Cognition and Self-Efficacy of Stratigraphy and Geologic Time: Implications for Improving Undergraduate Student Performance in Geological Reasoning

Erin Peters Burton[1,a)] and G. K. Mattietti[2]

ABSTRACT

In general, integration of spatial information can be difficult for students. To study students' spatial thinking and their self-efficacy of interpreting stratigraphic columns, we designed an exercise that asks college-level students to interpret problems on the principles of superposition, original horizontality and lateral continuity, and geologic time using text and symbols. The exercise was designed with two goals in mind: to determine the level of student confidence and cognition and to test the effectiveness of this type of exercise in large-enrollment courses. Overall, students performed well on symbolic representations of the columns, but reported low self-efficacy of their interpretations. The opposite occurred with the short-answer questions. Results suggest that these students are more comfortable with verbal questions, but they lack the ability to synthesize complete answers to diverse questions. Students also tended to feel less comfortable with questions where they had to convert text to a symbolic representation. We found this type of assignment to be extremely useful with a large class, as it elicited much information about student learning without taking extensive time to evaluate. Implications for geoscience educators include the need to incorporate techniques to improve the completeness of student responses on problems that require synthesis. © *2011 National Association of Geoscience Teachers*. [DOI: 10.5408/1.3605042]

INTRODUCTION

Learning environments in science strive to set two goals for students: (1) to master knowledge and understandings constructed by previous generations of scientists, and (2) to be able to construct new scientific knowledge themselves (Gilbert, 2008). Understanding prior knowledge and constructing new scientific knowledge requires background knowledge and the ability to think at high cognitive levels. Bloom's taxonomy of the cognitive domain can be a helpful tool in categorizing lower level and higher level of thinking. This taxonomy was created in 1956 by a team of educational psychologists when they first noticed that over 95% of the questions they encountered on tests in college classes were lower levels of cognition. The original Bloom's taxonomy of the cognitive domain (1956) categorized cognition levels from lowest to highest as knowledge, comprehension, application, analysis, synthesis, and evaluation. More recently, Bloom's taxonomy has been revised to include the latest findings from cognitive research and to reflect a more active type of thinking. The new categories of cognition from lowest to highest order of thinking are: remembering, understanding, applying, analyzing, evaluating, and creating (Anderson *et al.*, 2003).

Both spatial reasoning and an ability to think at Bloom's taxonomic level of synthesis are prerequisites for knowledge construction in the field of geology. The discipline draws principles from other physical sciences and relies heavily on observation, and deductive and inductive thinking skills. The use and correlation of stratigraphic columns are skills employed extensively in geology and are based on the ability to draw on multiple symbolic representations used to interpret the geologic conditions and the evolutionary history of a region.

It is known that humans in general find that it is difficult to integrate spatial information from two-dimensional to three-dimensional representations, and that individuals vary widely in their spatial integration ability (e.g., Black, 2005; Duesbury and O'Neil, 1996; Ishikawa and Montello, 2006). Study of students' thinking regarding stratigraphic columns can lead to new information about the teaching and learning of these important cognitive tasks.

Typical course content from an introductory level historical geology course addresses concepts such as correlation, "deep time" (McPhee, 1982) and evolution; these core concepts are not a part of everyday thinking. Research shows that diachronic thinking, one component in the ability to fully understand geologic time, develops in children by 10–12 years of age (Dodick and Orion, 2003a, 2003b, 2006). In addition, it has been shown that although students may not have experience with the scale of geologic reasoning and with the length of time between geologic events, they may hold preconceived knowledge that includes misconceptions about time and Earth's history (Libarkin *et al.*, 2005, 2007). Such misconceptions seem to be invariant across gender, race, age, and social status (Schoon, 1992).

A correct understanding of geologic time has a fundamental role in the learning about Earth Systems (Zen, 2001; Libarkin *et al.*, 2005, 2007). With respect to understanding geologic time, there is a growing body of evidence showing that field activities related to regional geology settings are the most effective springboards to developing diachronic reasoning (Miller, 2001; Thomas, 2001). Because of logistic challenges and lack of resources, these field benefits are rarely available to the majority of students enrolled in general geology classes.

Received 20 November 2009; revised 7 December 2010; accepted 2 February 2011; published online 1 August 2011.

[1]Department of Science Education and Educational Psychology, George Mason University, 4400 University Dr., MSN 4B3, Fairfax, Virginia, USA
[2]Department of Atmospheric, Oceanic & Earth Science, George Mason University, 4400 University Dr., MSN 4B3, Fairfax, Virginia, USA
[a)]Author to whom correspondence should be addressed; electronic mail: epeters1@gmu.edu

Students find it very challenging to visualize the nonvisible sides of three-dimensional geologic structures (Kali and Orion, 1996). To this already challenging three-dimensional thinking, historical geology adds the variable of "deep time," resulting in the need for four-dimensional thinking (Ault, 1982). In spite of the useful reasoning skills and high cognitive load required to master the study of geology, the majority of students approach the study of geology as a purely academic activity rather than a way to understand the world, so they resort to superficially learning the material (Pinet, 1995; Prothero, 2000). Students often perceive that they can merely memorize the information to pass the exams, and entirely miss out on the rich cognitive experience the geosciences have to offer (McManus, 2002).

Science education reform documents, such as the *Benchmarks for Scientific Literacy* (1993), stress the importance of actively engaging students in problems involving realistic scenarios that are consistent with the nature of scientific investigation. In large introductory geology courses, often populated by over 100 students, it is difficult to promote an exchange of questions and answers that enables the instructor to understand the cognitive level at which students are mastering the subject and to illuminate the reasoning behind the content. Because of logistical reasons and lack of resources, most instructors of large classes choose the instructional delivery method of lecturing, supplemented by student questions from the floor (Prothero, 2000). This pedagogical choice makes it very difficult to gauge a students' level of confidence in understanding the material; oftentimes the only cues to the instructor are the students' expressions of discouragement or disinterest. Assessment of student geologic reasoning in a large class poses problems because to gather meaningful student reasoning and to provide timely feedback is very time consuming (McConnell *et al*., 2003). Classroom communication systems like clickers are a popular tool for giving students timely feedback on their thinking during a lecture. This formative assessment, however, does not consistently access student cognition at a high level because the question format must be forced-response. In other words, the answer is always present in the choices so students do not construct the answer in their own words.

