# A NEW CENTURY FOR GEOSCIENCE EDUCATION RESEARCH

Prepared for the National Academies
Board on Science Education
Committee on the Status, Contributions, and Future
Directions of Discipline-Based Education Research

January 2011

Michael D. Piburn, Arizona State University*
Kaatje van der Hoeven Kraft, Mesa Community College
Heather Pacheco, Arizona State University

*Inquiries about this work should be directed to mike.piburn@asu.edu

## Introduction

This review of research in geoscience education has been conducted at the request of the Board on Science Education of the National Research Council. The challenge was to identify a representative group of recent studies, to summarize typical research methods, to categorize the studies by type and summarize the results, to state the conclusions we have reached, and to make recommendations for further work. Because of the breadth of this review, its depth is necessarily limited. We have not included older articles, those from outside the geosciences, those from primary school settings, or opinion pieces. Our apologies if we have omitted a body of work that the reader might consider critical, or a favorite piece of research that is thought to be seminal.

## Trends in Geoscience Education Research

There are two groups of scholars currently conducting research in geoscience education. The first are college and university faculty who are interested in teaching. They disseminate the results of their work primarily by presentations at the National Association of Geoscience Teachers (NAGT), which meets concurrently with the Geological Society of America (GSA), and by publication in the Journal of Geoscience Education (JGE). The second is a more diverse set of geoscientists, science educators and cognitive scientists who are interested in cognitive issues in teaching and learning. They present their work at meetings of the American Educational Research Association (AERA), the National Association for Research in Science Teaching (NARST), and a variety of other professional associations, and publish in journals such as Cognition and Instruction, International Journal of Science Education (IJSE), Journal of Science Education and Technology, Journal of Research in Science Teaching (JRST), and Science Education. While there is overlap between these groups, the literatures that they have created are relatively distinct.

Of the research studies chosen for review here, most were published or presented during the period 2001-2010, and two are in press. Most are from the JGE, GSA Special papers, or were presented at GSA meetings. The rest are from a variety of other sources. It is clear that publication of geoscience education research is dominated by JGE and geoscience organizations.

Perkins (2004) reviewed more than 300 articles published in the JGE between 1998 and 2004, and concluded that "209 are success stories…describing some sort of innovative project. 165 are about college education, 38 deal with primary or secondary education." His rating of these articles is shown in Table 1.

Table. 1 Research Ratings of JGE articles (Perkins, 2004)

| RATING | ASSESSMENT | % OF ARTICLES |
|---|---|---|
| 0 | Failed to mention anything about how the project affected student/participant learning | 21% |
| 1 | Included comments or assertions about learning but gave no evidence in support | 51% |
| 2 | Gave results of evaluations or some other kind of data to support conclusions about learning | 12% |
| 3 | Included a complete and well thought out assessment | 10% |

The impressions gained during this review were quite similar to those documented by Perkins.

Although JGE has primarily been a publication through which college and university geoscientists share ideas for creative teaching methods, there does appear to have been a noticeable increase in the frequency of research articles within the last 10 years. There is also an increase in both the JGE and the wider literature in research papers on the subject of cognitive issues in teaching and learning.

There is no special pattern to the research reported here. Sample size ranges from four (Riggs, et al., 2009a) to more than ten thousand (Braun, et al., n.d.). Instruments may include questionnaires, interviews, concept inventories, drawings, or grades and other in-class performance, and variables may be either qualitative or quantitative. Designs may be descriptive, comparative, correlational (including regression analysis), or quasi-experimental. Comparison groups may be students in different courses of study, pre- and post-tests or control and experimental groups. Many of the studies are difficult to categorize because they involve a variety of instruments, methods and comparisons. There are no truly experimental studies, using pre- and post-tests with randomly assigned control and experimental group subjects.

This presentation is organized into five categories; conceptual frameworks, the introductory course, field based coursework, affordances and constraints, and the affective domain. The category of conceptual frameworks is dominated by descriptive studies, with guiding questions but no hypotheses. Studies of the introductory course include descriptive, quasi-experimental, and correlational designs. Studies of field based coursework are primarily descriptive. Among studies of affordances and constraints there are descriptive, comparative, correlational and quasi-experimental examples. Studies of affect are descriptive or comparative.

## Conceptual Frameworks

Many college science teachers became aware of the profound difficulties that their students experience with fundamental concepts in science when they saw “A Private Universe” for the first time. In this film, graduating Harvard University students were unable to explain, among other things, the origin of the phases of the Moon or Earth’s seasons. Atwood & Atwood (1996) pursued this idea with interviews of 49 college students who were preparing to be elementary teachers, and found only one who was able to give scientifically correct answers. An on-line survey of more than 10,000 High School teachers (Braun, et al., n.d.) revealed that fewer than 40% correctly responded to a question about Earth’s orbit around the sun. The remainder of those who completed the survey chose from among three alternative answers that are consistent with the common idea that the seasons are the result of changing Earth-sun distance.

One of the most robust bodies of science education literature during the late 20$^{th}$ century was about the "misconceptions" that students bring to their science classes. Although some relevant papers had been published earlier, interest in the topic was enhanced by an international seminar hosted by Joseph Novak in 1983 at Cornell University. Fifty-five papers were presented and 118 people registered for the seminar. This was followed by three additional meetings in 1987, 1994 and 1997. Research on what were later called naïve conceptions, or alternative frameworks, drove a new body of inquiry into worldview theory and conceptual change teaching.

By 1991 the bibliography on student alternative frameworks that was maintained by Helga Pfundt and Reinders Duit contained more than 2,000 entries, of which only 27 could be clearly identified as targeting geoscience concepts (Smith, et al., 1999). Results for other disciplines are shown in Table 1.

Table 1. Percentage of articles about misconceptions that can be classified by subject area (Smith, et al., 1999)

| | n | % |
|---|---|---|
| Physics | 650 | 57 |
| Chemistry | 290 | 25 |
| Biology | 175 | 15 |
| Earth Science | 27 | 2 |

The area in which the greatest amount of research about student misconceptions has been conducted is physics. More is known about the knowledge base of physics students than in any other field, and the level of sophistication about teaching models is very high. Chemistry and biology have also developed a substantial knowledge base about student misconceptions and constructivist teaching. In contrast, surprisingly little is known about misconceptions in geosciences. That appears to be changing now, but more work needs to be done. Without a deep understanding of the ideas students bring with them to their classes, and the barriers to learning that misconceptions present, progress in research into teaching and learning will be seriously impeded.

Three aspects of geoscience: Earth's crust, Earth's interior and geologic time were assessed by a series of questionnaires completed by 265 college students as part of a larger program to develop a conceptual inventory in geoscience (Libarkin, et al., 2005). A sample of 105 students was then chosen for interview. The study revealed many misunderstandings about geological concepts. For instance, fewer than 50% of the students believed that Earth was 4.5 billion years old, and 11-13% thought that Earth was less than 100,000 years old. Alternative explanations for the origin of earthquakes involved heat, temperature, climate, weather, people and animals. Student description of Earth's interior were based on analogies like a "dartboard," a "jawbreaker," or a "baseball." Conceptual barriers to geoscience learning were the subject of questionnaires and interviews of college students by Kortz & Murray (2009). They identified the following topics that students have difficulty with: time; changing Earth; large spatial scale; bedrock; materials; atomic scale, and; pressure

In addition to questionnaires and interviews, concept maps have been important tools for revealing misconceptions. Rebich & Gautier (2005) had students make concept maps before and after a 3-week mock summit on climate change. They found that students had inappropriate models of radiative processes, and thought that global warming resulted from sunlight entering the atmosphere through a hole in the ozone. They were also confused about aerosols and greenhouse gases. The authors state that the post-instruction concept maps revealed large increases in concepts and useful links, and a decrease in misconceptions. Engelbrecht, et al. (2005) examined student-drawn concept maps over a two-semester sequence of introductory geology lectures, and found that there was an increase in geological concepts identified, but "a disproportionately small increase in integration of those concepts into frameworks of understanding."

In a review commissioned by the Board on Science Education, Libarkin (2008) identified five concept inventories in physics, three in astronomy and six in biology, but only one in the geosciences.

