ABSTRACT

CZAJKA, CHARLES DOUGLAS. Assessing Learning and Teaching across Geoscience Courses and Curricula. (Under the direction of Dr. David McConnell).

Discipline based education research arose out of the realization that students face unique learning challenges in the different Science, Technology, Engineering and Mathematics (STEM) disciplines and that knowledge about teaching and learning coupled with content knowledge can be used to improve student learning across STEM courses. Common education research goals across the STEM disciplines include understanding how people think and learn within a discipline, assessing instructional strategies that enhance student learning, identifying effective strategies for guiding findings into classroom practice, and identifying approaches to make STEM education more inclusive and diverse. The work here describes three studies that contribute to these goals within the geosciences.

Efforts to promote the use of student-centered learning environments in STEM education have been tempered by a lack of professional development strategies that help faculty overcome common barriers to reform. Chapter 1 investigated the impact that adopting a suite of student-centered teaching materials had on the teaching practices and beliefs of eight geoscience faculty. Each instructor adopted 18 lessons developed as part of the InTeGrate (Interdisciplinary Teaching about the Earth for a Sustainable Future) project into their courses. A self-report survey and observational protocol were used to collect data on teaching practices whereas instructor beliefs were captured using an interview protocol. Data was collected over three semesters, a control semester using traditional materials, a pilot semester using the new materials, and a final treatment semester. While no self-reported changes in practices were reported, observations indicate the incorporation of more student-centered teaching practices. Interview results confirm that most instructors made at least moderate shifts toward more student-centered beliefs about
teaching and learning. These findings demonstrate that the adoption of well-crafted, student-centered instructional materials can have a positive impact on both the teaching practices and beliefs of college faculty.

The instructors above also collected student performance data on a sixteen question Geoscience Literacy Exam (GLE). Chapter 2 describes the impact that using socially relevant InTeGrate teaching materials had on changes in geoscience literacy of undergraduate students. The GLE was administered pre/post-course by all eight instructors during the project, and data were compared between the non-InTeGrate control semester and the InTeGrate treatment semester. While a significant difference in pre/post change scores on the GLE was not seen for the entire student population between the control and treatment semesters, two of the instructors did have significantly higher change scores among their students in the treatment semester. The use of InTeGrate teaching materials also had the effect of helping both female and Hispanic students close performance gaps that existed during the control semester.

The ability to think on geologic timescales is regarded as an important skill in the study of geology, yet little work has specifically addressed student understanding of this concept among geology students. Chapter 3 describes an exploratory, pre-experimental study investigating student knowledge of various geologic time concepts among geology majors. A 21-question pre/post-test was constructed to assess concepts related to landscape identification and formation rates, Earth history, the geologic timescale, and relative and absolute dating. Pre/post testing was conducted across four semesters in a variety of geology courses. Data were also collected from a group of non-majors in a Physical Geology course and from a group of geoscience faculty. Additionally, interviews were conducted with 11 senior geology majors to investigate their conceptions related to a sample of the concepts assessed. Results show that
students made the largest gains after taking Physical and Historical Geology courses and maintained their knowledge level during subsequent courses. Many geology students lacked familiarity with the geologic timescale and had difficulty estimating formation rates of landscapes formed on intermediate timespans.
Assessing Learning and Teaching across Geoscience Courses and Curricula

by

Charles Douglas Czajka

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Marine, Earth, and Atmospheric Sciences

Raleigh, North Carolina

2018

APPROVED BY:

________________________________________  ______________________________________
Dr. David McConnell                           Dr. Elana Leithold
Committee Chair

________________________________________  ______________________________________
Dr. Karen McNeal                              Dr. Karl Wegmann
DEDICATION

I dedicate this work to my grandfather. I’m glad I could share this achievement with him.
BIOGRAPHY

Should our hero’s hands be holding this blackest purse?

Mom, am I failing or worse?

Mom, am I failing?

What should these earnest hands be holding?

-Why?
ACKNOWLEDGMENTS

First and foremost I would like to acknowledge the role that my advisor David McConnell has played in getting me to this point. All of the advice and guidance he has given me over the last six years will ensure that I am successful in all of my future pursuits.

