the gain was significant: questions 1–20, $p = 0.2614$; question 21, $p = 0.0036$; question 22, $p = 0.3795$; question 23, $p = 0.0047$; question 24, $p = 0.0499$, and question 25, $p = 0.0046$.

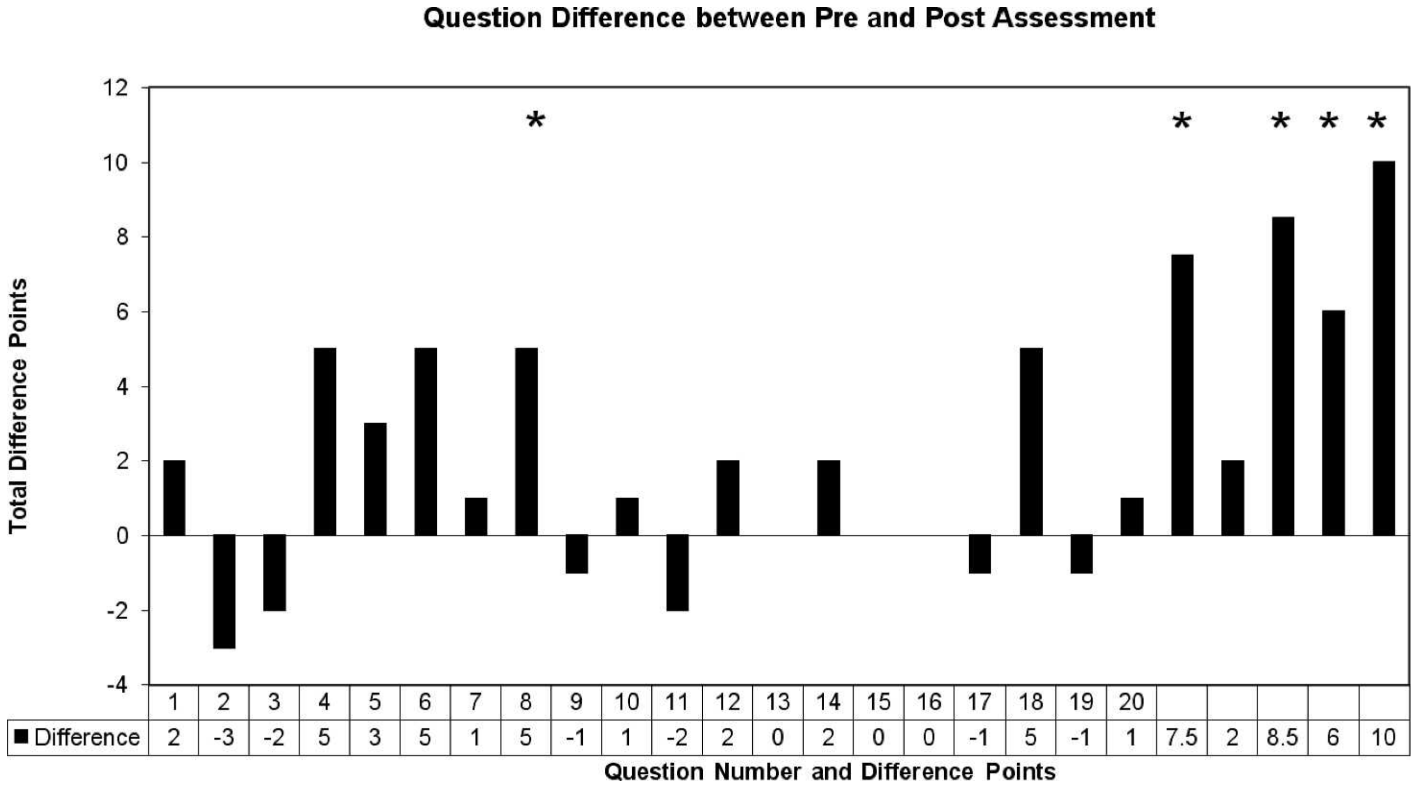


FIGURE 6: Questions net change, the difference between pre- and postassessment, organized by order of questions on the assessment.

graphs for answers, but even for the eighth graders, that was not until after the third or fourth experiment. This skill required time to develop, but even the third grade students learned how to interpret their graph to conclude the results derived from their data, as the standards indicate.