That was the Geoscience Concept Inventory (GCI), which she and a group of colleagues created. The GCI was developed through a cycle in which: a committee of college and university geosciences teachers examined the curriculum to identify content; students were interviewed in order to reveal misconceptions; a trial version of the instrument was created and administered; test statistics were created and the instrument revised; the new instrument was presented to students in an interview process, and revised once again, and; the subsequent version of the GCI was used in a series of pilot studies.

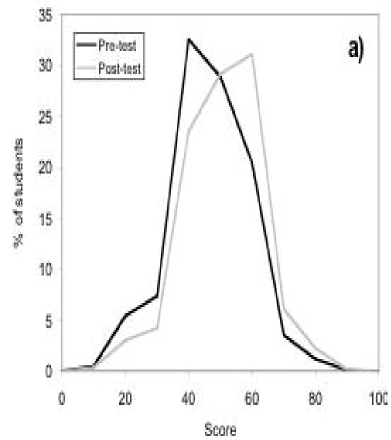
Libarkin argued that, because a test such as the GCI violates many of the assumptions of parametric statistics, and in particular the assumption of the interval nature of the items, it was much more appropriate to use Item Response Theory in analysis of the GCI. As a result, scores on the GCI are transformed in response to item difficulties, so that researchers may choose sub-tests of selected items that they believe reflect their own teaching objectives, and still compare results across separate studies with confidence that the scaled means and individual scores are completely comparable. It is also possible to compare gains from pre- to post-test as a percentage of the total possible improvement. The creation of the GCI has been an important first step in facilitating an important body of research in the field.

## The Introductory Course

Many articles in the Journal of Geoscience Education are devoted to instruction for introductory college courses. These range from tips about how to teach a difficult topics to descriptions of full curricula to research designed to demonstrate the superiority of one method over another. Generally presented in isolation, without meaningful connections to broader efforts or initiatives within geoscience education, the bulk of this literature does little to build a repertoire of best practices. However, within the last decade a body of work has begun to emerge that is more coherent and that directly addresses the question of teaching effectiveness in terms of student achievement and learning gains.

The research reviewed here has been deliberately arranged in the order of increasing student engagement and achievement. Thus, it presents a circumstantial case for the relationship between these two variables. However, from quasi-experimental studies alone it is not possible to estimate the amount of interaction that is required for meaningful improvements in student performance nor the variance shared between the two variables. These are addressed by a study discussed later in this paper in which an observational protocol is used to directly measure the interactions that occur within classrooms and compare them to student learning.

Two major studies examined the efficacy of the introductory course over a variety of contexts and approaches. In the first case, two GCI test versions of 29 items, 11 items common to each, were distributed to the instructors of 43 courses at 32 institutions in 22 states (Libarkin & Anderson, 2005). The topics of these courses included introductory physical and historical geology, oceanography, environmental science, and specialty topics. In the second study, one GCI version with four anchor items and one item from each difficulty category was administered to each of six sections of introductory



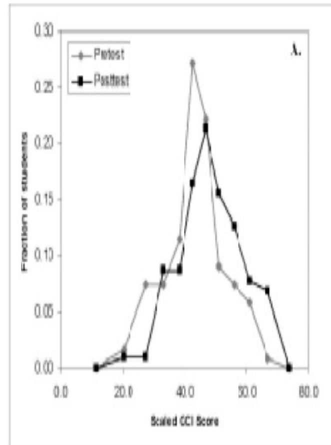




Figure 1a) Libarkin & Anderson (2005)      Figure 1b) Petcovic & Ruhf (2008)

physical and historical geology (Petcovic & Ruhf, 2008). In both studies, the GCI was administered as pre- and post-tests. The sample size in the first instance was 2493 (pre-test) and 1295 (post-test) students. In the second study, the sample consisted of 122 (pre-test) and 102 (post-test) students. The results of these studies are shown in Figures 1a and 1b (with permission of the authors).

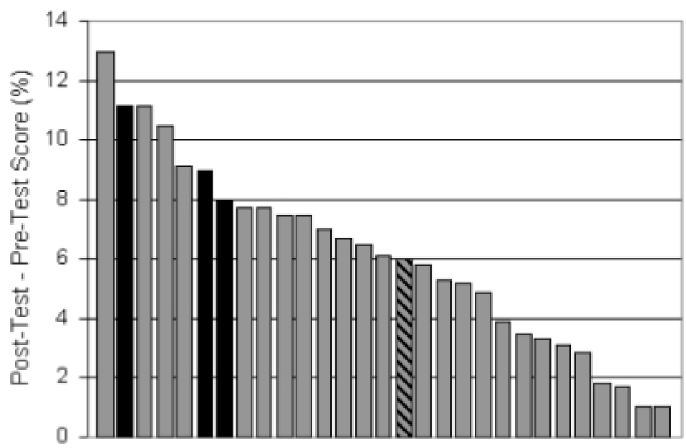
It is evident on inspection that there is little difference between pre- and post-test distributions of the samples in these two studies. Libarkin states that as many as half of the lecture sections studied showed no significant improvement in student scores, and that matched-pair t-tests indicated that students with

- the lowest initial scores showed the *largest gains*
- moderate initial scores showed *little improvement*
- the highest initial scores showed *no improvement*.
- after instruction, students were, on average, *familiar with only half the concepts on the CGI.*

The fact that these remarkable findings are by no means unique to the geosciences is discussed in greater detail later in this paper.

Two studies have been published in JGE in the last five years testing a model called Just-in-Time Teaching (JiTT). In this style of teaching, which has been popularized by Gregor Novak, teachers post "warm ups" on a web site that students must visit before class. Teachers review the web-postings and organize their classes based upon student responses. Luo (2008) compared the performance of groups of students who completed warm-ups with those who did not and found differences in favor of the first group. However, these results are difficult to interpret because of the unusual choice of comparison groups. Linneman & Plake (2006) compared control and experimental classes in which the treatment was JiTT and the control was traditional lecture. A step-wise multiple regression showed that grouping by control and experimental conditions did not predict post-test scores. The only variables that predicted exam averages were pre-test scores on a science measure and the score on a measure of attitude toward the subject administered at the beginning of class.

Several studies have examined the effectiveness of interactive lectures. Clary and Wandersee (2007) tested a model of integrated, thematic instruction in introductory geology lecture. The experimental condition created a “mini-lab” involving petrified wood. A station was set up in front of the classroom with specimens of petrified and modern wood, and students visited the station before and after class. Following class students participated in electronic discussions and visited the university's geology museum. They were allowed to complete an extra-point assignment in which they investigated a site where petrified wood could be viewed and they plotted these locations on a map at the museum during their visit. The authors had mixed findings, with one comparison showing the experimental group superior to the control group on the final exam and a Petrified Wood Survey, and the other comparison showing no significant differences on the final exam. Kortz, et al. (2008) followed brief introductory lectures with interactive sessions during which student groups discussed ideas and completed worksheets based upon the misconceptions literature. They found significant increases in both quiz and GCI scores between pre- and post-test. McConnell, et al. (2003) used a variety of "inquiry" strategies during lecture and were able to show significant differences between




Striped column=control group
Black columns= interactive lectures
Gray columns=comparison classes from other colleges or universities

Figure 2. McConnell, et al., 2006

control and experimental groups, with "inquiry" groups receiving higher exam scores. In a more carefully controlled study, McConnell, et al. (2006) were able to show that interactive courses in which students were given brief introductory lectures, followed by concepttests with student responses triggering immediate feedback and adjustment of instruction led to a substantial improvement in GCI scores. When compared with 25 other classes reported by Libarkin, the experimental groups were among the most successful (Figure 2, with permission of the authors). In a study of understanding of the seasons, students in the experimental group were given hand-held ball and stick models and flashlights. Their task was to work through a modeling activity in small groups, while the control group was given a lecture on the seasons (Gray, et al., 2010). Both experimental and control student groups were engaged in peer learning and concepttests. Normalized gain scores in the experimental group were larger than those in the control group, demonstrating the value of including physical objects in inquiry activities.

