Geoscience Conceptual Knowledge of Preservice Elementary Teachers: Results from the Geoscience Concept Inventory

Heather L. Petcovic — Department of Geosciences and the Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI 49008, heather.petcovic@wmich.edu

Robert J. Ruhf — The Mallinson Institute for Science Education, Western Michigan University, Kalamazoo, MI 49008, robert.ruhf@wmich.edu

ABSTRACT

Effective instruction hinges in part on understanding what prior knowledge students bring to the classroom, and on evaluating how this knowledge changes during instruction. In many disciplines, multiple-choice tests have been developed to gauge student prior knowledge and assess learning. In this study, a 15-item version of the Geoscience Concept Inventory (GCI) was used to assess the prior knowledge and learning of students enrolled in an introductory physical and historical geology course specifically designed for preservice elementary (K-8) teachers. Gains (pretest to posttest) among participants (n = 122) averaged 4%, similar to gains reported elsewhere. However, gains among participants enrolled in revised course sections (n = 84) averaged 7-8%. Detailed analysis shows that statistically significant gains occurred on test items related to geologic time, earthquakes, radiometric dating, and tectonics. Items for which the greatest gains were observed correlate with teaching method; classroom activities coupled with discussion and supplemental reading appear most effective in increasing student knowledge. Our interpretation of the GCI results suggests that students need multiple opportunities to work with geologic concepts in a variety of formats, and provides further evidence of the persistence of student prior knowledge in specific topics.

INTRODUCTION

Most instructors recognize that students enter their classes with at least some prior knowledge, understanding, and beliefs. Effective learning requires students to integrate new information with this prior knowledge, and in many cases, to correct prior knowledge that is incomplete or incorrect (e.g., Piaget, 1962; Posner et al., 1982; Vosniadou et al., 2001). Effective learning also requires students to move from what may be rote, disconnected knowledge to a deeper, interconnected understanding of key concepts. An important goal of effective teaching, therefore, is to ensure that students progress from incomplete or incorrect ideas to a deeper and more correct conceptual understanding of the topic. Consequently, instructors should be aware of student prior knowledge to ensure effective teaching. Although many instructors recognize the role and importance of prior knowledge in learning, it is not always easy to ascertain what students know. Moreover, instructors need to know if student ideas have changed in response to teaching. Clearly, if the goal of instruction is to have students leave a course with greater and more correct conceptual understanding, tools are needed to assess this.

Libarkin and Kurdziel (2001) reviewed literature related to student prior knowledge and alternate conceptions and made several recommendations for how college instructors can assess student ideas. Open-ended, pre-instruction questionnaires and concept maps can be highly effective tools for assessing student ideas, yet these open-ended tools can be time-consuming for instructors to interpret and use. Multiple-choice tests in which the distracters (incorrect responses) are drawn from research on student alternate conceptions have the potential to provide instructors with more rapid yet still useful feedback as to what pre-instruction ideas their students hold.

Other science disciplines have taken this approach and have seen the development and use of such multiple-choice tools for the purposes of assessing student knowledge prior to instruction, measuring student learning gains in a particular course, and determining the effectiveness of different instructional approaches. In physics, the Force Concept Inventory (Hestenes et al., 1992) is widely used to assess student learning in introductory courses. Other disciplines have followed suit, with research-based concept inventories now available in chemistry (Mulford, 1996), astronomy (Zeilik et al., 1998), and biology (Anderson, 2002).

The need for a tool to assess student knowledge of concepts commonly taught in introductory, college-level geoscience courses has recently been addressed by Libarkin and Anderson (2005, 2006, 2008) through development of the Geoscience Concept Inventory (GCI). Test questions (items) on this multiple-choice instrument were developed from extensive analysis of actual student responses to open-ended questions covering a range of geoscience content, and validated using statistical tests (see Instrumentation, below; see Libarkin and Anderson, 2006; 2008 for additional details). Libarkin and Anderson (2005) demonstrated the utility of the GCI in assessing student learning and the effectiveness of instruction across a variety of college types, locations, and specific introductory courses. Elkins and Elkins (2007) also used the GCI to assess student learning and instruction in an entirely field-based course.

In this study, we utilize the GCI to examine conceptual knowledge of geoscience topics among preservice elementary teachers. Specifically, we use the GCI to (1) examine the level of prior geoscience conceptual knowledge of future elementary teachers, and (2) assess changes in conceptual knowledge among future elementary teachers during instruction in an introductory geoscience course specifically designed for this population. Whereas the goal of teaching is to increase learning among all students, it is especially important that future teachers gain a deep understanding of geoscience content so that they can effectively teach their future students.

Student Conceptions in the Geosciences - Student learning and conceptual knowledge in the sciences is an

Theme	Example Conceptions and Areas of Difficulty
Earth structure, tectonics	Confusion over states of matter in the earth's interior (e.g., a liquid core or mantle); tectonic plates as decoupled from the earth's surface; heat, pressure, climate, and people as causes of earthquakes (Libarkin et al., 2005); magma originating from earth's core (Dahl et al., 2005)
Earth materials	Difficulty associating rock types with causal processes; pressure and "amalgamation" as producing igneous rocks, glaciers as producing metamorphic rocks (Stofflet, 1993); any crystal that scratches glass is a diamond (Schoon, 1992)
Surface processes	Groundwater hosted primarily in underground lakes and rivers; difficulty conceptualizing the scale of groundwater occurrence (Dickerson et al., 2005); flooding only occurs after spring snow melt (Schoon, 1992)
Geologic time	Difficulty placing geologic and evolutionary events in the proper relative sequence; difficulty assigning absolute ages to key geologic events (Trend, 2000; 2001; Dahl et al., 2005; Libarkin et al., 2005; Libarkin et al., 2007); life and earth originated at the same time (Trend 2001; Libarkin et al., 2007); coexistence of humans and dinosaurs (Schoon, 1992; Trend, 2001; Libarkin et al., 2007)

Table 1. Areas of conceptual difficulty and examples of specific geoscience conceptions held by college students, preservice teachers, and/or inservice teachers.

active field of research. Although theorists disagree on how exactly learning takes place, authors of most of this literature agree on several main points: (1) the knowledge, skills, experiences, and beliefs that students bring into the classroom exert a powerful effect over what students learn and retain from their classroom experiences (e.g., Piaget, 1978; Vygotsky, 1978; Posner et al., 1982; Pintrich et al., 1993); (2) students' prior ideas (variously termed "misconceptions," "prior conceptions," or "alternate conceptions") are deeply entrenched and resistant to change (e.g., Nersessian, 1989; Redish, 1994; Wandarsee et al., 1994); and (3) appropriately structured learning experiences that take into account student prior knowledge can lead to meaningful learning (e.g., Vosniadou et al., 2001; Duit and Treagust, 2003). In addition to studies of how prior knowledge interacts with learning, some of the research in this field focuses on cataloguing student ideas in particular fields, including earth science (e.g., Phillips, 1991; Schoon, 1992). The extensive alternative conceptions literature is a useful starting point for instructors who want to understand what ideas their students may hold, and provides a valuable framework from which instructors can build meaningful instruction.