After leading the students though the first four activities (reading myths, sorting skulls, developing a histogram on cranial variation, and inferring skull size based on a single tooth), the students developed a list of questions. Students in elementary school were eager to ask questions, and it usually became a competition as to who could ask the most interesting question. Middle school students in contrast were generally more reticent, and required much prompting in order to volunteer a question. It was apparent during these discussions that students struggled to understand what constitutes a scientifically testable question. For example, some students had a problem discerning opinion from fact. After the pre-assessment, one group of students discussed how to collect data on the question "Are ghosts real?" to determine what makes a question scientifically testable, we used:

1. Falsifiability (Popper, 1959): The hypothesis is the best guess to the question, and it can be proven as wrong.
2. Replication: We can repeat the experiment, and the results will be similar to the original experiment.
3. Definition: We used a narrow definition of science: discovery of the natural world through natural mechanisms; supernatural mechanisms cannot be used.
4. Objectivity: We discussed employing a systematic approach to answering "No," to the question "Is it personal and subjective?"
5. Anthropomorphism: During the reading of the xenarthran stories, we discussed attributing anthropomorphic qualities to animals in stories. These qualities cannot be tested in a scientific study, because we cannot enter the mind of an animal to verify.

Throughout the class, students struggled with the idea of a testable question, although they had a firmer understanding of what is science and what is not (Table III). Even though both elementary and middle school students had difficulty discerning the difference between science and nonscience questions, over the 12 h, they improved in designing and determining what is testable. Of the individual questions, only one non-xenarthran, science question was significant (question 8, "Can tarot cards tell the future?"). Students could reject it, because it was examining something from the supernatural realm (tarot cards); however, the question itself is scientifically testable. Evaluating if age is a factor in understanding what is a testable question, elementary students (grades three to five) and middle school students (grades six to eight) were analyzed separately, but age did not obtain as a significant factor.

Sorting the data by age, school, grade, gender, and other factors did not reveal any trends to our results. Figure 3 reveals that the younger students had more difficulty, but the results are not significant.

The boundary between science and pseudoscience is fuzzy, so what is the nature of science, and how does

5 Questions Xenarthran Science P=2.454x10^{-11}	5 Questions Xenarthran Non-science P=0.0309
6 Questions Other Topics Science P=0.0017	4 Questions Other Topics Non-science P=0.0004

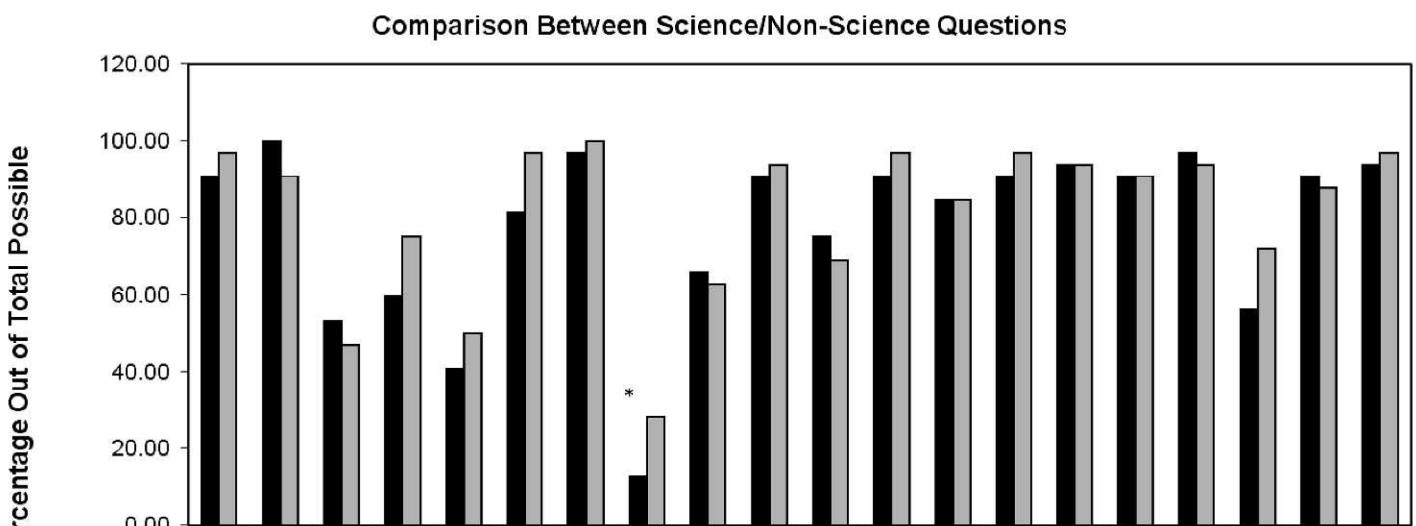
FIGURE 7: Likewise comparison of the 20 science/nonscience questions and their *p* values of students responding correctly at the end of the 12 h.