The studies just reviewed provide great insight into the relationship between teaching strategies and student achievement. They cannot reveal the full spectrum of events in the classroom that may also be important. In particular the interactions among students and between students and teacher may vary even among classes that use the same model of instruction. The only way to capture this dynamic is by classroom observation. Such information is contained in a report by Budd, et al. (2010) at the most recent meeting of the Geological Society of America. They used the Reformed Teaching

Observation Protocol (RTOP) to assess the level of interactivity in the classroom. The RTOP was developed at Arizona State University as part of the evaluation of the Arizona Collaborative for Excellence in the Preparation of Teachers, a project funded by the National Science Foundation (Piburn, et al., 2000; Sawada, et al., 2002). It is one of the only observational protocols for which there is substantial information on reliability and validity. It predicts student achievement in high school and college science and mathematics classes with a remarkably high degree of correlation, and has high face validity. For these reasons it has been widely used in the evaluation of funded projects at the state and national levels as well as in the research of individual scholars.

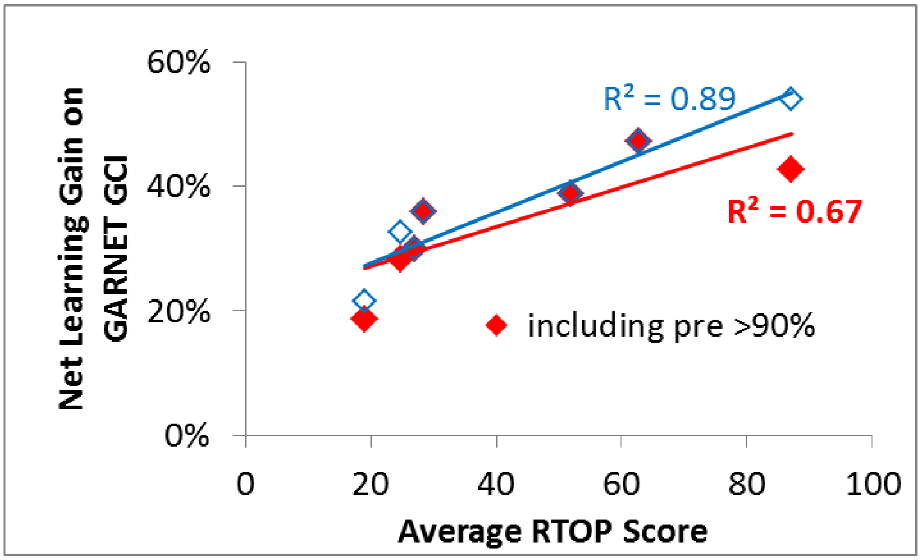


Figure 3: RTOP score for individual instructors vs. average normalized learning gain for eight classes in GARNET study

Eight instructors from five institutions who were teaching introductory physical geology reported modified GCI data from 407 students. Every participating instructor was observed at least twice with the RTOP. Learning gains ranged from 18.6% to 47.4% and RTOP scores ranged from 19 to 87. The authors report an r-squared value of 0.67 between the two variables. Furthermore, if the data are corrected for effects resulting from high initial concepttest scores of some students, the r-squared value rises to 0.89 (Figure 3, with permission of the authors). From this, they conclude that:

> "Learning gains are not high for traditionally taught courses – lecturing and limited use of clickers produced average learning gains <30%. Instructors with the highest RTOP scores (> 60) and learning gains > 50% used multiple strategies to engage students with content during class and force students to reflect on their understanding of critical concepts during and outside of class. If one is serious about educating students, the results indicate significant and substantial changes are needed beyond the traditional lecture. Just adding a few activities or interventions will not have great effect."

The growing body of research reporting similar findings about the relationship between teaching method and achievement gains leaves little question that the traditional lecture is the *least effective teaching method in the geosciences*. Many academic science departments are now considering quite different models of instruction, including cooperative learning, modeling instruction, and problem-based learning as alternatives.

## Field-Based Coursework

The geoscience community has long expressed support for field experiences as part of the curriculum. A common metric for measuring their success has been interviews and surveys, and students typically self-report that they like working in the field and that they learn more. Only recently has there been a more serious effort to validate the effectiveness of the field experience. Several main types of work in the field are being examined, including introductory-level experiences, field camps for majors and the development of problem solving in the field. These efforts seek to identify the nature and measurable value of field experiences and to establish an understanding of the processes and pathways associated with increasing knowledge and expertise in field settings at different levels.

Huntoon, et al. (2001) report on a small two-week long introductory-level intensive course in which a mix of 12 in-service teachers and $1^{st}$ and $2^{nd}$ year undergraduates were enrolled. The course goals were designed around both cognitive and affective objectives. Although the small sample size prevented statistical analysis, the authors found learning gains, an increased willingness to attempt more challenging problems, and an increase in interest in geosciences. Tretjniak & Riggs (2008) studied 36 pre-service elementary teachers in an earth science course. They found that virtual field trips can be sufficient for increasing basic geoscience knowledge, but that true field experiences are needed for student learning of concepts that entail hands-on work and spatial visualization. While there was no statistical difference in their course exams, follow-up interviews revealed that students with tangible field experiences had a deeper conceptual understanding of the content over those who just participated in a virtual field trip. In another qualitative analysis, field trips were identified as possible critical incidents, unique to the geosciences, that impact student retention in geoscience programs compared to other STEM disciplines (Levine, et al., 2007).

In the most definitive study for the introductory population, Elkins & Elkins (2007) examined the effectiveness of field-based delivery of the beginning course. There are two such programs: the University of Georgia's Interdisciplinary Field Program (IFP) and Bowling Green State University's "GeoJourney." In these courses, students travel for 23,335 kilometers (14,500 miles), camping across the U.S. for nine weeks. Both programs offer a full semester of general education credit in geosciences, Native American/culture studies and environmental studies/ecology. The authors report pre-post-test effect sizes with the GCI of 0.93 to 1.53 standard deviations, with an average of 1.08. These are quite large when compared with the average of 0.36 standard deviation reported by Libarkin for all classroom-based courses.

Research about the role of field camps in learning by geoscience majors is scarce. Using a qualitative approach Stokes & Boyle (2009) found that students participating in an upper-division field camp had positive gains in motivation that they propose leads to an environment that fosters deeper approaches to learning resulting in greater cognitive gains. Key quantitative findings support the beliefs about the value of field experiences.

Studies of problem solving and development of field expertise have been focused on student behaviors rather than instructional methods. Riggs et al. (2009 a,b) studied field-based geologic problem solving in an advanced field geology course held in the San Gabriel Mountains in California. Students were outfitted with GPS tracking units while they independently developed a geologic map. The field notes, exams, and navigational characteristics provided insights into problem solving in the field. The authors built upon findings from the research field of Naturalistic Decision Making (NDM). This identifies expert problem solvers as those who employ pattern recognition to make educated guesses about a situation and guide subsequent decisions. Most of the advanced students recognized important geologic features but low performers exhibited a lack of strategic navigation decisions, reflecting their inability to identify and test multiple interpretations of data. These findings both align with paradigms of expertise development and exemplify the unique challenges associated with geoscience-specific expertise development.

Petcovic et al. (2009) contextualize their “novice to expert” work with transdisciplinary findings about the development of expertise through: recognition of patterns in information due to familiarity with the subject; thinking revolving around big ideas, rather than memorization and recall; knowledge that is interrelated (chunking); ability to self-monitor and make decisions based on understanding (metacognition); flexible thought process; and awareness of complexities and abstractions. They studied three professional field geologists and four undergraduate field students, representing a continuum of prior geologic study and field mapping experience ranging from none to more than ten years. These participants mapped two field sites in the Upper Peninsula of Michigan. Outfitted with audio recording equipment, a camera and GPS trackers, they were given as much time as they liked to create bedrock maps of each site, recording their thoughts during the activities and taking photos of interesting geologic features. After the mapping exercises, participants clarified their maps through semi-structured interviews, responding to a set of open-ended probes, and answered interview questions about their strategies, decisions, confidence and perspectives post-mapping. This mixed-methods approach allowed the capture of fine details of participant thought processes during the activity. Findings from the “novice” data include complex movement tracks and high levels of backtracking, greater difficulty with determining location on the map, more distractions to the mapping task, complete lack of mental model synthesis from their observations and, subsequently, no novice testing of models against observations. Findings from the “experts’” include economy of movement, little backtracking, time spent at significant geologic features like contacts, expressions of degrees of certainty based on observations that helped to shape their mental models, well-defined (albeit differing) methods for developing and testing models and clear articulations about reasoning, suggesting higher levels of metacognition.