For those instructors who only use a traditional, multiple-choice exam to determine the level at which individuals and the class (as a whole) have learned the material, there is a lack of meaningful educational communication between the instructor and students. This exacerbates the lack of meaningful educational communication between the professor and students. The challenge for geology educators that teach large classes is to give and receive feedback quickly and effectively about their students' level of knowledge. Meaningful feedback between the instructor and students can help to better engage with the students' cognitive processes and permits assessments to align with course goals. Addressing this challenge at the introductory level is to approach the discipline as a way of knowing rather than just as a huge body of knowledge, where the understanding of basic principles and their application is contextualized in the discipline of geoscience.

Because introductory level geology courses fulfill general education requirements for the sciences, often students populating historical geology classes are not pursuing a major in a scientific discipline, and quite a few of them carry a self-defeating attitude toward their success in science classes (Wagner, 2000). Ideally, as students grapple with concepts and their application to geologic problems, they should acquire enough confidence to proceed to the correct solution by a set of well thought out steps rather than by enunciating memorized principles, rules, and nomenclature. Therefore, assessing student self-efficacy of reasoning in geoscience should be considered as important as assessing content knowledge.

Self-efficacy of reasoning refers to students' judgments of their capabilities to perform thinking tasks in scientific disciplines (Schunk, 1991). Self-efficacy is not an innate trait of students, but rather it can be changed by the learning environment in which students are participating (Schunk and Zimmerman, 1994; Dweck and Leggett, 1988). Self-efficacy has been related to student success in several academic domains. Zimmerman and Martinez-Pons (1990) found that verbal and mathematical self-efficacy correlated positively with the use of effective learning strategies. Pintrich and De Groot (1990) found that self-efficacy and cognitive strategy use were positively correlated and predicted achievement in the cognitive domain measured. In other domains, self-efficacy has been positively correlated with achievement in writing with college students (Zimmerman and Bandura, 1994), and it has been linked with students' choice to engage more significantly in written test items (Tuckman and Sexton, 1990). Thus, we suggest that improving the self-efficacy of geologic reasoning could be used to both boost student self-confidence about the subject and to enhance students' learning achievement.

For this purpose, a pilot self-assessment exercise was designed to review principles of superposition, original horizontality, lateral continuity, and concepts of geologic time, and to investigate the level of student confidence and cognition.

Additionally, the designed exercise was being tested to determine its usefulness as an assessment tool that provides feedback to students on their cognition and confidence of their answers. The exercise was offered as an in-class activity to the students of one section of historical geology ($n=134$ students) offered at a university in the mid-Atlantic region of the United States. On average, about 80% of the students taking this class are freshmen and sophomores; it is common for seniors to take this class as well in order to fulfill the second general education science requirement for a class with laboratory. Up to 20% of the students in this class were science majors, less than 5% had declared a major or minor in Earth Science related fields. In general, there were about as many female as male students. The ethnic distribution for the university during the semester of the study was 8% African American, 13% Asian American, 8% Hispanic, 1% Native American, 50% white, and 20% reported as other or not reported.

The exercise sought answers and student self-efficacy to a set of nine questions based on two greatly simplified stratigraphic columns (Fig. 1, Table I). The questions were offered in a table format, with a column for the questions, a second column for the corresponding answers, and a third column for the confidence for the answer.

The information gained by the instructor on student outcomes can help to design future instruction that better suits the needs of the students. As the outcomes of this

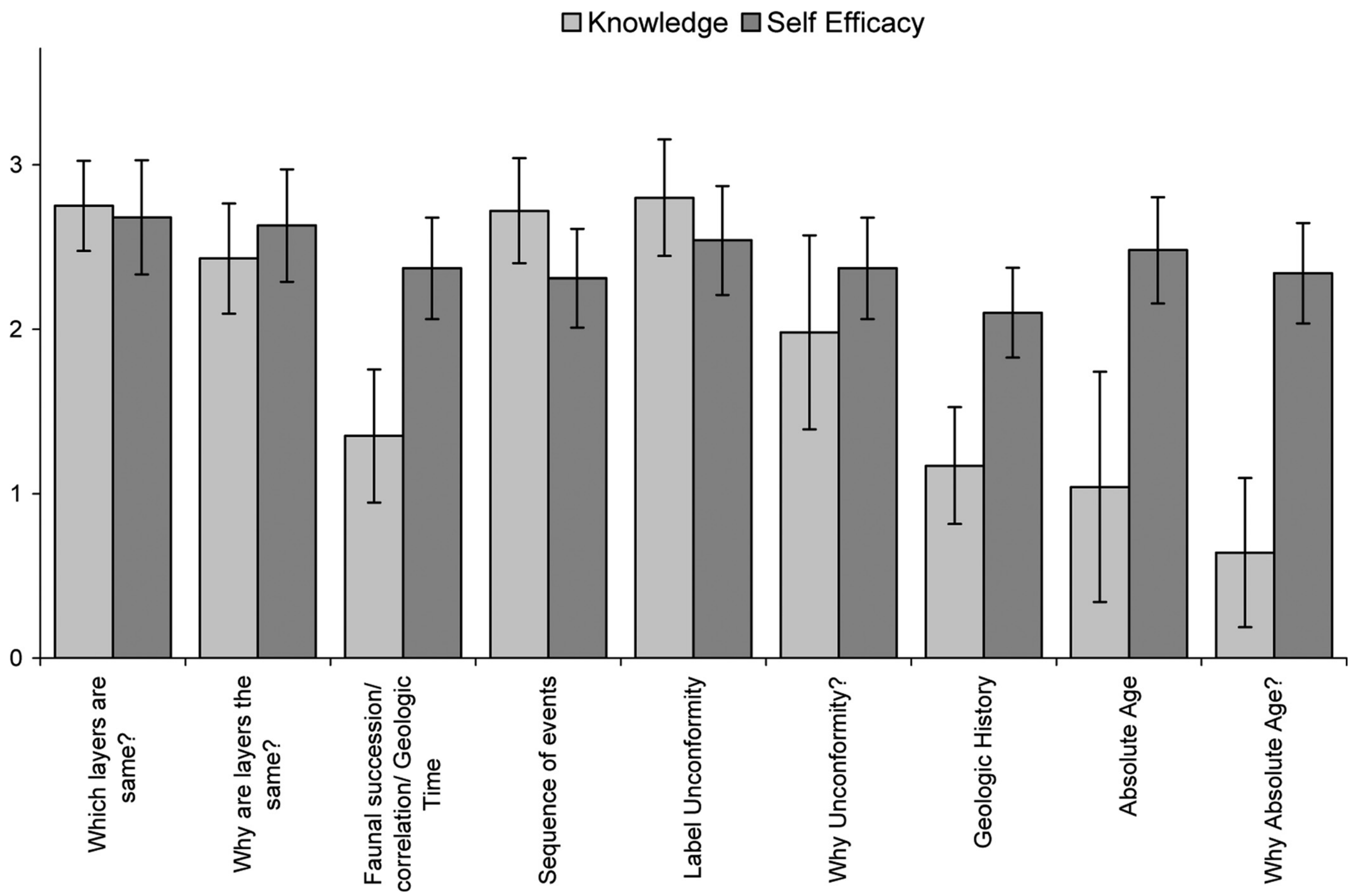

FIGURE 1: This example of a typical student's answers and rating of self-efficacy demonstrates that such an exercise can be useful in discerning the knowledge of students in large classes.