pseudoscience violate that nature? That is a challenge, because scientific philosophers debate ideas, but illuminating exactly what constitutes pseudoscience remains elusive (Lakatos, 1977). Science has the hallmarks of accurately predicting outcomes based on a series of known facts. Answers can be derived through testing, and the results are replicable; yet astrology makes those same claims. The difference is the quality of the prediction. Astrology's predictions are vague, and can be applied to many circumstances. In contrast, science is incredibly accurate. In 1924, Satyendra Nath Bose sent a paper predicting a new state of matter as atoms approach absolute zero to Albert Einstein, who translated it and had it published in *Zeitschrift für Physik* (Bose, 1924). Yet it wasn't until 1995, over 70 years later, when Eric Cornell and Carl Wiener cooled matter to a fraction of a degree to absolute zero, that the Bose–Einstein condensate finally was observed: That is specific prediction.

During the class, each group discussed what is science, and how can we test a question. Each of the questions the students asked was carefully dissected to understand if it was testable. Hints, such as the use of the words "how many" and "how long" as opposed to the word "why," give a strong indication if the question is testable, as well as applying our criteria for a scientifically testable question. We would then discuss how the experiment could be designed that would collect data to answer that question. Each group discussed reasons for dealing with the supernatural, i.e., whether trying to test for it or invoking it as an answer, is not testable.

Question 6 of the assessment was designed specifically to ask the same question as the second planned experiment, "By measuring the length of the last molar, can we predict how big a giant ground sloth was?" Using a *p* value of 0.05, this question neared statistical significance ($p = 0.0574$, and it could be argued that it is significant), but even after conducting the experiment, not all students correctly identified it as scientifically testable.

During the class discussions, students indicated that they could discern what was, and was not science. The answer for this could be simply that the students did not



	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
■ Pre %	90.6	100.	53.1	59.3	40.6	81.2	96.8	12.5	65.6	90.6	75.0	90.6	84.3	90.6	93.7	90.6	96.8	56.2	90.6	93.7
□ Post %	96.8	90.6	46.8	75.0	50.0	96.8	100.	28.1	62.5	93.7	68.7	96.8	84.3	96.8	93.7	90.6	93.7	71.8	87.5	96.8

FIGURE 8: Pre- and postassessment scores for each of the 20 science/nonscience questions. Xenarthra science questions are nos. 2, 6, 12, 17, and 20, and the other science questions are nos. 7, 8, 10, 14, 15, and 16. Xenarthran nonscience questions are 3, 5, 9, 18, and 19, and the other nonscience questions are nos. 1, 4, 11, and 13. An * indicates the gain was significant.

TABLE III: Science/nonscience question analysis based on the five criteria established during the class.[1]

Question	Criterion/a met	T/F
Are rabbits cute?	Subjective	F
Are the scratch patterns on glyptodon teeth more similar to the scratches on herbivore teeth or carnivore teeth?	Meets all criteria	T
Are armadillos empathetic?	Anthropomorphic	F
Are ghosts real?	Supernatural	F
Are sloths slow because they are deep thinkers?	Anthropomorphic	F
By measuring the length of the last molar, can we predict how big a giant ground sloth was?	Meets all criteria	T
Moles live underground. How much oxygen is present in their tunnels?	Meets all criteria	T
Can tarot cards tell the future?	The content is nonscientific, but the question itself is eminently testable.	T
Are sloths slow because they are lazy?	Anthropomorphic	F
Hyenas have the strongest bite force of any mammal. Do hyenas have more jaw muscles than other mammals?	Meets all criteria	T
Are people born under the sign of Leo the Lion courageous?	Supernatural	F
Did the attachment of the muscle on the jaw of tree sloths and ground sloths change even though they are/were herbivores?	Meets all criteria	T
Do skunks stink because they have bad spirits inside of them?	Supernatural	F
What is the average temperature in Portland, Oregon?	Meets all criteria	T
At what temperature do people begin to shiver to stay warm?	Meets all criteria	T
Do plants need oxygen to stay alive?	Meets all criteria	T
Sloths, anteaters and armadillos all have long claws. Which group has the longest claws proportional to their body size?	Meets all criteria	T
Can armadillos understand your vulnerability because they have shells?	Anthropomorphic	F
Did anteaters lose all their teeth as a punishment because they wanted to eat little animals?	Anthropomorphic	F
Can we analyze how armadillos and glyptodonts chewed by measuring the place on the jaw where the muscles attached?	Meets all criteria	T