## Affordances and Constraints

Cognitive studies in science education explore the influences on learning of the interactions between student psychological characteristics and the unique intellectual demands in the disciplines. Among these are affordances by which some people are better able to learn than others, and constraints that inhibit learning for some. Recent geoscience education literature provides confirmation of well documented findings in science education, including those that consider learner characteristics such as mental ability (Barba & Rubba, 1992), logical thinking (McConnell et al., 2005), inquiry (Apedoe, 2008), and non-scientific beliefs (Bickmore et al., 2009). Although important, these studies emphasize issues that are common to the study of all scientific disciplines, and they remain isolated in the geoscience literature.

Insight into the uniqueness of the geosciences can be obtained from the work of Robert Frodeman, currently the Director of the Center for the Study of Interdisciplinarity and Professor of Philosophy at the University of North Texas. While teaching philosophy at Fort Lewis College in Durango, Colorado, he completed a Master's degree in geology. He served subsequently for eight years as a consultant for the U.S. Geological Survey, and was the 2001-2002 Hennebach Professor of the Humanities at the Colorado School of Mines. His work remains perhaps the most complete and compelling description of how geologists think that is currently available.

Frodeman's view of the science and practice of geology is quite different than that often attributed to other scientific disciplines. It will not be possible here to completely describe his vision. However, it is important to note his opinion that "geologic seeing does not in the first instance consist of a series of deductions and inferences from a set of data," and his thesis that "geology depends upon a type of *visual intelligence* whereby the geologist applies a set of templates that organize sets of marks into a body of significant signs." The task of becoming an expert is to see systems in the rock. "I must look at the outcrop as a *language*, as a system of interrelating and counter-balancing processes, tectonic and climatic activity, Milankovitch cycles, and changes of sediment production and organic productivity. This was the most fundamental difference between my experience of the outcrop and that of the geologists who accompanied me in the field." (Frodeman, 1996).

Following Frodeman's lead, this section on Affordances and Constraints is sub-divided into three sections: Temporal Thinking, Spatial Visualization and, Systems Thinking. These are the areas judged to be most central to the unique thinking in the geosciences, and about which there is a rich research literature.

Temporal Thinking

Frodeman discusses geology as a science that depends upon the principle of uniformitarianism to co-ordinate our modern experiences of Earth processes with those of the distant past, but notes that "there are also inescapable disanalogies between our human experience of time and the vast expanses of geologic time". Time as it is experienced by most of us in our daily lives is a complicated subject and has been studied in detail by cognitive scientists. Geological, or "deep time", is a very different matter. It is of a scale and immensity that is well beyond the imagination of most people. Studies of temporal thinking in the geoscience literature are rooted in the challenges of identifying how people can and do understand the abstract concept of time and how these conceptual frameworks can be measured by researchers.

Evaluating people's understanding of deep time is difficult. Desired outcomes of instruction might include the ability to correctly order events, estimate the percentage of earth history occupied by certain periods, or state the absolute age of events. Furthermore, understanding geological time assumes comprehension of a large number of related concepts such as superposition and original horizontality, processes and rates, uniformitarianism, radioactivity and half-lives, etc. Facing this difficulty, Dodick and Orion (2003a) define two ways of understanding deep time. The first is "event" based, and refers to understanding the absolute and relative sequence of events, and the second is "logic" based and refers to "the cognitive processes undergone by students when solving problems involving geologic time."

One way to estimate understanding of the order and scaling of events in event-based time is to have subjects construct geological timelines (Catley & Novick, 2009; Libarkin, et al., 2007; Trend, 2001). These studies have asked students to order and estimate the age of events such as the origin of the earth, the first occurrence of life, evolution of eukaryotic cells, first fossils, the Cambrian explosion, the origin and extinction of dinosaurs, first mammals, and the first occurrence of humans. From these studies, we can conclude that students have a relatively good understanding of the order of geologic events, but are less able to correctly scale them. Trend argued that judgments of time fell into three categories: extremely ancient, moderately ancient and less ancient. Responses gathered by Catley & Novick clustered into six groups. In the first two, subjects overestimated the period of time between the origin of the earth and subsequent events. This is a well-known error, called "forward telescoping," that is often made by people when asked to describe and order events in their own lives. The third group gave time-lines that were scientifically correct. The fourth and fifth groups, which were much smaller, underestimated the spacing between the origin of the earth and later events. A final group, containing only three subjects, indicated that all events were essentially simultaneous (they were self-professed creationists). Using a similar procedure Libarkin, et al. were able to distinguish several groups of students: those with a "young Earth" viewpoint, a small number who gave scientifically correct estimates, and another similar to the "forward telescoping" group discussed above. It is especially interesting to note the comment by Catley and Novick that there was no apparent relationship between prior coursework and the kinds of timelines constructed by students. They added that "whatever macroevolutionary knowledge the stronger background students (including those who had taken evolution) obtained from their coursework, it was insufficient to help them make sense of the absolute timing or relative spacing of key evolutionary events examined in our study."

Dodick and Orion (2003b) identify another aspect of geologic time called "diachronic," referring to the development of a phenomenon over time. An example would be the sandstone created by a transgressive shoreline, in which an apparently continuous sedimentary deposit actually varies in age from one place to another. In order to examine diachronic thinking, Dodick & Orion (2003a,b) created three instruments: the Geological Time Aptitude Test (GeoTAT) consists of geological puzzles that ask about the temporal relationship between geological strata and their fossils; the Temporal Spatial Test (TST) consists primarily of spatial visualization puzzles, and; the Strategic Factors Test (SFT) presents pairs of three-dimensional representations of outcrops, and asks students to estimate which of the paired outcrops is older. A factor analysis of the GeoTAT revealed three groupings of items interpreted as isolated transformation, isolated temporal organization, and full diachronic schemes. In a series of validity studies, the GeoTAT discriminated between students from different grade levels, and also between those with and without a geology background. The easier puzzles were those involving transformational thinking, and the more difficult one required the use of a full set of diachronic schemes. The correlation between the GeoTAT and the TST was significant but small. Results from the SFT were mixed, and revealed a number of misconceptions.

### Spatial Visualization

Concepts of spatial reasoning find their origins in intelligence testing at the turn of the 20th century. In every analysis, spatial and verbal ability have been found to exist as separate factors of the intellect.  It has also been found repeatedly that scientists have exceptional spatial ability and that there is a relationship between ability in science and spatial reasoning.  Additionally, studies support the notion that spatial reasoning skills necessary for complex thinking in the geosciences can be improved through coursework and training.

There are many articles in the science education literature about the correlation between spatial ability, scientific ability and achievement, but few refer specifically to geosciences. In a recent study by Black (2005) of 97 undergraduate students enrolled in six undergraduate courses in geoscience, chemistry, physics and biology, correlations between a test of Earth Science Concept understanding (ESC) and three tests of spatial ability ranged from 0.34 to 0.52.  Spatial tasks included the Purdue Visualization of Rotations, Group Embedded Figures Test, and the spatial subtest of the Differential Aptitude Test.  Siemankowski & McKnight (1971) reported that levels of spatial ability varied among students in different sciences, with those in the geosciences surpassed only by physics students.  Titus & Horsman (2009) found that there was a relationship between grades and spatial ability in introductory courses, that students in higher level courses had better spatial ability than those in introductory courses, and that students' spatial ability can improve through practice in advanced geology courses.

Liben, et al. (in press) have been interested in the relationship between cognitive factors and very specific tasks that are required of advanced students in the geosciences.  Their analysis reveals that error patterns in the measurement of strike-and-dip are related to difficulties with measures of horizontality and verticality including the water level task and the rod and frame test.  Additional information about the relationship of spatial ability to more advanced geoscience classes comes from the work of Ozdemir  et al. (2004), who studied an upper division mineralogy class.  He reported correlations between scores on the surface development test and achievement on problems, laboratory exercises and the final examination of between 0.52 and 0.60.  He also reported significant improvements on both the cube comparisons and surface development test from pre- to post-test, suggesting that the course itself was an intervention on spatial ability.