Studies of student conceptions in the geosciences have largely focused on pre-college populations (e.g., Ault, 1982; Happs, 1985; Bar, 1989; Marquis and Thompson, 1997; Trend, 1998; Gobert and Clement, 1999; Dodick and Orion, 2003). Recent work has begun to examine geoscience conceptions among college students, as well as preservice and inservice teachers (populations most similar to ours). However, as noted by Orion and Ault (2007), many of the same preconceptions appear across grade levels K-16 and among adults, and many K-12 teachers hold the same non-scientific conceptions as their students. This research has revealed an array of student prior knowledge and conceptions, and has identified specific area of conceptual difficulty that can be organized around themes common to introductory geoscience courses (Table 1).

Given the diversity of prior conceptions (and conceptions retained by students despite instruction) reported in the literature (Table 1), it would be exceedingly difficult for a college instructor to ascertain and address each individual idea held by each student in his or her class. Research-based multiple-choice instruments such as the GCI, therefore, can provide rapid but valuable feedback as to what conceptions students hold and how (or even if) these ideas change during instruction

METHODS

Classroom Context - All study participants were enrolled in an entry-level physical and historical geology course at a mid-sized, Midwestern university. The course, *Earth Science for Elementary Educators II* (ESEE2), is specifically designed for undergraduate students pursuing teacher certification programs in elementary education or in integrated elementary science (certification for teaching grades K-8, children aged 5-14). This course is one of six entry-level science courses in the subject areas of life, physical and earth science. It is also one of two earth science courses offered in the program; the other is an entry-level physical geography course (*Earth Science for Elementary Educators I; ESEE1*). Students pursuing elementary teaching certification must take four of the six science courses, with at least one course from each subject area. Students pursuing the elementary integrated science certification must take all six science courses.

The course format consists of two, 2-hour and 20-minute sessions per week. In-class student learning occurs predominantly through laboratory activities, small group discussions and written assignments, and brief lectures integrated with whole class discussion. Regular readings and homework assignments accessed via an on-line system supplement the in-class learning and provide opportunities for additional instructor feedback. Enrollment is limited to 24 students per course section; typically 2 course sections using identical curricula are offered per semester. Sections are taught by a geoscience/science education faculty member, geoscience education doctoral students, and a part-time instructor. Although this is a content course, instructors utilize methods such as inquiry teaching to model teaching practices appropriate to an elementary classroom. In addition to geoscience content, the course explicitly addresses topics of interest to future elementary teachers, such as children's ideas and misconceptions in earth science, state and national earth science benchmarks, and lesson plans and activities appropriate for K-8 students. For additional course information, including a detailed description of the course content, goals, context, and a syllabus, see (http://serc.carleton.edu/teacherprep/courses/WMU-ESEE2.html).

Instrumentation: The GCI - Development and evaluation of the GCI is detailed in Libarkin and Anderson (2006, 2008) and summarized here. The GCI was developed through an iterative process of qualitative data collection, item development and expert review, pilot testing, statistical analysis, item revision, and further testing. Although the topical content of the GCI was predetermined by the authors, individual items were derived from open-ended surveys and interviews of more than 1000 students from a wide variety of college settings. Utilizing a grounded theory methodology, the authors mined interview and survey data for fundamental concepts and authentic student responses, thus ensuring that items were based on existing student conceptions (as well as misconceptions), rather than rote knowledge (Libarkin and Anderson, 2008). Items were then administered as a pretest to more than 3500 students, reviewed by both content and education external experts, and reviewed by faculty participating in the pilot testing. Testing and feedback guided revision of individual items; revised items were then administered to participating students as a post-test. Follow-up think-aloud interviews, in which students were asked to explain their responses to individual items, were then used to gauge student interpretation of questions and the extent to which questions evaluated alternate conceptions. Together, this iterative development procedure established the construct and content (face) validity of the instrument (Libarkin and Anderson, 2008).

Additional validity and reliability measures, as well as the scaling function for the GCI, were determined by Rasch and differential item functioning (DIF) analysis (Libarkin and Anderson, 2006). From a theoretical perspective, the Rasch approach, a type of Item Response Theory (IRT) analysis, assumes that: (1) test item statistics will fit the Rasch model, (2) student performance on the test instrument will be improved once a concept is learned, (3) the characteristics of specific test items are independent of the test subject's ability, and (4) that both overall test and individual item performance are directly related to the trait being measured (conceptual understanding of the geosciences) (Libarkin and Anderson, 2006). The Rasch analysis also considers the difficulty of individual items on the test, and the scaling function developed from the statistical analysis correlates test score directly with a subject's conceptual understanding. Thus a subject's scaled test score and response to particular items directly correlates to his or her level of conceptual understanding of these topics (Libarkin and Anderson, 2006; 2008)

These development and validation procedures have resulted in the current version of the GCI, which includes 73 items arranged into 11 difficulty categories. Users of the GCI construct a subtest by choosing a question from each of the 11 categories plus 4 questions required of all subtests, resulting in a 15-item instrument (test questions and directions for subtest creation are available from the authors at http://www.msu.edu/~libarkin). The GCI is developed such that although an individual user can tailor items to the content of his or her courses, the scoring of the instrument is tied to a single scale (Libarkin and Anderson, 2006). This allows users to compare student learning and instructional approaches nationwide.

The GCI subtest used in this study was constructed from the 4 required anchor items (1, 2, 37, and 73) plus 1 item selected from each of the 11 difficulty categories. Items chosen were representative of content addressed in the course, with the majority of the items directly related to content encountered by students in the laboratory activities or class discussions. Thus the test items we chose represent what we believe are some of the key concepts covered by the course.

Research Procedure - The 15-item GCI was administered as a pretest in each participating course section during the first week of the semester. An identical version of the GCI was administered 15 weeks later during the last week of the course. In general, students took 15-20 minutes to complete the test. The course instructors did not view participant results until after each semester had ended.