[1]Criteria: (1) falsification, can be proven wrong; (2) replication, can be repeated with similar results; (3) definition, must be elucidating a natural mechanism in the real world; (4) objectivity, based on observable phenomena and not emotion or personal opinion; and (5) anthropomorphism, not interpreted with human emotion, judgment, or feelings.

apply what they learned in the course to a novel scenario, i.e., the questions as posed in the postassessment. Perhaps this approach just needed more time than the allotted 12 h for students to clarify in their minds what is and is not science, as were applied in the 20 questions.

Questions 21 and 22 related to identical concepts. (21, "As a scientist, describe the picture of this mammal, a giant armadillo," and 22, "You are the lead paleontologist at this dig site. Area A was dry land. Area B was shallow water. The location of the fossil animal and the fossil eggs were in the same layer. An analysis of the eggs indicates that they are the same species as the fossil animal. As a scientist, describe what you have found.") In both cases, the students needed to make objective observations about the giant armadillo picture (large claws and scratch marks in the soil), and most did. The drawing of a dinosaur dig (dinosaur and egg fossils were located on an island); however, students either left the question blank, or they made stories about the eggs and dinosaur.

Students gained significantly on the giant armadillo question [$t(31) = -3.1499$, $p = 0.0036$], but did not on the dinosaur question [$t(31) = -0.8916$, $p = 0.3795$]. These two questions were addressing if the students were able to conceptually take information learned from one source (make an observation) and apply it to another source. They were developmentally capable of making these connections, but did not. There is not much research available on students transforming concepts from one scenario and applying it to another, similar scenario. In elementary school, students spend very little time conducting science experiments and learning science. According to a survey conducted by the Maine Education Policy Research Institute, 0.3%–23% of classroom time is spent on science (Poliquin, 1997). Science, however, is taught in isolated packages (for example, the Full Option Science System, a highly acclaimed hands-on science curriculum, teaches individual units) rather than approaching science in a more integrative manner. From middle school, students are taught in discrete units: geometry, physics, English, German, language, art, etc. This downplays the interconnection among all knowledge, especially those among the different branches of science. This underscores the fact that there needs to be more research on how students learn science (Slutskin, 2003), although there is a direct correlation between the amount of time spent on a single subject and making significant academic gain in that subject (Raizen et al., 1985).

In addition, we found that students are not applying knowledge learned in one disciplinary area to other areas—that is, somehow, knowledge is compartmentalized rather than integrated. This has important implications. Even students engaged in comprehensive science curricula are not necessarily guaranteed to perform adequately in standardized tests. Instead, there is a disconnection between what they learn in a specific topic and the ability to translate and integrate that information into other fields of knowledge. For example, students measured the skulls of 15 nine-banded armadillo specimens. We discussed how to display the data and how to make a histogram. The students generated a histogram; some students had access to computers and used Excel to do so. When asked, "What was the variation in the skulls?" they did not look to the graph and table to answer their question until directed to do so. Instead, they answered with their preconceived notions of what they thought the answer would be. They had to be led through the same process in the course of the next experiment they conducted. In that subsequent experiment, students were asked to determine the approximate length of a ground sloth skull, based on a single tooth. The students discussed how to accomplish this task, measured all the teeth of the specimens available, but they did not know how to proceed thence. This was true for all students, regardless of age. After discussing how scientists display their data in graphs and tables, the students were led to developing a scatter plot graph. At this point, only one group of students was able to successfully use their resulting graph to determine the approximate size of the "unknown" skull. The other groups made guesses, or "knew" the answer. After showing them how to read the scatter plot they had generated, students demonstrated that they could in fact estimate skull length, based on their data, and were more likely to look at the graph for their answer only when they developed a question similar to the missing skull question (i.e., "Can we determine the size of the missing glyptodon skull from a tooth?"). Eventually, after conducting four or five experiments and analyzing the results with graphs, students started to look to their data to answer their question. Even the third graders were equally adept at reading a graph, and eventually looked at graphs to find their answer.