Visualizing 3-dimensional structures is a particularly important ability to geologists.  Kastens, et al. (2009) completed a study in which geologists and undergraduate students attempted to reconstruct the geologic structure represented by an array of artificial outcrops.  Subjects were then asked to choose from among 14 possible models the one that they thought best represented the available information. Science students produced almost twice the number of evidence-based claims as non-science students, and students who chose a correct model gave more evidence than those who did not. Challenges included "identifying appropriate observational evidence, combining multiple lines of reasoning, and understanding the scale relationship between candidate models and the full-scale structure."

Piburn, et al. (2005) constructed a quasi-experimental study designed to test the hypothesis that specific training in spatial reasoning would transfer to improved learning in introductory laboratories. They used four laboratory sections taught concurrently during the summer, with two identified as control and two as experimental.  The laboratory itself was situated in a computer-based terrain called *Painted Canyon*, created to resemble as authentically as possible the task of a field geologist.  Students could move around the terrain, exploring it from all points of view, and most laboratory tasks were set within that context.  In the experimental sections, students were removed for 30 minutes to 1 hour at a time to complete a set of computer-based spatial activities involving topographic maps and geological block diagrams. A Geospatial Examination was created for the experiment, and given as pre- and post-test along with the cube rotation and visualization tests. Analysis of Variance revealed that while all students improved on their Geospatial score, improvement among students in the experimental condition was greater than in the control condition, and the gap between females and males that existed at the beginning of the experiment no longer existed at the end. From this, the authors concluded that *spatial reasoning can be taught, and that improved ability transfers to improved performance on closely related achievement measures.*

Systems Thinking

The general ideas of modeling in science, integral to providing evidence for the acceptance of scientific theories, contain close connections to the nature of science as it is practiced in the geosciences. Studies of modeling instruction have tended to focus on the Earth-Moon or Solar systems, and to target misconceptions and improvement of conceptual understandings (Atwood & Atwood, 1996; Bell & Trundle, 2008). Bell and Trundle used Moon watches, the computer simulation *Starry Night Backyard* and a series of Moon phase cards from *Physics by Inquiry* to improve the conceptual understanding of a sample of 50 pre-service teachers. The Journal of Geoscience Education also contains many ideas about teaching with stream tables or analogs of volcanoes, but little research on such methods, and no indication that they involve modeling in the complex sense that is implied here. Unfortunately, many of the systems encountered in the geosciences are much too elaborate to be reduced to simple models that can be easily manipulated in the classroom.

Although science educators from many disciplines agree that systems thinking is an important element of scientific thought, work in this area is not well developed and there are no standardized instruments to be found to measure the ability. However, a small number of studies shed light on its general characteristics and the challenges that we face to measure student abilities in this domain.

The systems concept is itself a difficult one, and efforts to define it characterize the research in the field. All agree, however, that systems consist of interacting sets of phenomena that tend toward some kind of equilibrium. Thinking about systems involves identifying individual elements, the interconnections between the parts of the system and, most importantly, a unifying synthesis that characterizes the system itself. It is not surprising that systems thinking is one of the most cognitively challenging elements of scientific reasoning.

Working with Israeli 8$^{th}$ graders who were studying the water cycle, Assaraf and Orion (2005) derived the idea that systems thinking involves "several sequential stages arranged in a hierarchical structure." Using a combination of ten qualitative and quantitative instruments, they found a high correlation between the number of systems components identified and connections among the components. However, students had difficulty stating the dynamic relationships of the system. Even after instruction, "66% of the students identified groundwater as a static, subsurface lake," and 62% said that the water cycle has a beginning and an end. The authors concluded that the stages in systems thinking involved: the ability to identify system components and processes; the ability to identify dynamic relationships between system components; the ability to understand the cyclic nature of systems, and; perception of temporal elements of the system. Among the students they studied, 70% were proficient at the first level, 50% at the second, 30-40% at the third, and only 10-30% at the fourth.

After interviewing students in her college course on System Analysis of the Earth, Raia (2005) tentatively proposed two common mental models. The first was "a static view that did not take into account temporal and spatial scales of our planet," and the second was a "linear mono-causal approach to Earth processes." She interviewed sixteen undergraduates, ten of whom were earth science majors and six were engineering students. Fourteen of these students gave evidence of a linear-causal approach to causality, looked for unique causes, and never considered observed patterns as emergent properties of dynamic systems.

The idea that systems thinking is hierarchical is common in the literature. Libarkin and Kudziel (2006) proposed four levels of understanding (figure 3, with permission). Using this taxonomy to code interview transcripts of 61 college students, they concluded that students are predominantly at the

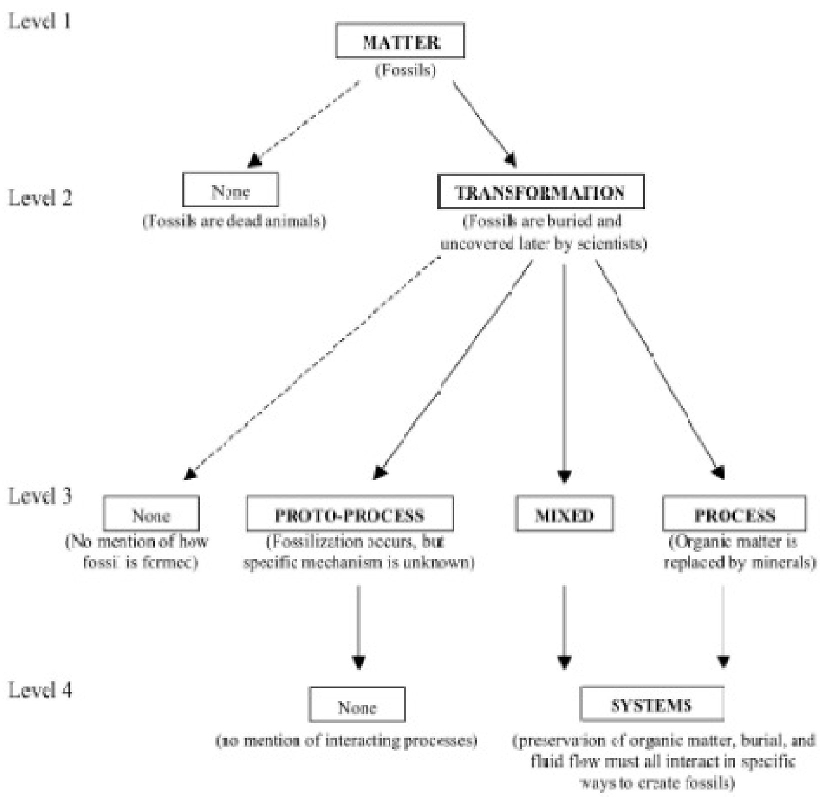



Figure 3. Levels of understanding of systems (Libarkin & Kurdziel, 2006)

Proto-Process ontological level, "acknowledging that a process must exist to cause a Transformation, including mention of a process-related word, such as subduction, but without a clear explanation of what subduction actually is." Those thinking at this level "generally do not or cannot explain underlying processes responsible for geologic phenomena." A startling result of this study was that only one of the 61 students interviewed gave evidence of the highest level of systems thinking. From this, Libarkin & Kudziel concluded that *most students are not ready for geoscience instruction at the start of an introductory course*.

## Affect in the Geosciences

The role of the affective domain in the context of teaching and learning are gaining attention in geoscience education research. The connection between cognitive gains and the underlying motivation, attitudes and self-efficacy that students bring to and develop in their coursework is increasingly recognized as a powerful key to understanding achievement and perseverance.

Recent work funded by NSF grants (McConnell et al., 2009, McConnell, et al., 2010) is an important start in gaining insight into the attitudes and motivations of introductory students. Those who have low prior knowledge, but high self-efficacy, are just as likely to be successful as those students who have high prior knowledge and high self-efficacy. Research with women and minorities indicates that these populations enter into the classroom with significantly lower levels of self-efficacy and are more likely to use peers as a resource for studying (van der Hoeven Kraft et al., 2010). The use of interactivity in the classroom may be just as important from an affective perspective as from a cognitive one.