The GCI was administered to a total of six course sections of ESEE2 over three semesters. One author (Petcovic) taught five of the course sections, and a part-time instructor taught the remaining section (section 2). Due to a realignment of course objectives between ESEE1 and ESEE2, significant content revisions took place between Spring 2005 (sections 1 and 2) and Fall 2005 (sections 3 and 4). These included the introduction of several new activities particularly in the area of plate tectonics, introduction of the online homework component, and introduction of a semester project where students investigate children's misconceptions of specific earth science topics. Because the content of the course changed, we replaced three GCI items to better reflect revised course content. The new items were of equivalent difficulty to the removed items, thus preserving the overall scoring of the GCI. Minor revisions to the homework and semester project components occurred between Fall 2005 (sections 3 and 4) and Spring 2006 (sections 5 and 6), but the GCI test was not changed during this time. Because of the changes to the course, data are reported individually for each course section, in aggregate for all course sections, in aggregate for sections 1-2, and in aggregate for sections 3-6.

RESULTS

Study Population - ESEE2 students present on the testing days and consenting to participate in the study took the GCI. Based on demographic data collected during the pretest, the aggregate study population (n = 122) is about 85% female, 98% white, traditionally-aged college juniors and seniors (mean age 21.8). Individual course sections range from 74-90% female. Approximately three-quarters of participants reported that they had taken an earth science course in high school, and 95% reported that they had taken another earth science course in college. Nearly all of these students have taken ESEE1 or the equivalent general education course in physical geography covering topics of atmosphere, weather, and climate. Only 9 students reported having taken a prior college-level course in physical or historical geology.

Overall Results of the GCI - The mean scaled score on the GCI pretest among all participants (n = 122) was 43.9%, with individual scores ranging from 20.3% to 67.0% correct on both the pretest and posttest (Figure 1a). However, even though the range of scores was similar between the pre- and posttests, the mean scaled score on the posttest (n = 102) was 48.1%, showing an overall gain (pre to post) among all participants of 4.2% (Figure 1).

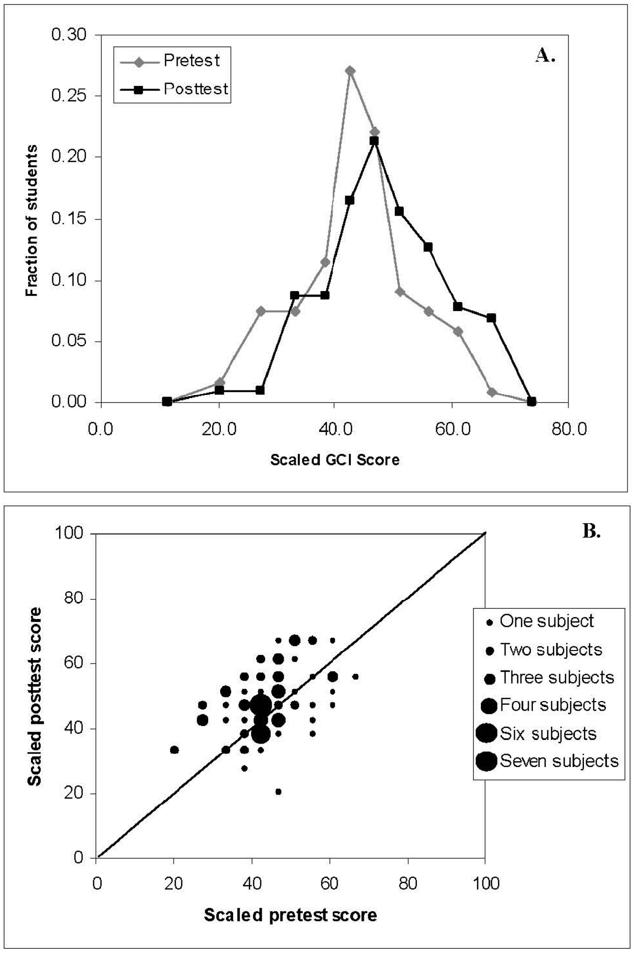


Figure 1. A. Distribution of scaled GCI scores on the pretest and posttest, results shown for all class sections (pretest n = 122; posttest n = 102). B. Matched pretest and posttest pairs for individual students from all class sections (n = 89). Points falling on the line indicate no change between the pretest and posttest scores. Points falling below the line indicate a decrease between pretest and posttest scores, and points falling above the line indicate an increase between pretest and posttest scores. The size of the symbol indicates the number of students achieving a particular pretest-posttest score combination. Note that the majority of tests fall above the line.

Students in sections 3-6 had a slightly greater overall mean gain of 6.5%. Improvements were seen among both high- and low-scoring students, but the greatest normalized gains were seen among students who scored 30-40% correct on the pretest (Figure 1).

We were able to match pretests and posttest for 89 individual students (Figure 1b); thus subject attrition between the pretest and posttest was 27%. Attrition was due to either students being absent on a testing day, or students choosing not to consent to participate in the study.

Student responses to individual GCI items ranged from 0-83% correct on the pretest, and 0-92% correct on the posttest (Figure 2). Items covered a wide range of difficulty, as determined by fraction of correct responses on the pretest; 50% or more of pretest responses were correct for about half of the test items, and fewer than 50% of the pretest responses were correct for the remaining test items (Figure 2). Pre- to posttest gains occur among both easy and difficult test items (Figure 2).