I would also like to acknowledge my other committee members, Dr. Lonnie Leithold, Dr. Karl Wegmann, and Dr. Karen McNeal. Their valuable assistance, input, and conversations on my research and dissertation have helped to strengthen it beyond what I could have done alone.

Much of my research could not have been conducted independently, and so I need to acknowledge all the members of my research group, both past and present. This includes LeeAnna Chapman, Jason Jones, and Michael Pelch. I am grateful for your help in co-coding interviews, being available to talk through research issues and brainstorm ideas, and being the best travel companions to conferences and the associated adventures (long live the Orpheum Theater & RIP David Johnston).

Education research is also not possible without willing participants, and I’d like to thank both the InTeGrate research team and all the geology majors at NCSU. Both of these groups submitted themselves to multiple instances of either being interviewed by me, or taking an assessment multiple times. This wouldn’t have been possible without your participation, so thank you.

Finally, I need to thank Rachel Atkins for all the help through this whole process, especially the end, where trying to finish writing and prepare my defense was made even more hectic by having to prepare for job interviews. She made getting through the constant 16 hour work days much more bearable.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................................. vii
LIST OF FIGURES ................................................................................................................................. viii

CHAPTER 1: The Adoption of Student-Centered Teaching Materials as a Professional Development Experience for College Faculty ................................................................. 1
Introduction .............................................................................................................................................. 1
Methods .................................................................................................................................................. 6
Results .................................................................................................................................................... 11
  Teaching Practices .................................................................................................................................. 11
  Teaching Beliefs .................................................................................................................................... 13
  Qualitative Results by TBI Questions ................................................................................................. 14
Discussion ............................................................................................................................................... 18
  Practices ............................................................................................................................................... 18
  Beliefs ................................................................................................................................................... 20
  Beliefs changes through the Interconnected Model .............................................................................. 21
Limitations ............................................................................................................................................... 23
Conclusions ............................................................................................................................................ 24
Acknowledgements ................................................................................................................................. 24
Funding .................................................................................................................................................... 25
References ............................................................................................................................................... 26

CHAPTER 2: Can Teaching the Geosciences in the Context of Societal Issues Have an Impact on Student Geoscience Literacy Gains? .................................................................................. 31
Introduction ............................................................................................................................................ 31
Methods ................................................................................................................................................. 34
  Study Participants ................................................................................................................................. 34
  Study Design .................................................................................................................................... 35
  Quantitative Instrument ...................................................................................................................... 35
Results .................................................................................................................................................... 39
Discussion ............................................................................................................................................... 45
  Demographic Factors ........................................................................................................................ 47
Limitations ............................................................................................................................................... 48
Conclusions ............................................................................................................................................ 49
Funding .................................................................................................................................................... 51
References ............................................................................................................................................... 52

CHAPTER 3: An Exploratory Study Examining Geology Students’ Conceptions Related to Geologic Time and Rates .................................................................................................................. 56
Introduction ............................................................................................................................................ 56
Research Design and Methods ................................................................................................................ 59
  Study Population ................................................................................................................................. 59
  Study Design .................................................................................................................................... 60
  Instrument Design ............................................................................................................................. 61
Data Analysis .......................................................................................................................................... 64
Results .................................................................................................................................................... 65
  Quantitative Results ........................................................................................................................ 65
  Qualitative Interview Results .......................................................................................................... 69
Alluvial Fan .................................................................................................................. 69
U-Shaped Valley ........................................................................................................... 71
Event Timeline ............................................................................................................ 72
Discussion .................................................................................................................. 74
Implications and Future Work ..................................................................................... 79
Limitations .................................................................................................................. 81
Conclusions ................................................................................................................ 82
Acknowledgements .................................................................................................... 82
References .................................................................................................................. 83
Appendices .................................................................................................................. 87
Appendix A .................................................................................................................. 87
Appendix B .................................................................................................................. 92
LIST OF TABLES

CHAPTER 1: The Adoption of Student-Centered Teaching Materials as a Professional Development Experience for College Faculty

Table 1  Study Participants .............................................................................................................. 6
Table 2  Teaching Practices Inventory (TPI) categories and points ............................................. 8
Table 3  Teaching Beliefs Interview Questions ............................................................................ 10

CHAPTER 2: Can Teaching the Geosciences in the Context of Societal Issues Have an Impact on Student Geoscience Literacy Gains?