exercise were studied, four questions drove the analysis of data collected:

1) How confident are undergraduate students in their ability to apply basic stratigraphy knowledge to two simplified stratigraphic columns in their second semester of geology?
2) How proficient are students in synthesizing prior knowledge in problem solving relative to their interpretation of stratigraphic columns?
3) Will students' knowledge match their confidence level?
4) Are there particular levels of cognition, as defined by the revised Bloom's Taxonomy (Anderson *et al.*, 2003), where students have equal confidence levels and levels of declarative knowledge about the content?

METHODS

The Self-Assessment Exercise

The assignment consisted of answering nine questions found in Table I about the stratigraphic columns shown in Fig. 1. Table I includes the cognitive loads of the questions and the highest level of revised Bloom's taxonomy of each question. Cognitive load and level of taxonomy were pieces of information not provided to the students in the assignment. Students rated their level of confidence of their answer, with 1 for being absolutely sure, 2 for being hesitant, and 3 for just guessing. Students in this course had used this scale previously on a similar exercise. However, because we now wanted to compare student self-efficacy with cognition scores that represented highest cognition with the highest score, we reversed the scale in our analysis tool, SPSS, so that the scales matched. In the analysis, scores were 3 for being absolutely sure, 2 for being hesitant, and 1 for just guessing. Zero was assigned to students who did not answer or declared having no idea. This 4-level scale was chosen because a more traditional 5-point Likert scale includes a "no opinion" option. The authors felt it was important to provide only options where students needed to consider their self-efficacy of the problem rather than allowing students to avoid providing a substantial answer with a "no opinion" option. Writing a short answer was chosen because it provided students with an active way to apply their knowledge in a more engaging way, and helped to reveal more of their thinking than a passive multiple choice option. This exercise was offered right after the fourth week of class, after students encountered lectures with the course's basic principles and tools for understanding Earth history: sedimentary environments, facies, fossilization, principles of original horizontality, superposition and lateral continuity, laws of faunal succession, biostratigraphic correlation, unconformity, and ways to measure geologic time.

The self-assessment exercise was purposely composed of a very simple scenario so students could focus on

TABLE I: Cognitive levels for questions.

Question	Cognitive load	Key words for correct answer	Highest Bloom's taxonomy level
1. Which layers are the same?	Pattern recognition No discipline specific knowledge	B=D; C=F	Understanding
2. Why do you say so?	Correlation of lithologies and fossils	Same fossils, same lithologies	Applying
3. What do the fossils in these rocks indicate with respect to the geologic time when these rocks are formed?	Faunal succession principlePrinciple of superposition Index fossil	Same fossils + same lithologies indicate that the rocks were forming at the same time in the same type of environment. Based on principle of superposition it is possible to establish which rocks/events are older/younger in each column	Analyzing
4. What is the correct sequence of rock layer from oldest to youngest?	Understanding the meaning of correlation – Use concept of question 3 to hypothesize the succession of events.	A B=D C D=F	Analyzing
5. On the figure mark the unconformity with an arrow	Definition of Unconformity What an unconformity looks like – definition	Student clearly marks the unconformity (the wiggly line is the symbol)	Remembering
6. Why did you place the arrow there?	Definition of Unconformity	Comparison of the two columns, missing lithology E from left column Identification of the "squiggly" line	Understanding
7. What does the unconformity indicate about the geologic history represented by these two stratigraphic columns?	Understanding , Application and analysis of the Unconformity using data from both columns	Missing lithology E A time-gap in the geologic history of the left column.Action of geologic forces (surface processes or plate tectonics).Surface or erosion or non deposition	Analyzing
8. Can these rocks give you the information on the absolute age of their formation?	Relative vs. Absolute age determination	No	Remembering
9. Why yes or why not?	Significance of closed systems	Sedimentary materials are non suitable for absolute age determination	Understanding

explaining their reasoning and apply their problem solving skills in a relatively short time frame (15 min at the end of the lecture). It was emphasized that their answer was to be concise (fit the assigned space) but articulate, so they could clearly justify their reasoning as much as possible. Students were also informed about the value of self-efficacy as a tool for enhancing the effectiveness, motivation, and engagement connected to their learning experience. Active involvement with the subject (problem solving) was presented as a key to ownership of learning. In addition to the opportunity to practice with exam-like questions, the student received a small extra-credit bonus granted on completion of both sets of questions and confidence level. The authors felt that by explaining the rationale of the exercise with respect to the discipline and to the reflective practice of self-assessment and by providing a small extra-credit reward, students were more likely to be motivated to try their best and to be candid about their level of preparation.

The students turned in their exercise and were informed that the key was going to be available to them on a class website right after class, so they could check their answers. The graded worksheet was returned to the students 48 h later, during the following lecture meeting. Students were instructed to compare their exercise to the key to make sure that they had answered correctly, and to compare the results with their level of confidence. A correct answer accompanied with a high level of confidence meant that the students' preparation was good, and thus encouraged the student to work on maintaining their study strategies. A tentative answer meant that more attention needed to be devoted to the topic of the question, whether or not the answer given was correct. Students in this category were encouraged to spend a little more time in focused studying or to find strategies for optimizing their efforts. Guessing indicated that significantly more studying was needed, so students scoring low in both knowledge and confidence were invited to spend significantly more effort studying for the exam. They were also encouraged to see the instructor for assistance to strategize preparation for exam in the short time left. Finally, a high confidence level on a wrong answer identified a potential for students to have a secure feeling that they understand the information, when they did not understand the content correctly. It was explained to the students that if that was the case, they needed to meet with the instructor as soon as possible in order to address the misunderstanding. In the end, each student had a personalized snapshot of their level of knowledge so they could take advantage of the two week time period before the exam to work at remediating, consolidating, or maintaining the level of preparation.