ANALYSIS

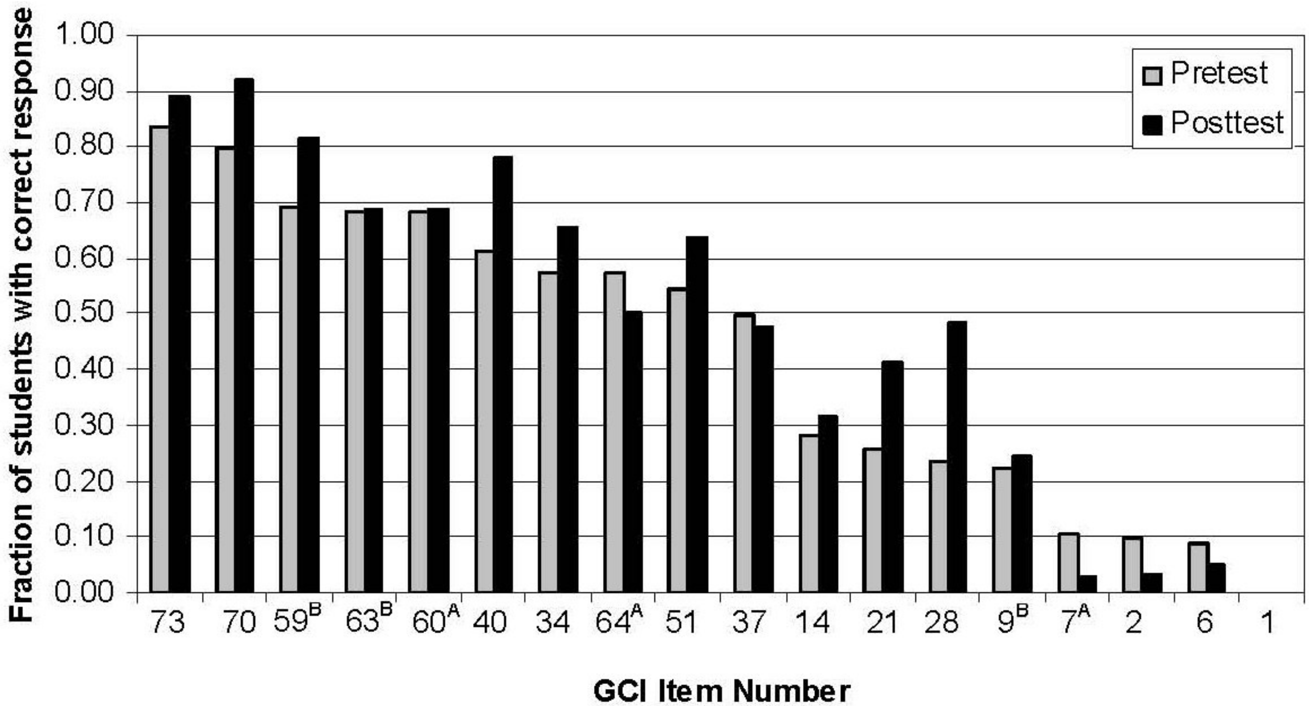
We analyzed the GCI data collected during this study quantitatively to determine (1) whether statistically significant differences on pretest performance existed between ESEE2 course sections, (2) whether observed pretest to posttest changes in mean GCI scores were statistically significant for individual course sections, and (3) for course sections with significant pretest to posttest change, which individual GCI items were causing this change. All analyses utilized matched pretests and posttests for individual students (n = 89).

One-way Analysis of Variance (ANOVA) tests on pretest scores were used to determine whether significant differences in pretest performance, and thus the initial level of geoscience conceptual knowledge, existed between course sections. A one-way repeated measures ANOVA on pretest to posttest difference scores using course section as a covariate was performed to determine whether course section had a significant effect on student performance. Significant results prompted further testing, so paired sample t-tests on the difference scores for each course section were performed to determine whether pre- to posttest performance changes were significant within any of the course sections. Thus, these tests track overall pretest to posttest learning as measured by the GCI. For these analyses, a p-value of 0.05 or less (95% or greater level of confidence) indicated significance.

For the course sections in which the previous tests revealed a statistically significant overall change in mean pretest to posttest difference score, we analyzed the fraction of students responding correctly to each GCI items using the McNemar chi-square test for matched pairs (Glass and Hopkins, 1999). The chi-square test enabled us to identify individual GCI items for which we observe a statistically significant change between the pretest and posttest, hence indicating a significant change in students' level of knowledge of individual GCI topics.

Analysis of Pretest GCI Scores - Table 2 reports the mean pretest, posttest, and matched pre-to posttest difference scores for all six course sections. The p-value obtained from the one-way ANOVA performed on all raw pretest scores was 0.54, indicating that there were no statistically significant differences on the pretest among the course sections. An effect size of $\eta^2 = 0.045$ was observed, thus the section factor explains 4.5% of the variance in the pre-test scores. The initial level of geoscience concept knowledge as measured by the GCI was therefore similar across all six course sections.

Analysis of Pretest to Posttest Mean GCI Difference Scores - The repeated measures ANOVA results revealed a lack of significant difference between pretest and posttest scores for matched pairs of tests, as indicated by an insignificant p-value for test effect of 0.364, with an effect size of $\eta^2 = 0.009$. However, the section covariate had a significant interaction with the difference scores, as verified by the p-value of 0.006 with an effect size of $\eta^2 = 0.084$. The average posttest score was higher than the average pretest score for five of the course sections (Table 2). Further analysis of paired



[A]Item used in Sections 1-2 test version only
[B]Item used in Sections 3-6 test version only

Figure 2. Fraction of students responding correctly to each GCI item, reported in aggregate for all class sections (pretest n = 122; posttest n = 102). The number on the horizontal axis refers to the original GCI item number. Items are arranged by difficulty, as determined by the fraction of correct pretest responses. Our difficulty ranking largely agrees with the ranking reported by Libarkin and Anderson (2006). Statistically significant improvements between the pretest and posttest were noted for item 14 in section 5, item 21 in section 4, item 28 in sections 3 and 4, item 70 in section 3, and item 73 in section 3. Statistically significant decreases were noted for item 2 in section 5 and item 64 in section 1.

Section No.	N.	Pretest	Posttest	Difference	p-value	t-statistic±
1	15	0.413	0.360	-0.054	0.14	1.550
2	10	0.447	0.487	0.040	0.37	-0.947
3	16	0.375	0.475	0.100	0.00**	-3.586
4	21	0.451	0.511	0.060	0.05*	-2.130
5	14	0.448	0.557	0.110	0.01**	-3.371
6	13	0.400	0.477	0.077	0.11	-1.702
All	89	0.422	0.479	0.056	0.00**	-3.803
3-6	64	0.421	0.505	0.084	0.00**	-5.204

Table 2. Paired sample t-tests for the pretest/posttest raw scores. *Significant at the 95% level of confidence **Significant at the 99% level of confidence ±In the case of paired-sample t-tests, the paired t-test statistic is reported as a measure of effect size.

sample t-tests (Table 2) revealed that the pre- to posttest difference in mean scores was significant for three of the course sections (3, 4, and 5). The only course section for which the average posttest score was lower than the average pretest score was section 1 (Table 2); however, the paired sample t-test indicated that this drop was not statistically significant.

Analysis of Correct Response to Each GCI Item - McNemar chi-square tests for matched pairs were performed for all 15 test items for the three course sections in which we found statistically significant pretest to posttest improvement (sections 3, 4, and 5; Table 3). Based upon the chi-square test results, statistically significant increases in the number of correct responses were observed for specific items in course sections 3 (GCI items 28, 70 and 73), 4 (items 21 and 28), and 5 (item 14) (Table 3). Thus, it appears that these items were responsible for the significant increase in overall posttest performance within these sections. Interestingly, we observed a statistically significant decrease in the number of correct responses to item 2 in Section 5, although this section did have an overall mean pretest to posttest increase.