The affective components of undergraduate research and field experiences are also important. Maguire (1998) used post-field experience interviews and surveys of 100 students from a small university in the United Kingdom. They found that students consistently valued the experience because of the practicality of fieldwork, the opportunities for socializing and exposure to real situations not accessible by reading a textbook. Huntoon, et al., (2001) reported a positive impact of a field experience on interest in geosciences as well as a high value associated with the development of peer-relationships.

Boyle, et al., (2007) measured levels of anxiety, motivation, deep vs. surface learning, and performance in a field experience. Key findings included elevated anxiety prior to the field experience and a subsequent decrease in anxiety and increase in confidence. They concluded that fieldwork is a successful way to engage the spectrum of students in effective approaches to learning because of its inherent socially integrative nature and remarkable effectiveness in increasing student confidence in group work. Stokes and Boyle (2009) followed this work by examining experiences in an extended field class. They employed a modified version of the Boyle et al. (2007) instrument, as well as interviewing and observing students. Overall, attitudes became increasingly positive through the course of fieldwork. One of the highest-ranking items from their survey was group work. However, while students saw value in fieldwork for increasing knowledge and usefulness for a degree, they did not see a link to use in future classes. Most students indicated that learning in the field was more interesting than other contexts and they exhibited an increase in desire to learn. The other great change was in the confidence to work independently. Social and cultural events were the most highly rated events, and they helped with maintaining motivation throughout the experience. Stokes and Boyle propose that because student's affect increased positively, students are more likely to engage in deep learning.

In these studies, it is often the non-academic experiences dimensions that create an atmosphere conducive to learning. While the field is not the only place where these interactions occur, it is clearly one that fosters these types of experiences more readily. The same may be said for undergraduate research experiences. Students who participated in undergraduate research (n = 30), findings indicate that fun and playfulness were critical to the success and persistence of participants in addition to their learning of the content (Jarrett and Burnley, 2010). Jarrett and Burnley (2010) reported on an undergraduate research summer project in which students said that their most memorable experiences were due to the fun that they experienced and the collegiality that resulted from the experience. All of these factors become more apparent and important with women and possibly minorities in the classroom, in the field, or doing research.

Vislova, et al. (2010) identified gender differences in attitudes associated with geoscience courses. Their study of 539 males and 607 females from 14 classes in seven institutions ranging from research universities to community colleges used a version of the Motivated Strategies for Learning Questionnaire as a pre/post course attitude assessment. Incoming students showed significant gender differences in cognitive and metacognitive strategies, self-efficacy, test anxiety, time and study environment and effort regulation. Females report lower self-efficacy and higher test anxiety than their male peers. Findings of outgoing attitudes include decreases in female self-efficacy, task value and critical thinking while males were found to have decreases in self-efficacy but increases in intrinsic goals orientation and critical thinking. Females also reported lower likelihood of engaging in future geoscience courses compared to their male peers despite similar course grades.

Gender differences have also been examined in several studies of the field experience. Maguire (1998) found females to be more concerned about the cost associated with field camp. Females also assigned higher value to group work, rated field activities as requiring high levels of fitness, and rated their own fitness lower than that of males. While females report higher pre-trip anxiety, the process of engaging in the field experience decreases their stress and helps them to be equally prepared to engage in deep learning (Boyle et al., 2007; Stokes & Boyle, 2009).

The value of sense of place and place attachment has been well documented in fields outside of the geosciences. Efforts by several researchers in the geoscience education community have resulted in the development of a theoretical framework linking engagement of minorities to these measures of value, appreciation and connection to physical places. Semken and Butler-Freeman (2008) report on findings from a locally-based, trans-disciplinary and cross-cultural course they developed that was an equivalent alternative to the conventional introductory geoscience course. Their subjects included 27 undergraduates who were randomly selected from a pool of volunteers who were originally enrolled in the conventional large-lecture introductory geoscience course. Their instruments were the Place Attachment Inventory (PAI) and the Young Place Meaning Survey (PMS). Comparison data came from a body of PAI post-conventional course data (n = 753) gathered over previous years at the same institution. Findings include significant gains in both place attachment and place meaning for the experimental group. Findings from a recent ethnographic qualitative study of a place-based Earth science course presented to a diverse group of in-service teachers indicate that the approach can enhance personal relevance of the discipline, and appreciation for surrounding geological features, systems, and processes (Semken and Williams, 2008; Williams and Semken, in press).

## Summary

Educational research in the geosciences has a shorter and more recent history than in other content areas. Nevertheless, our work here has convinced us that the last ten years have brought a dramatic improvement, both qualitatively and quantitatively, in the literature of the field. For this reason we have titled this paper *A New Century for Geoscience Education Research*.

Most college teachers assume that their students come to them without prior knowledge and conceptions, but that they are fully ready to learn. Neither of these is true. The work of McConnell, et al. (2005) indicates that no more than 1/3 of all college students demonstrate the ability to use fully abstract thought. This is a result that is so well known in the science education literature that many journals will no longer accept submissions on the topic. Perhaps even more startling is the assertion by Libarkin & Kurdziel (2006) that most students are not prepared to learn at the beginning of an introductory course. Best practice in science teaching suggests that student knowledge and abilities be assessed *before* instruction, and that teaching be tailored to accommodate the results. The idea of giving a test before class begins is completely foreign to most instructors, but we believe that with more familiarity with the literature, and the support of their colleagues, they will become comfortable with the practice.

We recognize that the most startling, and even controversial, aspect of this paper is the review of teaching and student success in the introductory course, and particularly what may be perceived as an attack on the traditional lecture. Major studies in other areas of science education, such as Hake's six-thousand student survey of mechanics test data for introductory physics courses, or the meta-analysis by Springer, Stanne & Donovan of small-group learning in college science courses, support the results found in the geosciences. Craig Nelson (2010) offers his summative analysis of these studies:

> "Hake's meta–analysis for introductory physics (1998) also changed my thinking. Standardized pretests and posttests of conceptual understanding had been used in a variety of introductory courses. For each course, Hake calculated that average normalized gain <g>, as the ratio of the actual average gain in class understanding (posttest mean minus pretest mean) to the maximum possible average gain for that class (100 minus the pretest mean). Traditional lectures produced an average normalized gain of 23 percent. Various forms of structured student–student interaction ("interactive engagement") produced an average of 48 percent. No traditionally taught class came near the mean for interactive engagement. There was comparatively little difference in gain between the worst and best of the standard lecture courses. Effort spent on improving lectures was a waste of time in comparison with that spent on transforming the pedagogy…… A meta-analysis for science and related fields (Springer, Stanne, & Donovan, 1999) found that the average effect of small-group learning would move a student from 50th percentile to the 70th."

These results virtually demand that geoscience educators turn their attention toward the evaluation of alternative instructional methods. We understand the cultural and economic factors that maintain the large group lecture as the dominant mode of instruction for introductory courses, but there is no logic to their continuation in the face of evidence that students are learning little or nothing.

Cognitive issues in teaching and learning have not been the primary concern of college and university geoscience instructors. Those which we have discussed here appear to be more salient in the geosciences than in other disciplines. However, the broader research in cognition and instruction has much to offer to those who are interested in course design and evaluation of instruction. As they become increasingly familiar with this literature, we anticipate that such topics will become more prominent in the educational literature of the geosciences.

While it is true that research in geoscience education has made great progress in the last decade, it is also true that much remains to be studied. Despite extreme interest in the subjects, there is little concrete evidence regarding field work or about equity and access. Introductory courses as well as those for majors should be reconsidered in the light of contemporary research, and we believe that the evaluation of instruction should become common practice in all geoscience departments. Based upon these concerns, we offer the following very specific conclusions and recommendations:

## Conclusions

In the area of student concepts and alternative frameworks:

- **we know** that students bring a great deal of prior knowledge and many alternative conceptions to their college science classes, and that what they know and believe after instruction is heavily influenced by their prior knowledge and misconceptions.
- **we do not know** the details of student worldviews or the kinds of conceptual change teaching strategies that are likely to bring them from naïve to scientifically correct understanding of the geosciences.

With regard to the introductory course:

- **we know** that there is no significant learning in many introductory university lecture courses, and that even in the best cases student gains rarely exceed 25% of what could have been learned. In lecture courses where interactivity is high, gains may rise to as much as 50%. To exceed that amount, it is necessary to use instructional strategies that minimize lecture and maximize other teaching methods. We know that students learn best when they are engaged with real objects or phenomena, working in cooperative groups, solving complex problems, and interested in what they are learning.
- **we do not know** what kinds of alternative strategies, such as studio teaching, modeling instruction or problem-based learning, might be more effective alternatives to the traditional format for the introductory course.