The time required for the instructor to read one sheet and to mark out misconceptions or other major need of improvement, and record the extra-credit points for each student took approximately 1 min, for a total time of 1 ½ h of work. Overall, the activity took about 30 min of lecture time: 5 min to present it, 15 min to carry it out, and another 5 min of comments when it was returned; another 5 min were counted for distribution and picking up the worksheet. The instructor optimized the class time for the assignment by making the exercise key and commentary available on the course Blackboard page. End of term feedback about the course activities indicated that students found this assignment a very useful tool for exam preparation.

Rationale for the Assignment's Design

With this assignment the authors planned to achieve two goals: 1) to give the instructor a view of the cognitive level achieved by the class as a whole and of possible misunderstandings developing on the topics of the exercise; 2) to provide students with a reflective approach for preparing for their exams through the use of self-efficacy.

Students in introductory classes, especially non-science majors, approach learning in sciences as the memorization of concepts resulting in the ability to list and identify concepts (Tobias, 1990). In the domains of cognitive performance, these concepts fall into the lowest category of remembering. The ability to recall names and definitions is indeed a fundamental step toward mastery of geology, however, if left at this stage, student learning may result in a useless exercise in memorization for the purpose of regurgitation on tests. This low-level type of knowledge is often forgotten shortly after tests are taken and frustrated students eventually question the purpose of such mental gymnastics (Baddeley and Della Sala, 1996). Student reporting of academic self-efficacy was used because it has been found to be negatively correlated with student disengagement (Bandura *et al*., 2003). That is, students who have low academic self-efficacy tend to make excuses for bad behavior, blame others, and do not take responsibility for the consequences of their lack of study skills.

Through this assignment, students had an opportunity to compare their level of cognition for each question with their confidence level. The instructor guided students in their comparisons between cognition and self-efficacy and explained further actions that students should take to enhance their learning based on the results of the comparisons.

The design of the assignment was based on the "backwards design" guidelines of Wiggins and McTighe (2005): first establishing the level of mastery for the concepts in the assessment, then crafting the questions around a problem solving approach: recognizing, interpreting, connecting, and hypothesizing about the stratigraphic columns. The cognitive load and level for each question are shown in Table I, while the conceptual expectations for each question are addressed below.

Questions 1–4 address the correct understanding of how faunal succession is used to establish correlations and, subsequently, how sequences of events are built from rocks cropping out at different sites. Question 1 requires a simple recognition that units C and F are the same because they have the same pattern representing the rock and the sketch representing the fossil. To answer question 2, however, students needed to articulate the fact that the geologic equivalence of the strata was given by both the fossils and the rock facies; identifying only the rock is not sufficient. For question 3, students were asked in an open ended way to analyze the information from the two columns. Students were required to perform a correlation based on the application of principles of stratigraphy and faunal succession. Students were expected to also conjecture that the fossils could have been index fossils, a concept they were familiar with from the lecture. Question 4 required students to correlate the two columns to reconstruct a sequence of events. Note that both questions 2 and 4 could be answered correctly just by identifying patterns and by understanding sequencing.

Questions 5–7 focus on the significance of an unconformity. An unconformity is a multidimensional concept, it requires the understanding of physical change over time. Open-ended answers to questions on this topic are most likely to reveal student thinking because the student must construct the answer rather than being forced into a particular response. Question 5 shows a surface that is disrupted in a more irregular way than other boundaries among strata in the two columns. Students were asked to explain this reasoning in question 6, i.e., how did they identify the position of the unconformity in the figure. Question 6 could be answered by merely stating the criteria for labeling the unconformity, thus with minimal discipline knowledge. Question 7, however, asked for an explanation of the existence of the unconformity and its interpretation using both columns.

The last two questions were meant to probe the understanding of how "deep" time builds up on two different approaches: the correlation of faunal successions, which gives a relative timeline, and the numerical measurement of age based on the relative abundances of parent–daughter isotopic pairs. These questions implied the understanding that geologic time is a measure of change recorded in the facies of the examined rocks and that only rocks (or minerals) that can be considered chemically closed systems are suitable for numerical age determination.

Answering without the option of the multiple choice caused the students to make their reasoning more transparent by communicating ideas in their own words. Because students knew that they were not penalized for recording a low level of self-efficacy or for giving the wrong answer, the instructor was able to place emphasis on learning as a progression rather than as a one-time event.

RESULTS

In this section of the paper the results of the cognition and self-efficacy of the group are examined. Out of 134 students participating in this exercise, 84 returned their worksheet for this study. The levels of cognitive domain were determined through collaboration of both an educational psychologist with a science education background (first author) and a geoscientist who is an educator at the college level (second author). There was full agreement on the levels of cognition for the questions between the two authors and the scoring rubric was written based on this information. The rubric and the rating rationale were also reviewed and agreed upon by three other geoscience faculty not involved with this study. Student answers were scored separately by the authors, who then met to discuss their scorings. Ratings were recorded when researchers reached a consensus on the scores. Answers that displayed a deep understanding of the content received a score of 3. Answers that demonstrated an incomplete view of the content received a score of 2. Answers that had a small amount of correct factual information, but did not meet expectations received a score of 1, and answers that were entirely incorrect received a score of 0. The third column in Table I indicates the phrases that represent the cognition required for a maximum score of a 3. Ratings for level of confidence used a scale from 3–0. As mentioned previously, the scale was reversed for the purpose of analysis so that 3 indicated that the student was absolutely sure of the answer, 2 indicated that the student thought that they could be correct, and 1 indicated that the student had significant doubts about the correctness of the answer. Finally, 0 indicated that the student had no idea whether the answer they provided addresses the question.

Descriptive statistics were calculated for cognition (scoring raging from 0–3) and self-efficacy (scoring ranging from 0–3), and 2-tailed Pearson correlations were performed between the class scores for the questions dealing with cognition and self-efficacy. Paired-sample t-tests were performed to compare the means of the cognition and self-efficacy for each question. Cronbach's alpha test of reliability for the self-efficacy ratings (α=0.75) showed high reliability among the responses. The inter-rater reliability for the scores of cognition was found to have 97% alignment over 100% of the scoring samples. Each rater utilized a rubric which ensured that ratings were consistent.