GCI Item Number	Section 3			Section 4			Section 5		
	Pre-Post Diff	Chi-sq	Effect Size	Pre-Post Diff	Chi-sq	Effect Size	Pre-Post Diff	Chi-sq	Effect Size
2	0	0	0	-14.3	3.00	0.143	-28.6	4.00*	0.286
14	-12.5	0.40	0.125	0	0	0	35.7	5.00*	-0.357
21	12.5	0.67	-0.125	28.6	4.50*	-0.286	21.4	1.80	-0.214
28	37.5	6.00*	-0.375	38.1	6.40*	-0.381	35.7	3.57	-0.357
70	25.0	4.00*	-0.250	9.5	1.00	-0.095	21.4	3.00	-0.214
73	25.0	4.00*	-0.250	4.8	1.00	-0.048	14.3	2.00	-0.143

Table 3. Matched pretest to posttest difference scores and McNemar chi-square test results for GCI items in course sections 3-5 (observed values). *Significant change in the number of correct responses at the 95% level (critical value = 3.84, df =1).

DISCUSSION

Our study has utilized the GCI to (1) examine the level of prior geoscience conceptual knowledge of future elementary teachers, and (2) assess pretest to posttest changes in conceptual knowledge among future elementary teachers enrolled in ESEE2. Thus our analysis of GCI results serves as a tool to assess both the prior knowledge of our students, and the efficacy of our course designed for this population of students. We acknowledge that using a multiple-choice instrument to examine student ideas is limited; however well-designed, a multiple-choice test is unlikely to reveal the full extent to which students truly understand concepts, and the extent to which these concepts are interconnected to create conceptual frameworks. However, because of the way in which it was developed and validated, the GCI is particularly useful in assessing whether students have correct conceptual (as opposed to rote) knowledge of particular geoscience topics.

Prior Geoscience Knowledge of Preservice Elementary Teachers - Overall, our students performed poorly on the GCI pretest, with a mean scaled score of 43.9%. Other users of the GCI report similar scaled pretest scores of 39-43% (Libarkin and Anderson, 2005; Dahl et al., 2005; Elkins and Elkins, 2007), suggesting that our population of preservice elementary teachers have a pre-instructional level of geoscience knowledge that is similar to the national sample of undergraduate students, as measured by the GCI. Additionally, we did not find any statistically significant differences on mean pretest scores between course sections (Table 2), suggesting that our student population as a whole has similar levels of pre-college geoscience preparation.

A widespread perception exists that future elementary teachers have weak academic backgrounds in science. Research comparing the academic science preparation of preservice elementary and secondary teachers (Book and Freeman, 1985) and comparing earth science content knowledge of preservice and inservice teachers (Dahl et al., 2005) does confirm that preservice elementary teachers typically are weaker in science than other populations of students. However, the results of our study suggest that our population of preservice elementary teachers has pre-instructional knowledge of geoscience equal to the national sample of undergraduates reported by Libarkin and Anderson (2005).

Until 2006, earth science was a high school graduation requirement in Michigan, where most of our students attended high school. We suggest that the similarity in pretest performance between our students and the national sample of undergraduates reflects that most of our students (about 75%) have taken an earth science course in high school. Our student population of future elementary teachers, clearly, is no weaker in earth science than the national sample of undergraduates, based on GCI pretest score.

Overall Changes in Geoscience Knowledge among ESEE2 Students - The overall gains made by students enrolled in our course (Table 2; Figure 1) are similar to the average scaled gain of 4% reported by Libarkin and Anderson (2005), which utilized a 20-question version of the GCI to examine over 2500 students enrolled in 29 courses nationwide. Among the significant findings from their study, these authors report that students with minimal geoscience knowledge as measured by the pretest (means of <40% correct) made the greatest conceptual gains in their geoscience courses. Students who scored higher on the pretest (40-60%) exhibited minimal changes in score, whereas the highest scoring students exhibited no change. We observed a similar result in which students scoring less than ~40% on the pretest had the greatest gains (Figure 1B), and students scoring 60% or higher demonstrated little gains.

Comparison between GCI results for our students and the national sample suggests that, overall, ESEE2 is no more effective than other introductory geoscience courses in increasing student knowledge of the geosciences. However, we note that students in sections 3-6 have overall learning gains on the GCI that are greater than gains for students in sections 1-2. Three out of the four sections, in fact, have statistically significant pretest to posttest gains of 7-8% (Table 2). Additionally, for nearly all items common to both versions of the test, student learning as measured by the GCI was greater for students enrolled in these sections. We interpret these finding to indicate that the revisions we made to ESEE2 were effective in increasing overall student learning in this course. However, as our students are still achieving only 50% correct responses on the posttest, we clearly have a long way to go toward improved student learning.

Significant Changes on Individual GCI Items - A detailed look at the fraction of students responding correctly to each GCI item suggests that student conceptions changed for some of the course content but not for other content. Furthermore, statistically significant pretest to posttest learning gains were distributed among both easy and more difficult items (Table 3; Figure 2); in other words, our students were not