Students in the public education system are taught disjunctive facts. They become quite adept at learning and reciting facts, and almost certainly use this as a strategy for doing well on the standardized tests. The xenarthran curriculum was specifically taught as a project-based inquiry unit rather than fact-based curriculum. The facts supplied in this unit were specific to the questions being asked. No additional facts were given outside those explicitly needed for the students to conduct their experiment. Questions 23, 24, and 25 were specifically designed to assess how the students gain factual information, even from a program designed for project-based inquiry. Question 23 asked for any fact. Question 24 asked the students to identify a fact shared by at least two groups of xenarthrans. Question 25 asked for a fact shared by the extant group of xenarthrans (tree sloths, anteaters, and armadillos) and not shared by the extinct group (ground sloths, glyptodonts, and pampatheres). In all three questions, students gained significantly [$t(31) = -3.0440$, $p = 0.0047$; $t(31) = -3.1499$, $p = 0.0036$); and $t(31) = -2.0406$, $p = 0.0499$, respectively]. Indeed, of the major areas of the assessment, "Facts" was the only section that was significant between pre- and postassessment, and the principal reason that the overall assessment showed a significant gain.

Between these two sections of the assessment, transforming information learned to another area, and reciting facts, it is evident that U.S. K–8 students are not given enough time to think critically, a condition necessary to the undertaking of science. Higher education academics expect students to be prepared to think critically. It is apparent that in elementary and middle schools, students are not yet taught to apply critical thought to problems. If K–12 teachers wait until high school to begin this training, it is too late. Students need to learn and practice these techniques from elementary grades. All students demonstrated that they were capable of thinking critically when lead through the process, even the third graders.

CONCLUSION

The students in this study represent only a small subset of the number of students who participated in the class. Our results with these TAG students emphasized that there are several important and independent concepts at work in the preparation students for college while helping them survive the standardized testing they must endure.

There were no differences in results based on gender, ethnicity, grade, or age of the students. However, students from schools with more than 75% of the student body participating in the FRMP had a much higher average (4.29) when compared with the average (2.93) of students from schools with lower numbers of students participating in that program. There is, however, no way to determine what percentage of the students in this study were part of the FRMP, or whether it is proportional to the overall school percentages.

Students struggled to identify what is science and what is not science through selecting questions that could be scientifically tested, individually, but when analyzed by grouping the questions according to the type of question, the results reviled that the students did, indeed, show statistically significant results. Younger students are more comfortable asking questions, regardless of whether they are asking the so-called right questions. This allowed for lively discussions of what constituted a testable question, and what constitutes science. Middle school students were, in contrast, more reticent to volunteer questions. It was therefore more difficult to work though the specifics of what is testable and what is science. Since younger students are more eager to question, it seems a perfect platform for them to practice asking questions, and determining if they are or are not testable questions.

Students did not immediately understand that the answers to their scientifically testable questions are answered by means of the data that they collected. Each new experiment generated new data. These data were developed into graphs and tables. The students' initial question was restated, and at first they had to be led through by using the data to answer it, although in time, students began to look at their data to answer the underlying question. The 12 h they spent in this class were vastly different from the "science" they carry out in their classrooms. Without experience and practice, these results are not surprising.

In the current U.S. educational framework, students are taught isolated facts in order to prepare for their high-stakes standardized tests, rather than the idea that concepts in science can be applied to many distinct questions and areas of science. Students therefore perceive science as discrete units of information, and do not understand that each branch of science is interconnected to all other branches of science. They therefore must rely on, and continue to learn, facts in order to perform adequately on the assessment tests currently administered in the educational system.