About field work:

- **we know** that teaching in the field can be an exceptionally effective strategy for presenting content in both introductory and advanced courses. The rationale for other undergraduate field studies courses appears to be to increase student expertise in process skills, such as mapping or interpretation of relationships among variables.
- **we do not know** what kind of path students take on the journey from novice to expert, or how to organize field-mapping and similar courses in order to facilitate and maximize the results of such experiences.

From studies of affordances and constraints to learning that appear to characterize the unique nature of the geosciences:

- **we know** that experts are exceptionally capable in temporal, spatial and systems thinking, and that students who can access and apply these modes of thought are more successful in their work than others. We know that students can be trained to use these thought processes, and that there is some transfer to their academic work. However, there are other types of reasoning that may allow students and professionals to be equally successful in their work.

- **we do not know** whether temporal, spatial and systems thinking are essential to learning in the geosciences, or whether they are just an alternative modality, and we do not know how to incorporate instruction in these forms of thought into our traditional courses.

From studies of the affective dimensions and the geosciences:

- **we know** that women and minorities do not succeed in the sciences or enter into the professions in the same proportion as majority men. These populations come to the classroom with significantly lower levels of self-efficacy than others, and may have poorer attitudes toward the subject. Additionally, females have higher levels of anxiety, which inhibits deep learning and encourage surface learning. For all students, levels of self-efficacy appear to influence achievement.
- **we do not know** the full range of affective variables that are important to learning and persistence in the geosciences. We do not know how learning environments interact with affective variables, or how trait-treatment interactions influence achievement. We do not know how attitude toward the geosciences varies by student population, or how that influences their achievement and career choice.

## Recommendations

Based upon these conclusions, we recommend that:

- Studies of student alternative frameworks should be continued. These should lead to the construction and evaluation of conceptual change teaching models that can be validated as examples of best practice.
- Large data-bases of test items should be developed for use in geoscience education research. This would allow for the conduct of research and comparisons across a variety of settings and contexts.
- More attention should be given to interactive methods of instruction such as studio teaching, modeling instruction and problem-based learning. One basis for evaluation of such methods should be student achievement expressed as gain scores or other comparable metrics.
- Research studies should emphasize the differences between introductory courses for non-science students and advanced courses for students who will become geoscientists. Distinctions should be made between such goals as scientific literacy vs. professional preparation.
- Expert-novice studies should be increased, especially in the area of learning in the field.
- Advanced course work should emphasize instruction in the cognitive strategies that are important to the practice of the geosciences.
- More research should be conducted on issues of affect, underrepresentation of women and minorities, and pipeline issues.

## References


Apedoe, S.S. (2008). Engaging students in inquiry: tales from an undergraduate geology laboratory-based course. Science Education, 92(), 631-663

Assaraf, O.B.-Z. & N. Orion (2005). Development of system thinking skills in the context of Earth Systems Education. Journal of Research in Science Teaching, 42, 518-560.

Atwood, R. & Atwood, V. (1996). Preservice elementary teachers' conceptions of the causes of seasons. Journal of Research in Science Teaching, 33(5), 553-563.

Barba, R. & Rubba, P.A. (1992). A comparison of preservice and in-service earth and space science teachers' general mental abilities, content knowledge and problem-solving skills. Journal of Research in Science Teaching, 29(10), 1021-1036.

Bell, R.L. & Trundle, K.C. (2008). The use of a computer simulation to promote scientific conceptions of Moon phases. Journal of Research in Science Teaching, 45(3), 346-372.

Bickmore, B.R., Thompson, R.R., Grandy, D.A. & Tomlin,T. (2009). Geology as story-telling for teaching the nature of science and the science-religion interface. Journal of Geoscience Education, 57(3),178-190.

Black, A.A. (2005). Spatial ability and earth science conceptual understanding. Journal of Geoscience Education, 53(4), 402-414.

Boyle, A.S., Maguire, S., Martin, A., Milsom, C., Nash, R., Rawlson, S., et al. (2007). Fieldwork is good: the student perception of the affective domain. Geography in Higher Education, 31(2), 299-317.

Braun, L., Braxton, J., Piou, E., Wood, M.G., Shor, L. & Stein, F. (n.d.) A Private Universe Project: the teachers' lab. The Annenberg/CPB Math and Science Projects. Retrieved November 15, 2010 from (http://www.learner.org/teacherslab/pup)

Budd, D.A., Van der Hoeven Kraft, K.J., Knight, C., Wirth, K.R., McConnell,D.A., Bykerk-Kauffman, A., Matheney, R.K., Perkins, D. & Stempien, J.A. (3 November 2010). How much non-traditional teaching does it take to dramatically improve learning gains in introductory physical geology?[abs] p Denver, CO: Geological Society of America.

Catley, K.M. & Novick, L.R. (2009). Digging deep: exploring college students' knowledge of macroevolutionary time. Journal of Research in Science Teaching, 46(3), 311-332.

Clary, R.M. & Wandersee, J.H. (2007). A mixed methods analysis of the effects of an integrated geobiological study of petrified wood in introductory geology classrooms. Journal of Research in Science Teaching, 44(8), 1011-1035.

Dodick, J. & Orion,N. (2003a). Cognitive factors affecting student understanding of geologic time. Journal of Research in Science Teaching, 40(4), 415-442.

Dodick, J. & Orion,N. (2003b). Measuring student understanding of geological time. Science Education, 87(5), 701-731.

Elkins, J.T. & Elkins, N.M.L. (2007). Teaching geology in the field: significant geoscience concept gains in entirely field-based introductory geology courses. Journal of Geoscience Education, 55(2), 55(2), 126-132.

Engelbricht, A.C., Mintzes, J.J., Brown, L.M. & Kelso, P.R. (2005). Probing understanding in physical geology using concept maps and clinical interviews. Journal of Geoscience Education, 53(3), 263-270.

Frodeman, R. (1996). Envisioning the outcrop. Journal of Geoscience Education, 44, 417-427.

Gray, K., Steer, D., McConnel, D. & Owens, K.(2010). Using a student-manipulated model to enhance student learning in a large lecture class. Journal of College Science Teaching, 40(1), 86-95.

Hake, R. (1998). Interactive-engagement versus traditional methods: A six-thousand student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64-74.

Huntoon J.E., Bluth, G.J.S. & Kennedey, W.A. (2001). Measuring the effects of a research-based field experience on undergraduates and K-12 teachers. Journal of Geoscience Education, 49(3), 235-248.

Jarret, O.S. & Burnley, P.C. (2010). Lessons on the role of fun/playfulness from a geology undergraduate summer research program. Journal of Geoscience Education, 58(2), 110-120.

Kastens, K. A., Liben, L. S., and Agrawal, S., 2009, How students and field geologists reason in integrating spatial observations from outcrops to visualize a 3-D geological structure: International Journal of Science Education, v. 31, no. 3, p. 365-393.

Kortz, K.M. & Murray, D.M. (2009). Barriers to college students learning how rocks form. Journal of Geoscience Education, 57(4), 300-315.

Kortz, K.M., Smay, J.J. & Murray, D.P. (2008). Increasing learning in introductory geoscience courses using lecture tutorials. Journal of Geoscience Education, 56(3), 280-290.

Levine, R., R. Gonzales, S. Cole, M. Fuhrman & K. Carlson LeFloch (2007). The geoscience pipeline: a conceptual framework. Journal of Geoscience Education, 55(6), 458-468.

Libarkin, J. (2008). Concept Inventories in Science: Manuscript prepared for the National Research Council. Retrieved September 2010 from http://www7.nationalacademies.org/bose/Libarkin_CommissionedPaper.pdf

Libarkin, J.C. & Anderson, S.W. (2005). Assessment of learning in entry-level geoscience courses: results from the Geosciences Concept Inventory. Journal of Geoscience Education, 54(3), 394-401.

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

Libarkin, J.C.& J.P. Kurdziel (2006) Ontology and the teaching of Earth System Science. Journal of Geological Education, 54(3), 408-413.