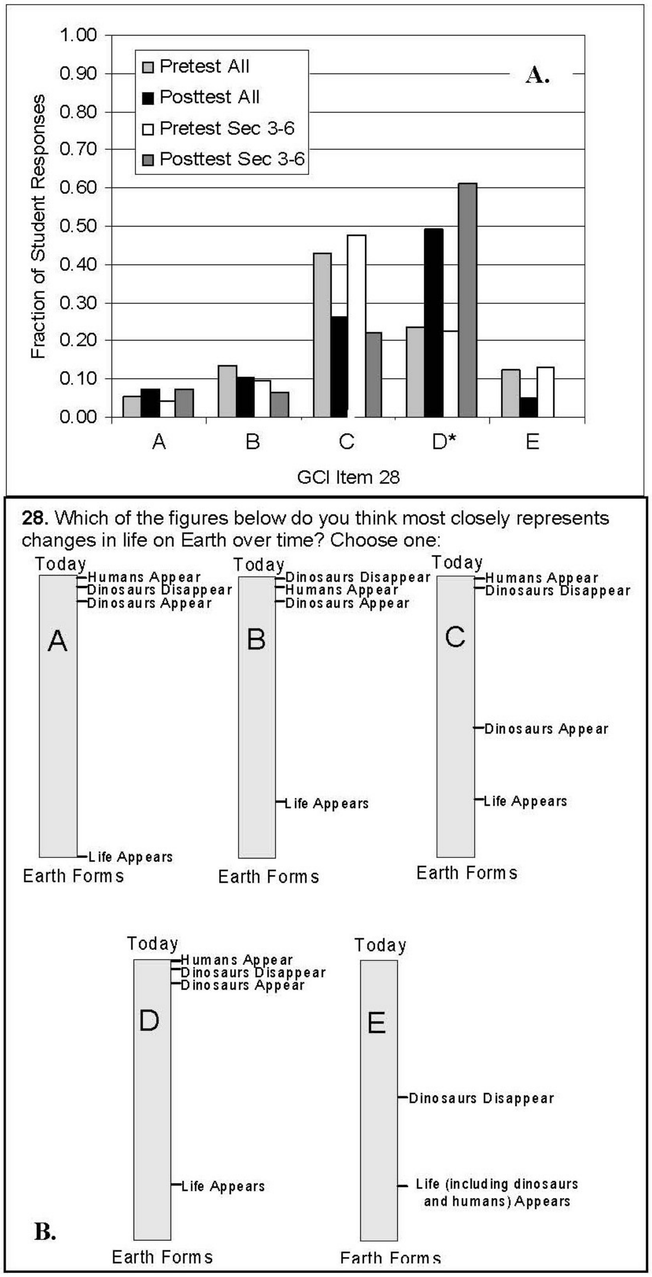


Figure 3. A. Fraction of students responding to each distracter for GCI item 28, which had the overall largest gains for any item used in this study. Correct response is indicated with *. Approximately 23.8% of students chose the correct response to this item on the pretest, but 48.6% of all students (and 60.1% of students in sections 3-6) chose the correct response on the posttest. B. GCI item 28.

merely performing better on the easy GCI items on the posttest. What do we think we are doing right in this course? An analysis of student responses to each GCI item can reveal changes in student thinking, presumably as a result of instruction in this course. This type of item-by-item analysis of GCI data is also reported by Dahl et al. (2005).

The difficulties that students, as well as preservice and inservice teachers, have grasping the scale of geologic time and understanding the sequence of events in geologic history is well documented (e.g., Trend, 2000; 2001; Dodick and Orion, 2003; Libarkin et al, 2005; Dahl et al., 2005; Libarkin et al., 2007). Our students, however, had some of the greatest (and statistically significant) learning gains on items related to geologic time - specifically item 28 (Figure 3). Prior to instruction about 45% of students could place a series of events (life appears, dinosaurs appear, dinosaurs disappear, humans appear) in the correct sequence; however, only 25% of students could place these events in the correct sequence and at the correct locations on a geologic timescale. On the posttest, however, over 50% of students (and over 60% of students in sections 3-6) could answer this item correctly (Figure 3).

A detailed description of the classroom activity we use to teach geologic time is available at (http://serc.carleton.edu/teacherprep/resources/activities/earthhistory.html), presented as part of the 2007 Teacher Preparation workshop. In brief, the activity follows a modified learning cycle format in which students are presented with a "hook" to engage their interest in the subject, share prior knowledge about the subject, manipulate models or data related to the content, work collaboratively to reach conclusions based on evidence collected, and further reflect on the content learned in a homework assignment. This framework is used in all activities in this course. In this particular classroom activity, students first use an interview with an older adult to construct a timescale, which they divide into periods and compare timescales among the class (modified after Hermann and Lewis, 2004). Next, students develop a geologic timescale of events provided by the instructor using only their prior knowledge. Students then use classroom and Internet resources to place the same events in the proper order and at the correct locations along the timescale. Finally, students investigate the geologic timescale and place eons and eras of geologic time on the same scale as earth events. Comparisons are drawn between the human life timescale and geologic time.

Changes in student responses to item 28 suggests that this particular class activity seems to be effective in helping students to understand the nature of geologic time. Other classroom activities also result in high learning gains, as measured by the GCI. Our evidence suggests a loose correlation between type of learning task and conceptual knowledge change among our students (Table 4).

During our course we spend at least one entire class period engaged in inquiry-based activities and group discussion of concepts directly related to geologic time (item 28), radioactive decay and absolute age dating techniques (items 21 and 1), plate tectonics and the structure of the earth (items 51 and 14), fossils (item 40), and groundwater (item 59). Concepts directly related to these GCI items are introduced to students through in-class activities and discussions, and reinforced through homework reading and quizzes. The greatest learning gains on the GCI were among these items (Table 4); items 28, 40, 21, 59, and 51 had pretest to posttest learning gains of over 5% among all students tested (Figure 2), including some course sections with statistically significant gains on these items (Table 3).

Topics related to earthquakes (item 70), the cause of ocean waves (item 34), the breakup of Pangea (item 73), and the location of glaciers (item 9) are discussed in the class but are not the main focus of the classroom activity. Some of the concepts related to these GCI items are

GCI Item No.	% Pre- to Posttest Difference	Content Area	Type of Learning Task
28	+38.0*	Geological timescale	Classroom activity, homework reading
40	+19.9	Fossil definition	Classroom activity, homework reading
70	+18.8*	Earthquake definition	Class discussion, homework reading
21	+17.9*	Radioactive decay	Classroom activity, homework reading
34	+13.1	Cause of ocean waves	Class discussion
51	+12.4	Location of volcanoes, earthquakes at plate boundaries	Classroom activity, homework reading
59	+12.0	Groundwater definition	Classroom activity, homework reading
73	+9.5*	Cause of Pangea breakup	Class discussion
14	+4.6*	Interior of the Earth	Classroom activity, homework reading
9	+2.1	Location of glaciers	Class discussion, homework reading
63	+0.4	Past size of Earth	Homework reading
1	0.0	Techniques to date age of Earth	Classroom activity, homework reading
37	-0.1	Duration of Pangea breakup	Class discussion
6	-3.0	Tectonic plate definition	Homework reading
2	-8.9*	Factors effecting erosion rates	Homework reading

Table 4. Correlation between type of learning task and pretest to posttest change on the GCI, aggregated for course sections 3-6. * Statistically significant change in the number of correct responses at the 95% or 99% level.

covered in out-of-class reading assignments. Gains on these GCI items were lower than for the prior set of items, generally less than 5% (with the exception of item 70) (Table 4; Figure 2). Finally, concepts related to the past size of the earth (item 63), erosion rates (item 2) and the definition of a tectonic plate (item 6) are addressed in out-of-class homework reading assignments but class time is not spent discussing these concepts. All of these items show essentially no change or decreases between the pretest and posttest (Table 4; Figure 2).