Pearson correlations showed that two of the questions had statistically significant high negative correlations between the correct answer and the self-efficacy about the answer: question 3 on geologic time ($r = -0.236$), and question 9 ($r = -0.279$) on how absolute age might be calculated. Students' answers on questions 3 and 9 were low, 1.35 out of 3 and 0.64 out of 3, respectively, yet student self-efficacy was high, 2.37 and 2.34, respectively.

The next grouping of correlations showed mild connections between the answer and the self-efficacy of the answer: question 2 ($r = -0.172$), question 5 ($r = -0.172$), and question 8 ($r = -0.153$). Again the correlations between the score of the answer and the score of the self-efficacy were negatively correlated, and the answer in all three cases was a lower score than the student self-efficacy reported score. Table II shows the correlations between the self-efficacy score and the score of the correctness of the question.

All but one question showed significant differences between the correctness of the answer and the self-efficacy of the answer. The first question required students to identify the symbols in the diagrams and recognize patterns; this was the only question that did not require any discipline specific knowledge. This was the only question that did not show significant differences between cognition

TABLE II: Correlations between self-efficacy score and declarative knowledge score.

Question	Correlation
Which layers are the same?	0.058
Why are the layers the same?	−0.172
Faunal succession/correlation/geologic time	−0.236 [a]
Sequence of events	0.065
Label unconformity	−0.172
Why unconformity?	0.070
Geologic history	0.091
Absolute age	−0.153
Why absolute age?	−0.279[a]

[a]Indicates a significant correlation.

(M=2.75, SD=0.55) and self-efficacy (M=2.68, SD=0.49). Table III displays the results from the means and the paired-sample t-tests; the graphic representation of these data is shown in Fig. 2. The second question required students to explain that a lithological correlation is not sufficient for the rock beds to be of the same age, and it must be coupled with the fossil content for the same type of rock for it to be considered of the same age. Student self-efficacy (M= 2.63, SD = 0.58) was reported to be significantly higher than their cognition (M= 2.43, SD = 0.67) at a 0.05 level. Students were very sure of their answers, although roughly 60% of the answers were only partially correct. A minority of students had all of the components of the answer which would be indicated by a score of 3. Students often addressed only the fossil being the same or the type of rock in the layers being the same, but not both fossil and rock layer being enough evidence to show that indeed the layers were the same. For example, students reported, "[They] Have the same type of fossils that were buried during that period," "The fossils look exactly the same," and "It's the same type of rock." Another common aspect of the students' explanation was being vague and using pronouns instead of identifying either the fossil or the rock layer, and restating the question with a "They are identical to one another."

The third question required students to apply the faunal succession principle. Again, the students greatly overestimated their confidence in being right (M=2.37, SD=0.67) with their cognition (M=1.35, SD= 0.81) significantly ($t(85) = 10.31$, $p < 0.001$). In their answers, students often merely restated the question, gave vague and imprecise notions of when the fossil was formed, or identified rules without connecting it to the diagram. Students gave answers such as "The fossil represent[s] that geologic time in history," "They died and got buried beneath the layers," "[this indicates]The geologic era and which animals were around during that time," and "[this indicates]Geologic Era (depending on fossil depends on time of existence)." Only 1 student out of 84 mentioned that it was important for the fossils to be index fossils for the accuracy of the relative time determination.

The fourth question required a higher level of knowledge, as students needed to combine the superposition, original horizontality, and lateral continuity principles, and to be able to correlate the two columns to indicate the

TABLE III: Comparison of means for declarative knowledge and self-efficacy.

Question	Mean (Answer) (Range 0–3)	SD (Answer)	Mean (Self-efficacy) (Range 0 – 3)	SD (Self-efficacy)	*t*	*p*
1. Which layers are the same?	2.75	0.55	2.68	0.49	−0.91	0.374
2. Why are the layers the same?	2.43	0.67	2.63	0.58	2.19	0.032*
3. Faunal succession/ correlation/ Geologic time	1.35	0.81	2.37	0.67	10.31	0.001*
4. Sequence of events	2.72	0.64	2.31	0.71	−4.08	0.001*
5. Label unconformity	2.80	0.71	2.54	0.64	−2.42	0.018*
6. Why unconformity?	1.98	1.18	2.37	0.69	2.71	0.008*
7. Geologic history	1.17	0.71	2.10	0.69	9.04	0.001*
8. Absolute age	1.04	1.40	2.48	0.65	9.10	0.001*
9. Why absolute age?	0.64	0.91	2.34	0.68	16.15	0.001*

*Represents statistical significance at p < .0 = 5 level.

chronological sequence. In this case, students felt significantly less confident (M=2.31, SD=0.71) in their correct answers (M=2.72, SD=0.64) where $t(85) = -4.08$, $p < 0.001$. Most students had this answer entirely correct, however, they felt less confident in that answer.

The fifth question asked students to place an arrow on the unconformity. Students could answer the question using two strategies: 1) have knowledge about the symbol and identify the symbol used to indicate an unconformity, or 2) recognize the break in the pattern of the rock layer types. Scoring for this question resulted in either a 3 (entirely correct) or a 0 (not correct). There was a significant difference between the student answer and their self-efficacy about the answer [$t(64) = -2.42$, $p = 0.018$], and students were less confident in their answers (M=2.54, SD = 0.64) even though the answers were mostly correct

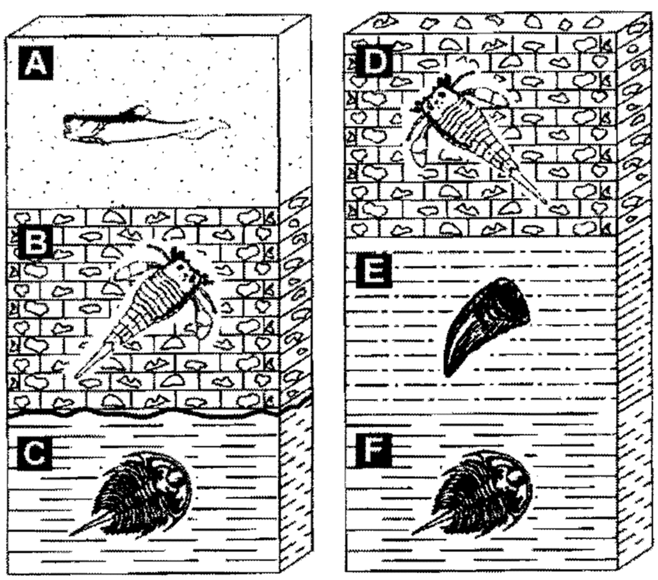

FIGURE 2: A graph of self-efficacy (light bar) and declarative knowledge (dark bar) of exercise illustrates the inverse relationship of the variables, particularly in the case of written answers. Vertical whisker bar represents the standard deviation.