If students are to learn the concepts necessary for performance in science, then teachers must be trained to teach science effectively. The amount of time spent on science varies greatly with each teacher, yet it is clearly established that greater exposure to a subject will favorably help the student in testing outcomes.

Standardized tests unfortunately will not go away. They are clearly not adequate to evaluate student learning, and no standardized test can really assess what a student knows. Instead, they are good at assessing how many facts a student has memorized. It cannot assess students' understanding of specific concepts. Some of the questions attempt to determine conceptualization, but once again, if the fact has been memorized, the student will correctly guess the answer.

What is the goal of education? If it is a solid base of facts, the current system can remain in place. If, however, the goal of education is to teach students to *think,* then educational goals are not being achieved. If our goal is to impart the logos that science is a process based on testable questions, we are not succeeding. We need to begin training elementary teachers to be comfortable teaching science, understanding the concept of testable questions, and providing opportunities for students to engage in the inquiry-based aspects that make science what it is today: a solidly constructed framework built on an ever-increasing structure of questions and answers. We need our teachers to help students connect between the individual examples they conduct as science inquiry with the underlying concept, and, even more importantly, apply that understanding to similar novel situations.

Acknowledgments

The authors would like to thank Gail Pyle at Saturday Academy. Without her help, this project could not have been done. We also would like to thank Dr. Michael Cummings for his constant support.

REFERENCES

Abd-El-Khalick, F. 2000. Improving science teachers' conceptions of nature of science: a critical review of the literature. *International Journal of Science Education,* 22:665–701.

Abd-El-Khalick, F., and Akerson, V. 2009. The influence of metacognitive training on preservice elementary teachers' conceptions of nature of science. *International Journal of Science Education,* 31:2161–2184.

Akerson, V., and Donnelly, L.A. 2010. Teaching nature of science to K-2 students: What understands can they attain.? *International Journal of Science Education,* 32:97–124.

Bose, S.N. 1924. Plancks Gesetz und Lichtquantenhypothese. *Zeitschrift für Physik,* 26:178–181.

Brown, P.L., and Abell, K. 2007. Project-based science. *Science and Children,* 45:60–61.

Chalmers, A.F. 1999. *What* is this thing called science? Indianapolis, IN: Hackett Publishing Company.

Chapin, J.R. 2006. The achievement gap in social studies and science starts early: evidence from the Early Childhood Longitudinal Study. *The Social Studies (Washington, DC),* 97:231–238.

Conchas, G.Q. 2001. Structuring failure and success: understanding the variability in Latino school engagement. *Harvard Educational Review,* 71:475–504.

International Association for the Evaluation of Educational Achievement. 1995. Trends in Mathematics and Science Study-Advanced (TIMSS). Available at http://timss.bc.edu/timss1995i/TIMSSPDF/C_Hilite.pdf (accessed 11 December 2007).

International Association for the Evaluation of Educational Achievement. 2007. Trends in Mathematics and Science Study (TIMSS). Available at http://nces.ed.gov/timss/results07_science07.asp (accessed 11 December 2007).

Johnston, J.S. 2009. What does the skill of observation look like in young children? *International Journal of Science Education,* 31:2511–2525.

Jordan, D.L., Henry, M.A., and Sutton, J.T., eds. 2000. Changing Omaha classrooms: Collaborative action research efforts studies of twelve teacher-research projects. Aurora, CO: McREL, p. 64–75.

Khishfe, R., and Lederman, N. 2007. Relationship between instructional context and views of nature of science. *International Journal of Science Education,* 29:939–961.

Koba, S.B. 1996. Narrowing the achievement gap in science. *Educational Leadership,* 53:14–17.

Lakatos, I. 1977. Philosophical papers, vol 1. Cambridge, UK: Cambridge University Press, p. 20–26.

Lowery, C.T., and Mattaini, M.A. 1999. The science of sharing power: Native American thought and behavior analysis. *Behavior and Social Issues,* 9:3–23.

Martin, M.O., Mullis, I.V.S., Gonzalez, E.J., and Chrostowski, S.J. 2004. TIMSS 2003 International Science Report: Findings from IEA's trends in international mathematics and science study at the fourth and eighth grades. Chestnut Hill, MA: TIMSS & PIRLS International Study Center, Boston College.

Mervis, J. 2007. U.S. says no to next global test of advanced math, science students. *Science,* 317:1851.