Table 1  Study Participants ........................................................................................................ 34
Table 2  InTeGrate modules and units utilized by the participating instructors  ...................... 35
Table 3  Test statistics from the pre-test administration during the control semester. Sample size n = 343 students. The desired values are those suggested by Ding and Beichner (2009). * = correlation of difficulty and discrimination indices (respectively) between semesters .......................................................................................... 38
Table 4  Summary of ANOVA results on GLE normalized change scores. Asterisks indicate significant main effects (*) and significant interaction effects (**) at the 95% confidence level .................................................................................................................. 40

CHAPTER 3: An Exploratory Study Examining Geology Students’ Conceptions Related to Geologic Time and Rates

Table 1  Participant Demographic Characteristics. Fr = Freshman, So = Sophomore, Ju = Junior, Se = Senior, M = Male, F = Female ................................................................. 60
Table 2  Subscales and concepts present on the pre-post assessment instrument ................. 62
Table 3  Test statistics calculated from the post-test administration in Historical Geology. Sample size n = 52 students. The desired values are those suggested by Ding and Beichner (2009) ............................................................................................................ 63
Table 4  Themes identified from each question during think aloud interviews and the number of students who were coded for each particular sub-theme ......................... 70
LIST OF FIGURES

CHAPTER 1: The Adoption of Student-Centered Teaching Materials as a Professional Development Experience for College Faculty

Figure 1  The Interconnected Model of Professional Growth from Clarke and Hollingsworth (2002) ......................................................... 4

Figure 2  Teaching Practices Inventory scores for each instructor .............................................. 12

Figure 3  Reformed Teaching Observation Protocol scores from seven of the eight instructors. One instructor was not observed .............................................. 12

Figure 4  Teacher Beliefs Interview scores for all four interviews from each instructor ...... 13

Figure 5  Histograms showing the distribution of codes on the pre- and post-InTeGrate interviews for all seven TBI questions. Questions in the left column show substantial change from pre- to post-ITG, questions in the middle column show moderate change, and questions in the right column show no change.................... 14

CHAPTER 2: Can Teaching the Geosciences in the Context of Societal Issues Have an Impact on Student Geoscience Literacy Gains?

Figure 1  Line plot showing mean normalized change scores for male and female students from control to treatment. Population sizes: Control Male (n=88), Control Female (n=155), Treatment Male (n=98), Treatment Female (n=135) ...... 42

Figure 2  Line plot showing mean normalized change scores for Hispanic and Non-Hispanic students from control to treatment. Population sizes: Control Non-Hispanic (n=181), Control Hispanic (n=46), Treatment Non-Hispanic (n=159), Treatment Hispanic (n=60) ................................................................. 43

Figure 3  Line plot showing mean normalized change score by student race from control to treatment semester. Population sizes: Control = White (n=133), Black (n=51), American Indian (n=8), Asian American (n=20); Treatment = White (n=120), Black (n=41), American Indian (n=4), Asian American (n=22)................. 44

Figure 4  Line plot showing mean normalized change score for by student academic rank from control to treatment semester. Population sizes: Control = Freshman (n=59), Sophomore (n=84), Junior (n=73), Senior (n=28); Treatment = Freshman (n=61), Sophomore (n=55), Junior (n=85), Senior (n=27)............... 45
CHAPTER 3: An Exploratory Study Examining Geology Students’ Conceptions Related to Geologic Time and Rates

Figure 1  Pre-Post test results from all courses and faculty experts who were only tested once. The score for each subscale of the assessment is shown within each bar. * = significant gains from pre- to post- on the subscale (p<0.05). ** = significant gains from pre- to post- on the entire assessment (p<0.0005). Subscales are scored out of 7 points, except for the Geo Time Concepts which is scored out of 13 points.................................................................66

Figure 2  Histograms showing the distribution of scores for Physical Geology students, geology majors, and faculty experts. Dashed lines represent the means of each group, and the differences between all means were significant (p<0.001)..................67