(M=2.80, SD=0.71). This finding indicates that being correct does not necessarily indicate high confidence in an answer, and that low confidence in an answer does not necessarily indicate an incorrect answer.

The sixth question asked students to justify why they marked the arrow on the diagram as they did. This question gave some insight into which strategy students used to identify the unconformity: identifying the symbol for unconformity or recognizing a break in the pattern in the diagram. Thirty percent of the students who had a correct answer recognized the symbol of the jagged line to identify the unconformity, and the remainder used the break in the pattern to identify this phenomena. Students had a significantly higher confidence (M=2.37, SD=0.69) than their ability to report a correct answer (M=1.98, SD = 1.18), where $t(83) = 2.71$, $p = 0.008$.

The seventh question required students to analyze the differences between the rock layers in the two columns, and students tended to recognize the differences, but began to make unsupported claims by speculating about the reasons for the differences. As indicated above, the most common answer was that a catastrophic event, such as an earthquake, was responsible for the change. Some representative answers included, "Something happened, maybe an earthquake or massive storm," "Perhaps a huge catastrophic even happened and blew out [layer] E," and "That weathering or natural disasters have altered the composition." Some other surprising inferences were revealed from this question such as, "It represents that something happened to change the strata in the column that didn't happen to the column 30 miles away, making them different," and "One column has some event which caused no fossil to form at one point" as if the rock layer were responsible for the fossil formation. There was a highly significant difference between student confidence level and cognition for this question ($t(83) = 9.04$, $p < 0.001$), as the students reported a higher self-efficacy (M=2.10, SD=0.69) than was achieved by their performance (M=1.17, SD=0.71).

The eighth and ninth questions yielded the most surprising results, as the students were extremely confident of their answers that were almost entirely incorrect. The eighth question asked students if the rocks depicted within the columns could give the absolute (numerical) age, and a majority of students answered yes (M=1.04, SD = 1.40) and reported high confidence in this incorrect answer (M=2.48, SD= 0.65), where $t(83)= 9.10$, $p<0.001$. Looking at the answers to the ninth and last question, some reasons why students rated their confidence so high when they were incorrect so many times became evident. Students did not consider that fossils are not the original organisms but are part of a chemically open system, thus not suitable for numerical age determination. Only one student answered correctly, stating "No[,] there are no layers shown that are testable for rate of decay. Only igneous and metamorphic can give the accurate actual age." The remaining 83 students incorrectly wrote that the presence of fossils made the rock datable by numerical techniques and were very confident that this answer was appropriate. This was shown in the quantitative data where the mean score for the last question was M=0.64, SD = 0.91 and the mean for the self-efficacy of the answer was M=2.34, SD=0.68. The difference between the two scores held the highest significance at $t(84) = 16.15$, $p<0.001$.

DISCUSSION

In this section, the discussion will be guided by the research questions that drove this study in order to discern reasons for the results and implications for teaching and learning.

How confident are undergraduate students in their second semester of geology in their interpretation of Stratigraphic columns?

When looking at the questions given to the students, it was found that students recorded higher scores for their self-efficacy than they scored on their knowledge on the questions that required a written response. Students recorded lower self-efficacy than the scores they earned on the three questions that required symbolic representation (choosing the layers that were the same, representing the chronological sequence with letters, and labeling the unconformity with an arrow). This suggests that students are less confident with questions that have convergent answers, that is, questions where there is one best answer, such as having to choose corresponding layers. Students in this study tended to be more confident with answers that have more opportunities to communicate using their own words, such as short answer questions (divergent questions where there is more than one way to produce a correct answer), regardless of their level of knowledge. Students may have higher self-efficacy with short, written responses because they feel they can express themselves in a variety of ways. However, because the open questions were higher level, students often missed all parts of the questions and answered with only partial answers, unaware that they were incorrect.

Will students' knowledge match their confidence level?

Student achievement matched the reported level of confidence in only one case, the first question, and this question did not require discipline knowledge because students could match the pattern on the rock and the fossil without understanding that the fossil and rock together could identify layers that occurred at the same time frame. On all the other questions, content knowledge in the discipline was required to answer the question correctly, and student self-efficacy reports were significantly different.

Two questions had mildly significant differences between cognition and self-efficacy (question 2 where $p = 0.032$ and question 5 where $p= 0.018$), which meant that their self-efficacy, while not matched, approached the level of cognition. Perhaps this occurred with question 2 because students could begin to explain the relationship between the layers in the two columns due to their ability to interpret the symbols in the drawing. That is, each rock type was illustrated by a different pattern and each fossil representation had a different shaped organism, and the question was less complex because students were not asked to identify the organism. Students could infer the answer from the drawing with little discipline knowledge, as the question did not require a theory to back up the evidence in the drawing. The same could be said about question 5, where the unconformity was identified with a wavy line. This was the only wavy line in the drawing, and the word unconformity connotes something that is not regular, such as the only wavy line in the drawing. Thus both questions 2 and 5 could be answered with only a little content knowledge in geology.

The remaining six questions had very high levels of significant differences between the achievement of the student and their reported self-efficacy. Question 4 was the only question from this group that had a higher level of achievement than self-efficacy, meaning the students underestimated their confidence in being correct. This question required students to put together information about the two columns and to represent the chronological sequence with symbols. As discussed previously, students tended to be less confident of their ability to answer questions that had a forced answer. Conversely, results showed a highly significant difference in achievement versus self-efficacy for questions 3, 6, 7, 8, and 9, where students overestimated their confidence in their answers. All of the questions from this group were short answer where students were free to write down any answer. Students may have felt confident in their answer because they used technical terms or were able to rephrase the question, but did not recognize that they needed to synthesize several different pieces of information to answer the question correctly. For example, question 7 asked students to explain what the unconformity meant in terms of geologic history for the two columns. Students actually noticed that something different happened to the two columns, but did not recognize that the columns represented samples in context of areas. Meanwhile, other students reported that something catastrophic must have happened to the column with the layer missing, and they ignored that the change could have also happened slowly throughout time. Perhaps students felt that by getting one answer onto the paper, it fulfilled the requirements of science and neglected to think about all facets of the answer.