National Assessment of Educational Progress. 2005a. Executive summary: Science results for grades 4, 8, and 12. Available at http://nationsreportcard.gov/science_2005/s0101.asp (accessed 11 December 2007).

National Assessment of Educational Progress. 2005b. Investigating the potential effect of exclusion rates on assessment results. Available at http://nces.ed.gov/nationsreportcard/about/2005_effect_exclusion.asp (accessed 11 December 2007).

National Science Foundation. 2004. Science and engineering indicators: Public attitudes and understanding of science and technology. Available at http://www.nsf.gov/statistics/ (accessed 18 August 2005).

Oliveira, A., Wilcox, K., Angelis, J., Applebee, A., Amodeo, V., and Snyder, M. 2012. Best practice in middle-school science. Orlando, FL: Paper presented at the annual meeting of the National Association for Research in Science Teaching, Orlando, FL.

Olszewski-Kubilius, P. 2006. Addressing the achievement gap between minority and nonminority children: Increasing access and achievement through Project EXCITE. *Gifted Child Today,* 29:28–37.

Poliquin, R. 1997. Time spent on core discipline areas in elementary schools. Maine Education Policy Research Institute, Center for Education Policy, Applied Research and Evaluation, College of Human Development, University of Southern Maine, Gorham.

PollingReport.com. 2007. An independent, nonpartisan resource on trends in American public opinion. Available at http://www.pollingreport.com/index.html (accessed 11 December 2007).

Popper, K. 1999. The logic of scientific discovery. New York: Rutledge, p. 27–49.

Raizen, S.A., Jones, L.V., Anderson, R.D., Bradburn, N.M., Johnston, D.F., Kerins, C.T., Nisselson, H., Rubin, D.B., Smith, M.S., Tenopyr, M.L., and Welch, W.W. 1985. The schooling process: Instructional time and course enrollment. Indicators of pre-college education in science and mathematics: A preliminary report. Washington, DC: National Academy Press, p. 83–109.

Sarkar, S., and Frazier, R. 2008. Place-based investigations and authentic inquiry. *The Science Teacher,* 75:29–33.

Schwartz, R., and Lederman, N. 2008. What scientists say: Scientists' views of nature of science and relation to science context. *International Journal of Science Education,* 30:727–771.

Secada, W.G. 1992. Race, ethnicity, social class, language, and achievement in mathematics. *In* Grouws, D.A., ed., Handbook of research on mathematics teaching and learning. New York: Macmillan.

Slutskin, R.L. 2003. The reality of the K–12 science public education system. National Office for the Integrated and Sustained Ocean Observations. Available at http://72.14.205.104/search?q=cache:DWupCpvigAkJ:www.ocean.us/system/files%3Ffile%3D2Formal-Reality_of_K-12_Science.pdf+What+is+the+average+time+elementary+students+spend+learning+science%3F&hl=en&ct=clnk&cd=9&gl=us (accessed 11 December 2007).

Somnath, S., and Frazier, R. 2008. Place-based investigations and authentic inquiry. *The Science Teacher,* 75:29–33.

U.S. Department of Education. 2001. No Child Left Behind Act of 2001. Washington, DC: U.S. Department of Education.

U.S. Department of Education. 2007. About ED: Overview and mission statement. Available at http://www2.ed.gov/about/landing.jhtml (accessed 11 December 2007).

Visone, Jermey D. 2010. Science or reading: What is being measured by standardized tests? *American Secondary Education,* 39:95–112.

Appendix 1— Complete List of Casts and Skulls:

Superorder Xenarthra Skulls and Casts
- Order Pilosa
  - Family Bradypodidae
    - *Bradypus tridactylus*
  - Family Megalonychidae
    - *Choloepus didactylus*
    - *Choloepus hoffmanni*
    - †*Megalonyx leptostomus* — Florida Pleistocene
    - †9 unknown small ground sloths specimens — Argentina Miocene
  - Family Mylodontidae
    - †*Catonyx tarijensis* Bolivia Pleistocene
    - †*Glossotherium chapadmalensis* — Florida Pliocene
    - †*Scelidodon* sp. — Argentina Pleistocene
  - Family Myrmecophagidae
    - *Myrmecophaga tridactyla*
    - *Tamandua Mexicana*
- Order Cingulata
  - Family Glyptodontidae
    - †*Glyptodon calvipes* — Uruguay Pleistocene
    - †*Panochthus tuberculatus* — Argentina Pleistocene
  - Family Pampatheriidae
    - †*Holmesina septentrionalis* — Florida Pleistocene
  - Family Dasypodidae
    - *Dasypus novemcinctus* (15 skulls as a control)
    - *Cabassous unicinctus*
    - *Euphractus sexcinctus*
    - *Priodontes maximus*