Libarkin, J.C., Kudziel, J.P. & Anderson, S.W. (2007). College student conceptions of geologic time and the disconnect between ordering and scale. Journal of Geoscience Education, 55(5), 413-422.

Liben, L.S.,Kastens, K.. & Christensen, A.E. (in press). Spatial foundations of science education: the illustrative case of instruction on introductory geology concepts. Cognition and Instruction.

Linneman, S. & Plake, T. (2008). Searching for the difference: a controlled test of Just-in-Time Teaching for large-enrollment introductory geology courses. Journal of Geoscience Education, 54(1), 18-24.

Luo, W. (2008). Just-in-Time Teaching (JiTT)improves students' performance in classes adaptation of JiTT in four geography courses. Journal of Geoscience Education, 56(2), 166-171.

Maguire, S. (1998). Gender differences in attitudes to undergraduate fieldwork. Area. 30(3). 207-214.

McConnell, D.A., Steer, D.N., & Owens, K. (2003). Assessment and active learning strategies for introductory geology courses. Journal of Geoscience Education, 51(2), 205-216.

McConnell, D.A., Steer, D.N., Owens, K. & Knight, K.C. (2005). How students think: implications for learning in introductory geosciences courses. Journal of Geological Education, 53(4), 462-470.

McConnell, D.A., Steer, D.N., Owens, K.D., Knott, J.R., VanHorn, S., Dick, J., Foos, A., Malone, M., McGrew, H., Greer, L., & Heaney, P.J. (2006). Using concepttests to assess and improve student conceptual understanding in introductory geosciences courses. Journal of Geoscience Education, 54(1), 61-68.

McConnell, D., Jones, M.H., Budd, D.A., Bykerk-Kauffman, A., Gilbert, L.A., Knight, C., Kraft, K.J., Nyman, M., Stempien, J., Vislova, T., Wirth, K.R., Perkins, D., Matheney, R.K., Nell, R.M. (2009). Baseline Data on Motivation and Learning Strategies of Students in Physical Geology Courses at Multiple

Institutions: GARNET Part 1, Overview: Geological Society of America Abstracts with Program, v. 41, no. 7, p 603.

McConnell, D., Stempien, J.A., Perkins, D., van der Hoeven Kraft, K.J., Vislova, T., Wirth, K.R., Budd, D.A., Bykerk-Kauffman, A., Gilbert, L.A., and Matheney, R.K. (2010) The little engine that could – less prior knowledge but high self-efficacy is equivalent to greater prior knowledge and low self-efficacy. Geol. Soc. Amer. Abst. with Prog, 42(5).

Nelson, C.E. (2010). Dysfunctional Illusions of Rigor: Lessons from the Scholarship of Teaching and Learning. Chapter 10 in L. B. Nilson & J. E. Miller (Eds.), *To Improve the Academy: Resources for Faculty, Instructional, and Organizational Development*, Volume 28, John Wiley & sons, Inc

Ozdemir, G., Piburn, M. & Sharp, T. (March 2004). Exploring the use of visualization in a college mineralogy course. [abs] Vancouver, B.C.: National Association for Research in Science Teaching.

Perkins, D. (2004). Scholarship of teaching and learning, assessment, and the Journal of Geoscience Education. Journal of Geoscience Education, 52(2), 113-115.

Petcovic, H.L., Libarkin, J.C. & Baker, K.M. (2009). An empirical methodology for investigating geocognition in the field. Journal of Geoscience Education, 57(4), 316-328.

Petcovic, H.L. & Ruhf, R.J. (2008). Geoscience conceptual knowledge of preservice elementary teachers: results from Geoscience Concept Inventory. Journal of Geoscience Education, 56(3), 251-260.

Piburn, M., Sawada, D., Turley, J., Falconer, K., Benford, R., Bloom, I. & Judson, E. (2000). *Reformed Teaching Observation Protocol: Reference Manual* (Technical Report No. INOO-3). Tempe, AZ: Arizona Collaborative for Excellence in the Preparation of Teachers.

Piburn, M., Reynolds, S.J., McAulliffe, C., Leedy, D.E., Birk, J.P. & Johnson, J.K. (2005). The role of visualization in learning from computer-based images. International Journal of Science Education, 27(8), 513-527.

Raia, F. (2005). Students'understanding of complex dynamic systems. Journal of Geoscience Education, 53(3), 297-308.

Rebich, S. & Gautier, C. (2005). Concept mapping to reveal prior knowledge and conceptual change in a mock summit course on global climate change. Journal of Geoscience Education, 53(4), 355-365.

Riggs, E.M., R. Balliet & C. Lieder (2009a). Geologic problem solving in the field: analysis of field navigation and mapping by advanced undergraduates. Journal of Geoscience Education, 57(1), 48-63.

Riggs, E.M., Balliet, R. & Lieder, C. (2009b). Effectiveness of problem solving during geologic field examinations: insights from analysis of GPS tracks at variable time scales. In S.J. Whitmeyer, D.W. Mogk & E. Pyle (Eds.), *Field Geology Education: Historical Perspectivesand Modern Approaches.* G.S.A. Special Paper461, 223-240. Boulder, CO: Geological Society of America.

Sawada, D., Piburn, M., Judson, E., Turley, J., Falconer, K., Benford, R. & Bloom, I. (2002). Measuring reform practices in science and mathematics classrooms: The Reformed Teaching Observation Protocol. School Science and Mathematics, 102(6), 245-253.

Semken, S.C. & Butler-Freeman (2008). Sense of place in the practice and assessment of place-based science teaching. Science Education, 92(6), 1042, 1057.

Semken, S., and Williams, D., 2008, Situating Earth and space sciences in Superior: Toward place-based K-12 education in a culturally and geologically diverse Southwestern region: Geological Society of America Abstracts with Programs, v. 40, no. 1, p. 90

Siemankowski, F. & MacKnight, F. (October, 1971). Spatial cognition: success prognosticator in college science courses. Journal of College Science Teaching, 56-69.

Smith, M., Piburn, M. & Reynolds, S. (1999). Research for earth science learning. Geotimes, 44(8), 27, 35.

Springer, L., Stanne, M., & Donovan, S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering and technology: A meta-analysis. Review of Educational Research, 69(1), 21-51.

Stokes, A. & Boyle, A. (2009). The undergraduate geoscience fieldwork experience: influencing factors and implications for learning. In S.J. Whitmer, D.W. Mogk & E.J. Pyle (Eds.), Field Geology Education: Historical Perspectives and Modern Approaches. GSA Special Paper 461, 291-311. Boulder, CO: Geological Society of America.

Titus, S. & E. Horsman (2009). Characterizing and improving spatial visualization skills. Journal of Geoscience Education, 57(4). 242-254.

Trend, R.D. (2001). Deep time framework: a preliminary study of U.K. primary teachers' conception of geological time and perceptions of geosciences. Journal of Research in Science Teaching, 38(2), 191-221.

Tretinjak, C.A and Riggs, E.M. (2008). Enhancement of Geology Content Knowledge Through Field-Based Instruction for Pre-Service Elementary Teachers. Journal of Geoscience Education, 56(5), 422-433.

van der Hoeven Kraft, K. J., Stempien, J. A., Wilson, M.W., Wirth, K.R., Bykerk-Kauffman, A., Jones, M.H., Vislova, T, and Budd, D.A. (2010) Examination of the Motivation and Learning Strategies of Underrepresented Students in Introductory Geosciences. Geol Soc. Amer. Abst. with Prog, 42(5).

Vislova, T., Mcconnell, David, Stempien, Jennifer A., Benson, William M., Matheney, Ronald K., van der Hoeven Kraft, K. J., Budd, David A., Gilbert, Lisa A., Bykerk-Kauffman, Ann, and Wirth, Karl R. (2010) Role of Gender in Student Affect in Introductory Physical Geology Courses at Multiple Institutions. Geol. Soc. Amer. Abst. with Prog, 42(5).

Williams, D., and Semken, S., in press, Ethnographic methods in analysis of place-based geoscience curriculum and pedagogy, in Feig, A. P., and Stokes, A., eds., Qualitative Research in Geoscience Education: Geological Society of America Special Paper 474: Boulder, CO: Geological Society of America.