At the beginning of the course students were exposed to the different ways of understanding geological processes in the historical context of Catastrophism versus Uniformitarianism. They were also made aware of the different rates at which geology forces operate (e.g., very fast earthquakes and volcanic eruptions compared to very slow erosional processes, mountain building, etc…). These concepts were also reviewed when students were learning about unconformities. The most common interpretation of the unconformity given by the students to answer the question reflected a "Catastrophist" interpretation where explanations offered scenarios of short lived phenomena such as an earthquake, a volcanic eruption, or a tsunami with no reference to the amount of time that may have passed until the overlying rock in the sequence formed, as if the unconformity marked an instantaneous event. It is postulated that this interpretation comes from the sense of time measured in human terms rather than on a "deep time" planetary scale. In this light it is possible to understand the difficulty that students encountered in discerning absolute and relative time concepts when they had to provide an explanation in their own words.

Are there particular levels of cognition, as defined by the revised Bloom's Taxonomy, where students have equal confidence levels and levels of declarative knowledge about the content?

The questions were sorted into two groups in terms of the level of cognition based on the revised Bloom's Taxonomy (Anderson *et al*., 2000). Questions 1, 5, 6, 8, and 9 were categorized as lower levels of cognition such as remembering and understanding. These questions asked students to identify concepts and how they are represented on the graphic. It did not appear that the levels of cognition presented any correlation with students' ability to match their achievement to their own self-efficacy about answering the question. Question 1 was not significantly different, which means that there was a close match between achievement and self-efficacy, but questions 8 and 9 had the highest differences between student achievement and self-efficacy in answering the question. Questions 2, 3, 4, and 7 were sorted into the higher levels of cognition, where students had to analyze and synthesize ideas to construct an answer. Again, no correlations could be found with this level of analysis because question 4 revealed that students had a higher cognition than confidence for this type of symbolic question. In answering the other high levels of cognition, questions 2, 3, and 7, students showed that they had overestimated their confidence when compared to their achievement on the question.

IMPLICATIONS

Students had higher self-efficacy with the short answer questions than with the symbolic representation questions, regardless of their level of achievement on the question. Perhaps students believe that the answers to scientific questions are always solved with one simple idea rather than requiring several interconnected ideas to minimally answer a question. The idea that students expect one concept to be sufficient to address a problem is corroborated in the science education research literature. Instructors tend to use one-dimensional meaning relationships between concepts and meaning, rather than showing the intricate web of relationships among concepts (Lemke, 1990). Students may perceive that in science each question has but one correct answer. Patterns of language found in science classes tend to be broken into three stages: 1) the instructor posing a question, 2) a student responding, and 3) the instructor evaluating the correctness of the response (Lemke, 1990). This pattern of questioning and responding confirms that especially in the K–12 setting, one singular answer is sufficient to answer a question in science class. Most students in this study had only one full semester of experience beyond high school and may not yet understand how to synthesize ideas to form complete answers in geology.

Another viable explanation for the result that students had higher self-efficacy with short answer questions than with symbolic representation questions could be explained by anxiety toward symbolic-based information. Students could believe that they have higher difficulty answering questions that are based on symbolic information and then underestimate their ability to answer symbolic questions correctly. This trend is seen with "math anxiety" where students who do not consider themselves "math-minded" feel they cannot perform well under any conditions (Ashcraft, 2002).

Student satisfaction with a singular conceptual answer is especially dangerous in geoscience education, where rich description of phenomena is required as data. As noted before, geology is a discipline of synthesis that requires high cognitive skills to be mastered. Students, unfamiliar with new terms and concerned mostly about the multiple choice exam outcomes, tend to focus their efforts on passively memorizing the material rather than trying to understand the approach of the discipline to the understanding of the world. Students who have only experienced the study of science as a collection of facts do not realize the power of reasoning and need more guidance in this cognitive domain. Providing efficient exercises that elicit both student cognition and self-efficacy of answering can be one way to explicitly show students the value of complex reasoning in geologic science.

It is suggested that the approach to the study of introductory historical geology be presented as often as possible as a problem solving opportunity. Like in the case of this self-assessment assignment, students can be given a scenario to analyze and to elaborate upon. In addition, more attention should be dedicated to helping students explain the concept of time on a large scale (i.e., deep time) thereby giving them opportunities to reason and solve problems about this subject.

A key element in building the understanding of geologic time is the exposure to geologic structures in their natural settings, where students can connect their perception of space to the temporal understanding (Dodick and Orion, 2003a). A large scale experiment conducted on perception of geologic time aided by direct field experience at the Trail of Time (Grand Canyon National Park) provides support that pedagogy of deep time is more effective if presented in a spatial context (Karlstrom *et al*., 2008). Unfortunately the opportunity to do field-based activities in large lecture classes is almost nonexistent. Learning about geologic time in a lecture-based course is commonly based on two-dimensional representations. A weak understanding of the meaning of geologic time could also explain the "catastrophist" tendency that most students adopted to explain in their own words the presence of the unconformity. A correct understanding of geologic time allows for a better understanding of the origin and evolution of the planet and life on it at the scale at which nature works, which provides students with a reference frame for understanding and advocating for scientific views (Zen, 2001).

As for how the role of self-efficacy played in the learning experience proposed in this article, the first benefit for the students was to become aware of their own knowledge with regard to specific concepts. This exercise garnered more awareness of declarative knowledge than a general review where answers are given immediately to the students. It was hoped that students could see this exercise as a tool for learning, with no penalty for acknowledging lack of confidence in an answer. Students were encouraged to use this exercise as an opportunity to address specific mistakes to both improve exam results and to build confidence in their own knowledge. Unlike giving out a mock set of exam questions with a key, this active and reflective activity guided the students to an awareness of a deeper level of learning. Because the self-assessment exercise carried very specifically targeted goals for each questions, it was easy for the students to target the weakness of specific cognitive loads, rather than undergoing massive comprehensive reviewing.

The success of this type of activity for very large classes was an incentive for both instructors and students to carry out similar projects. The authors are designing a modification of this exercise, a set of increasingly complex columns to include other aspects of stratigraphic analysis including facies change and interpretation of sedimentary environments. Efficient and easily accessible activities used by instructors of large geology classes can help to increase cognitive levels of learning while not overwhelming students or adding unmanageable amounts of work to the instructor. Success in learning the complex concepts in geoscience relies on the balance between an effective amount of time commitment and rich questions that elicit student learning for the purpose of reflection and improvement.