Our analysis suggests that our class format in which students learn content through hands-on activities supplemented by class discussion and homework reading leads to the greatest conceptual gains, as measured by the GCI. The impact of class discussion on conceptual learning is inconclusive; of the 3 GCI items that tested content learned only via class discussion in our course, two resulted in gains and one in essentially no change as measured by the GCI. Out-of-class reading appears to be the least effective means of student learning. Our data suggest that students need multiple opportunities in different formats, such as classroom activities, discussions, and supplemental reading materials, to develop a better understanding of geoscience concepts.

Persistence of Prior Knowledge and Entrenchment of Student Ideas - Libarkin and Anderson (2005) report extreme persistence of preconceived ideas despite instruction, as measured by the GCI. Their work, along with additional work (Trend, 2000; 2001; Dodick and Orion, 2003; Dahl et al., 2005), points to deeply entrenched ideas related to geologic time, events in geologic history, and absolute age dating. Our analysis of several GCI items, as exemplified by response patterns to item 1, also illustrates the persistence of prior knowledge.

Item 1, the most difficult item on the GCI, asks students to choose which method(s) scientists use to determine the age of the earth (Figure 4). This item is difficult conceptually, but also because students are prompted to select all responses that are correct (yet only "C" is correct). In our study, no students chose only the correct response (C) on either the pretest or on the posttest (Figure 2). On the pretest, only 20% of students selected analysis of uranium-lead as a method for determining the age of the earth, in addition to choosing additional distracters (Figure 4). On the posttest, however, over 45% of all students (and over 55% of students in sections 3-6) selected the correct response (Figure 4), suggesting that a substantial fraction of students did learn that the uranium-lead dating method can be used to determine the age of the earth. However, all of these students also selected other distracters in addition to the correct response on the posttest, suggesting that they held on to their prior conceptions that fossils, sedimentary layers, and carbon-14 dating were also methods used to determine the absolute age of the earth. It appears that we were successful in introducing the correct concept to our students, yet the students did not abandon their prior conceptions. This example points to the extreme entrenchment of prior ideas and to the challenge of altering these ideas during instruction.

CONCLUSIONS

The GCI has been used in our course as a means to gauge students' conceptual understanding prior to instruction, and to help us understand where our instruction has, and has not, been effective. We find that the population of preservice elementary teachers in our study has the same pre-instructional concept knowledge of the geosciences as other undergraduates, as measured by the GCI. This is in contrast to many studies that report weak background in science among preservice and inservice elementary teachers. Large pretest to posttest gains on some GCI items suggest that our course is particularly

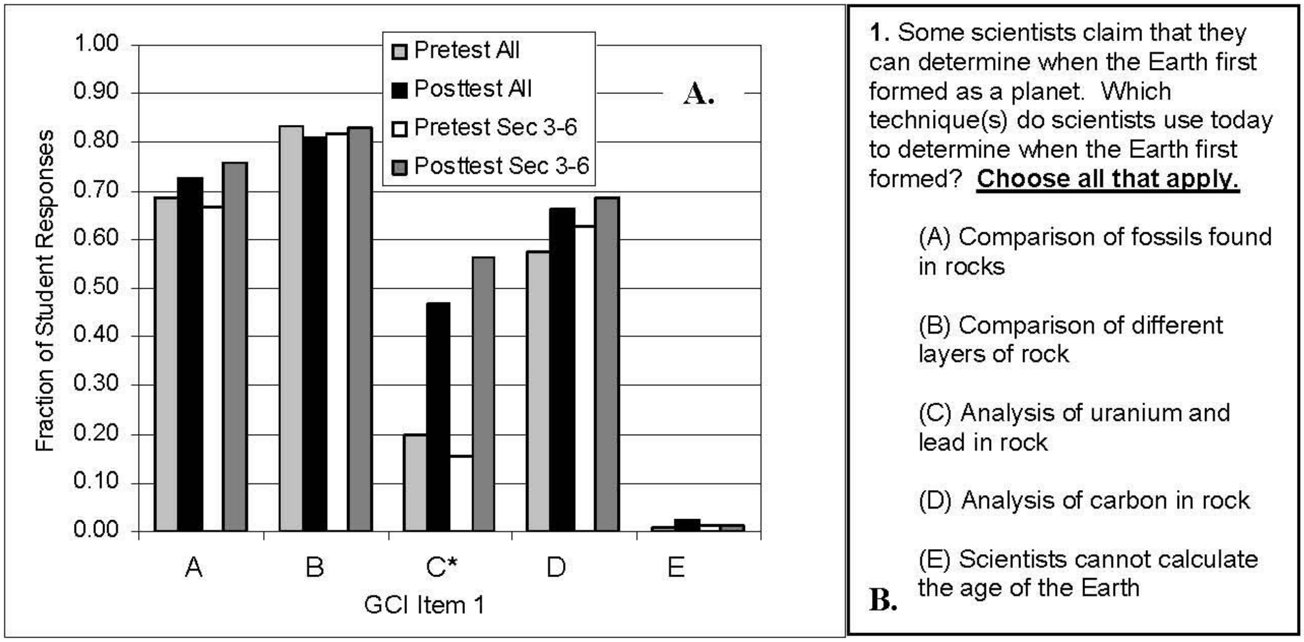


Figure 4. A. Fraction of students responding to each distracter for GCI item 1. Correct response is indicated with *. No students chose only the correct response on either the pretest or posttest. However, 19.9% of all students chose the correct response in addition to other distracters on the pretest. On the posttest, 46.5% of all students (and 58.8% of students in sections 3-6) chose the correct response in addition to choosing one or more incorrect distracters. B. GCI item 1.

effective in altering certain prior conceptions, especially those related to geologic time. However, analysis of other items points to the extreme entrenchment of student ideas, despite instruction. Furthermore, overall student performance on the GCI posttest remains poor, indicating that we have further work to do toward improving instruction and student learning in this course. Our analysis suggests that engendering conceptual development among students is difficult and that instructors may need to target certain concepts and spend more time on these. Class activities and group discussion, in particular, may also help students to develop scientifically correct conceptions. Out-of-class homework assignments appear to be least effective in bringing about conceptual understanding.

ACKNOWLEDGEMENTS

We would like to thank the many ESEE2 students who participated in the study. Julie Libarkin graciously supplied the current version of the GCI instrument as well as detailed instructions for constructing the subtest we used, plus gave us feedback on this manuscript. Herb Fynewever, Charles Henderson, and David Rudge provided valuable feedback on a draft version of this manuscript. We thank Roger Trend and an anonymous reviewer for their insightful comments that helped improve our work. We thank the statistics consultants at WMU for their assistance with running and interpreting the statistical tests. This study was conducted under Western Michigan University Human Subjects Institutional Review Board (HSIRB) protocol 04-12-12.