Outgroup Skulls and Casts
- Order Monotremata
  - Family Tachyglossidae
    - *Tachyglossus aculatus*
- Order Didephimorphia
  - Didelphidae
    - *Didelphis virginiana* (5 skulls as a control)

Appendix 2 — Complete Copy of the Assessment Instrument:

Questions scientists ask can be measured and analyzed. If the question is not testable, it is outside of science. Write "<u>YES</u>" if the question is testable; "<u>NO</u>" if the question is non-testable.

1. ____ Are rabbits cute?
2. ____ Are the scratch patterns on glyptodon teeth more similar to the scratches on herbivore teeth or carnivore teeth?
3. ____ Are armadillos empathetic?
4. ____ Are ghosts real?
5. ____ Are sloths slow because they are deep thinkers?
6. ____ By measuring the length of the last molar, can we predict how big a giant ground sloth was?
7. ____ Moles live underground. How much oxygen is present in their tunnels?
8. ____ Can tarot cards tell the future?
9. ____ Are sloths slow because they are lazy?

Fig. 1. Giant armadillo. Source: Wikipedia. Available at: http://upload.wikimedia.org/wikipedia/commons/5/5d/Tatucarreta.jpg

10. ____ Hyenas have the strongest bite force of any mammal. Do hyenas have more jaw muscles than other mammals?
11. ____ Are people born under the sign of Leo the Lion courageous?
12. ____ Did the attachment of the muscle on the jaw of tree sloths and ground sloths change even though they are/were herbivores?
13. ____ Do skunks stink because they have bad spirits inside of them?
14. ____ What is the average temperature in Portland, Oregon?
15. ____ At what temperature do people begin to shiver to stay warm?
16. ____ Do plants need oxygen to stay alive?
17. ____ Sloths, anteaters and armadillos all have long claws. Which group has the longest claws proportional to their body size?
18. ____ Can armadillos understand your vulnerability because they have shells?
19. ____ Did anteaters lose all their teeth as a punishment because they wanted to eat little animals?
20. ____ Can we analyze how armadillos and glyptodonts chewed by measuring the place on the jaw where the muscles attached?
21. ____ As a scientist, describe the picture of this mammal, a giant armadillo. See Figure 1.
22. ____ You are the lead paleontologist at this dig site. Area A was dry land. Area B was shallow water. The location of the fossil animal and the fossil eggs were in the same layer. An analysis of the eggs indicate that they are the same species as the fossil animal. As a scientist, describe what you have found. See Fig. 2.
23. ____ Please tell me something you know about tree sloths, anteaters, armadillos, ground sloths, glyptodonts, and/or pampatheres.
24. ____ What have you noticed about tree sloths, anteaters, armadillos that are <u>the same</u> as ground sloths, glyptodonts, and pampatheres?
25. ____ What have you noticed about tree sloths, anteaters, armadillos that are <u>different</u> than ground sloths, glyptodonts, and pampatheres?

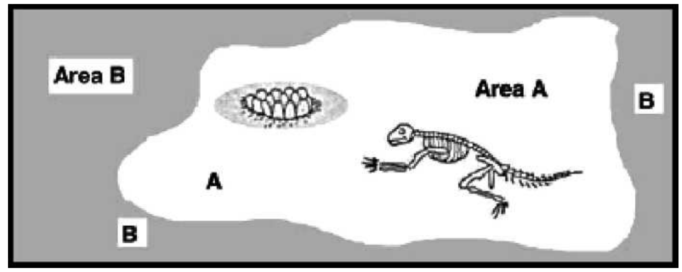


Fig. 2. GeoTAT instrument. Source: Dodick J., and Orion N. 2003. Measuring student understanding of geologic time. *Science Education* 87(5):708-731. Used with permission.