ACKNOWLEDGMENTS

The authors are thankful to Dr. Stacey Verardo, Dr. Germana Manca, and Dr. Giuliana Mele for the thought provoking conversations and the feedback relative to the exercise design and validation. The authors also are thankful to the students who participated in this exercise.

American Association for the Advancement of Science (AAAS), 1989, Project 2061: Science for all Americans: New York, Oxford University Press, 272 p.

American Association for the Advancement of Science, 1993, Benchmarks for scientific literacy, New York, Oxford University Press, 448 p.

Anderson, L.W., Krathwohl, D.R., Airasian, D.W., Cruikshank, K.A., Mayer, R.E.,Pintrich, P.R., Raths, J., and Wittrock, M. C., 2003, A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives: New York, Allyn and Bacon, 336 p.

Ashcraft, M.H., 2002, Math anxiety: Personal, educational, and cognitive consequences: Directions in Psychological Science, v. 11, p. 181–185.

Ault, C.R., 1982, Time in geological explanations as perceived by elementary school students: Journal of Geological Education, v. 30, p. 304–309.

Baddeley, A.D., and Della Sala, S., 1996, Working memory and executive control: Philosophical Transactions of the Royal Society of London, v. 351, p. 1397–1404.

Bandura, A., Caprara, G.V., Barbaranelli, C., Gerbino, M. , and Pastorelli, C., 2003, Role of affective self-regulatory efficacy in diverse spheres of psychosocial functioning: Child Development, v. 74, p. 769–782.

Black, A.A., 2005, Spatial ability and Earth science conceptual understanding: Journal of Geoscience Education, v. 53, p. 402–414.

Bloom, B.S., Engelhart, M.D., Furst, E.J., Hill, W.H., and Krathwohl, D.R., 1956, Taxonomy of educational objectives: Handbook I, the cognitive domain: New York, Longman, 156 p.

Dodick, J., and Orion, N., 2003a, Cognitive factors affecting student understanding ofgeologic time: Journal of Research in Science Teaching, v. 40, p. 415–422.10.1002/tea.v40:4.

Dodick, J.T., and Orion, N. 2003b, Measuring student understanding of "deep time": Science Education, 87(5), p. 708–731.

Dodick, J.T., 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 theEarth: Geological Society of America, Special Paper 413, p. 77–93.

Duesbury, R.T., and O'Neil, H.F. Jr., 1996, Effect of type of practice in a computer-aided design environment in visualizing there-dimensional objects from two-dimensional orthographic projections: Journal of Applied Psychology, v. 81, p. 249–260.

Dweck, C.S., and Leggett, E.L., 1988, A social-cognitive approach to motivation and Personality, Psychology Review, v. 95, p. 256–273.

Gilbert, J.K., 2008, Visualization: An emergent field of practice and enquiry in scienceeducation, in Gilbert, J.K., Reiner, M., and Nakhleh, M., eds., Visualization: Theory and Practice in Science Education: Dordrecht, Springer, p. 3–24.

Ishikawa, T., and Montello, D.R., 2006, Spatial knowledge acquisition from direct experience in the environment: Individual differences in the development ofmetric knowledge and the integration of separately learned places: Cognitive Psychology, v. 52, p. 93–129.

Kali, Y., and Orion, N., 1996, Spatial abilities of high-school students in theperception of geologic structures: Journal of Research in Science Teaching, v. 33, no. 4, p. 369–391.

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

Lemke, J.L., 1990, Talking science: Language, Learning and Values: Westport, CT, Ablex Publishing.

Libarkin, J.C, Anderson, S., Dahl, J., Beilfuss, M., Boone, W., and Kurdziel, J., 2005, Qualitative analysis of college students' ideas about the Earth: Interviews and open-ended questionnaires: Journal of Geoscience Education, v. 53, p. 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, v. 55, no. 5, p. 413–422.

McConnell, D.A., Steer, D.N., and Owens, K.D., 2003, Assessment and active learning strategies for introductory geology courses: Journal of Geoscience Education, v. 51, no. 2, p. 205–216.

McManus, D.A., 2002, The two paradigms of education and the peer review of teaching: Journal of Geoscience Education, v. 49, p. 423–434.

McPhee, J., 1982, Basin and range, New York, Farrar, Straus, and Giroux, 240 p.

Miller, M.G. 2001, Regional geology as a unifying theme and springboard to deep time: Journal of Geoscience Education, v. 49, no. 1, p. 10–17.

Pinet, P.R., 1995, Rediscovering geologic principles by collaborative learning: Journal of Geological Education, v. 43, p. 371–376.

Pintrich, P.R., and De Groot, E.V., 1990, Motivational and self-regulated learning components of classroom academic performance: Journal of Educational Psychology, v. 82, p. 33–40.

Prothero, W.A., Jr., 2000, Keeping our focus; a perspective on distance learning and the large introductory science class: Computers in Geosciences, v. 26, p. 647–655.

Schoon, K.J., 1992, Students' alternative conceptions of Earth and space: Journal of Geoscience Education, v. 40, no. 3, p. 209–214.

Schunk, D.H., 1991, Self-efficacy and academic motivation: Educational Psychologist, v. 26, p. 207–231.

Schunk, D.H., and Zimmerman, B.J., 1994, Self-regulation of learning and performance: Issues and Educational Applications: Hillsdale, NJ, Erlbaum.

Thomas, R.C., 2001, Learning geologic time in the field: Journal of Geoscience Education, v. 49, no. 1, p. 18–21.

Tobias, S., 1990, They're Not Dumb, They're Different: Stalking the Second Tier: Tucson, AZ, Research Corporation, 93 p.

Tuckman, B.W., and Sexton, T.L., 1990, The relation between self-beliefs and self-regulated performance: Journal of Social Behavior and Personality, v. 5, p. 465–472.

Wagner, J.R., 2000, Sneaking mathematical concepts through the back door of the introductory geology classroom: Mathematical Geology, v. 32, no. 2, p. 217–229.

Wiggins, G., and McTighe J., 2005, Understanding by Design: expanded 2nd edition: Upper Saddle, NJ, Pearson.

Zen, E., 2001, What is deep time and why should anyone care?: Journal of Geoscience Education, v. 49, p. 1–5.

Zimmerman, B.J., and Bandura, A., 1994, Impact of self-regulatory influences on writing course attainment: American Educational Research Journal, v. 31, p. 845–862.

Zimmerman, B.J., and Martinez-Pons, M., 1990, Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use: Journal of Educational Psychology, v. 82, p. 51–19.