REFERENCES

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

Anderson, D.L., Fisher, K.M., and Norman, G.J., 2002, Development and validation of the conceptual inventory of natural selection, Journal of Research in Science Teaching, v. 39, p. 952-978.

Bar, V., 1989, Children's views about the water cycle, Science Education, v. 73, p. 481-500.

Book, C., and Freeman, D., 1985, Comparing backgrounds and educational beliefs of elementary and secondary teacher education candidates, Research and Evaluation in Teacher Education: Program Evaluation Series No. 8, East Lansing, MI, Michigan State University, 22 p.

Dahl, J., Anderson, S.W., and Libarkin, J., 2005, Digging into Earth science: alternative conceptions held by K-12 teachers, Journal of Science Education, v. 12, p. 65-68.

Dickerson, D., Callahan, T.J., Van Sickle, M., Hay, G., 2005, Students' conceptions of scale regarding groundwater, Journal of Geoscience Education, v. 53, p. 374-380.

Dodick, J., and Orion, N., 2003, Cognitive factors affecting student understanding of geologic time, Journal of Research in Science Teaching, v. 40, p. 415-442.

Duit, R., and Treagust, D., 2003, Conceptual change: a powerful framework for improving science teaching and learning, International Journal of Science Education, v. 25, p. 671-688.

Elkins, J.T., and 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, v. 55, p. 126-132.

Glass, G., and Hopkins, K., 1996, Statistical Methods in Education and Psychology (3rd Edition), Boston, Allyn and Bacon.

Gobert, J.D., and Clement, J.J., 1999, Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in

plate tectonics, Journal of Research in Science Teaching, v. 36, p. 39-53.
Happs, J.C., 1985, Some aspects of student understanding of glaciers and mountains, In: Working paper of the Science Education Research Unit, University of Waikato, New Zealand.
Hermann and Lewis, 2004, A formative assessment of geologic time for high school Earth Science students, Journal of Geoscience Education, v. 52, n. 3, p. 231-235.
Hestenes, D., Wells, M., and Swackhamer, G., 1992, Force concept inventory, The Physics Teacher, v. 30, p. 141-158.
Howell, D., 2002, Statistical Methods for Psychology (5th Edition), Pacific Grove, CA, Duxbury.
Libarkin, J.C., and Anderson, S.W., 2008, Development of the Geoscience Concept Inventory, In: Deeds, D., and Callen, B., editors, Proceedings of the National STEM Assessment Conference (Assessment of Student Achievement), October 19-21, 2006, Washington DC, Drury University, p. 148-158.
Libarkin, J.C., and Anderson, S.W., 2006, The Geoscience Concept Inventory: application of Rasch analysis to concept inventory development in higher education, In: Liu, X., and Boone, W., editors, Applications of Rasch Measurement in Science Education, JAM Publishers, p. 45-73.
Libarkin, J.C., and Anderson, S.W., 2005, Assessment of learning in entry-level geoscience courses: results from the Geoscience Concept Inventory, Journal of Geoscience Education, v. 53, p. 394-401.
Libarkin, J.C., Anderson, S.W., Dahl, J., Beilfuss, M., Boone, W., and Kurdziel, J.P., 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., and Kurdziel, J.P., 2001, Research methodologies in science education: assessing students' alternative conceptions, Journal of Geoscience Education, v. 49, p. 378-383.
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, p. 413-422.
Marques, L.F, and Thompson, D.B., 1997, Portuguese students' understanding at age 10/11 and 14/15 of the origin and nature of the Earth and the development of life, Research in Science and Technological Education, v. 15, p 29-51
Mulford, D.R., 1996, An Inventory for Measuring College Students' Level Of Misconceptions in First Semester Chemistry [MS thesis], West Lafayette (Indiana), Purdue University. Available at (http://jchemed.chem.wisc.edu/JCEDLib/QBank/collection/CQandChP/CQs/ConceptsInventory/CCIIntro.html)
Nersessian, N.J., 1989, Conceptual change in science and in science education, Synthese, v. 80, p. 163-183
Orion, N., and Ault, Jr., C. R., 2007, Learning earth sciences, In: Abell, S.K., and Lederman, N.G., editors, Handbook of Research on Science Education, Lawrence Erlbaum Associates Inc., Mahwah, New Jersey, p. 653-687.
Piaget, J., 1978, Success and Understanding, Cambridge, MA, Harvard University Press.
Pintrich, P.R., Marx, R.W., and Boyle, R.A., 1993, Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change, Review of Educational Research, v. 63, p. 167-199.
Posner, G., Strike, K.A., Hewson, P., and Gertzog, W., 1982, Accommodation of a scientific conception: toward a theory of conceptual change, Science Education, v. 66, p. 211-227.
Redish, E.F., 1994, Implications of cognitive studies for teaching physics, American Journal of Physics, v. 62, p. 796-803.
Stofflet, R.T., 1993, Preservice elementary teachers' knowledge of rocks and their formation, Journal of Geological Education, v. 41, p. 226-230.
Schoon, K.J., 1995, The origin and extent of alternative conceptions in the earth and space sciences: a survey of pre-service elementary teachers, Journal of Elementary Science Education, v. 7, p. 27-46.
Trend, R.D., 1998, An investigation into understanding of geological time among 10- and 11-year old children, International Journal of Science Education, v. 20, p. 973-988.
Trend, R.D., 2000, Conceptions of geological time among primary teacher trainees, with reference to their engagement with geoscience, history, and science, International Journal of Science Education, v. 22, p. 539-555.
Trend, R.D., 2001, Deep time framework: a preliminary study of UK primary teachers' conceptions of geological time and perceptions of geoscience, Journal of Research in Science Teaching, v. 38, p. 191-221.
Vosniadou, S., Ioannides, C., Dimitrakpoulou, A., and Papademetriou, E., 2001, Designing learning environments to promote conceptual change in science, Learning and Instruction, v. 11, p. 381-419.
Vygotsky, L.S., 1978, Mind in Society: The Development of the Higher Psychological Processes, Cambridge, MA, Harvard University Press, 159 p.
Wandersee, J.H., Mintzes, J.J., and Novak, J.D., 1994, Research on alternative conceptions in science, In: Gabel, D., editor, Handbook of Research on Science Teaching and Learning, New York, Macmillan, p. 177-210.
Zeilik, M., Schau, C., and Mattern, N., 1998, Misconceptions and their change in university-level courses, The Physics Teacher, v. 36, p. 104.