Figure 3  Percent correct for each topic area on the assessment for the final attempt by all geology majors in the study and from faculty experts.........................................................68

Figure 4  Example timelines from student interviews, a) represents a student who used specific dates to create an accurate timeline and b) is a student with inaccurate conceptions about the timing of events.................................................................73
CHAPTER 1: The Adoption of Student-Centered Teaching Materials as a Professional Development Experience for College Faculty

Introduction

Student-centered teaching strategies play an important role in increasing undergraduate student learning (Freeman et al., 2014; Freeman, Haak, & Wenderoth, 2011; NRC, 2015). Student-centered classrooms typically feature teaching strategies that require active participation from the learner, often in collaboration with peers, and may include answering questions, peer discussion, problem solving, writing, and reflection on learning (Mcconnell et al., 2017). Such strategies can also help reduce the achievement gap among student populations, and have been shown to yield positive results for minority and first-generation students (Freeman et al., 2014; Haak, HilleRisLambers, Pitre, & Freeman, 2011). Various stakeholders have called for the increased adoption of these instructional strategies in college STEM (Science, Technology, Engineering, and Mathematics) courses (Handelsman et al., 2004; NRC, 1999; PCAST, 2012). However, survey results reveal that the systematic implementation of student-centered teaching is still some distance in the future as approximately half or less of faculty across a variety of STEM disciplines report using these methods (Borrego, Froyd, & Hall, 2010; Henderson & Dancy, 2009; Manduca et al., 2017).

Short term professional development workshops that last only hours to a few days are among the least effective approaches for promoting instructional change (Garet, Porter, Desimone, Birman, & Yoon, 2001; Stes, Min-Leliveld, Gijbels, & Van Petegem, 2010). More successful programs occur over longer periods of time and require active participation that may involve observing others, trying out new practices, and reflection on the process (Desimone & Garet, 2015; Garet et al., 2001). These components of professional development may be
presented in team-based or collaborative models, where instructors work with peers or instructional coaches (Czajka & McConnell, 2016; Gast, Schildkamp, & Veen, 2017). Henderson et al. (2011) reviewed strategies to promote instructional change in undergraduate STEM education and concluded that effective change required adopting programs that are uniquely tailored to each institution, occur over a semester or longer period, and align with or change the beliefs of the instructors. The National Research Council (NRC, 2012) proposed that changing instructional practice required programs that included two of the following strategies: (1) sustained, focused efforts, lasting from 4 weeks to a semester, or longer; (2) feedback on instructional practice; and (3) a deliberate focus on changing faculty conceptions about teaching and learning.

Henderson et al. (2011) and the NRC (2012) report share two features, a relatively lengthy professional development process and a focus on the beliefs of instructors. Various hypotheses have been proposed to explain the relationship between an instructor’s pedagogical beliefs and their teaching practices. Several of these hypotheses view the relationship in a very linear fashion, where a change in beliefs is what drives a change in practice (Fullan, 1982) or, alternatively, that positive classroom outcomes resulting from instructional changes and/or reflection on practice drive a change in instructional beliefs (Eley, 2006; Guskey, 1986). The concept that beliefs are a determinant of practices is supported by research showing that instructional beliefs can impact the implementation of curricula (G. H. Roehrig & Luft, 2004; G. Roehrig & Kruse, 2005) and that changes in an instructor’s beliefs leads to a change in their practice (Ho, Watkins, & Kelly, 2001). Alternatively, teaching practices driving changes in beliefs is supported by research showing that instructors’ beliefs about teaching and learning change as the result of change in their practice (Devlin, 2003; Hativa, 2000). Ultimately, the
complex and nuanced relationship between beliefs and practices is not well understood (Devlin, 2006), and there is often a disconnect between an instructor’s teaching beliefs and their classroom practices (Mansour, 2013).

Clarke and Hollingsworth (2002) proposed the Interconnected Model of Teacher Professional Growth (Figure 1) to describe the internal and external factors that influenced the complex relationship between an instructor’s beliefs and practices. They proposed four domains that comprised key aspects of the teaching experience. In this model the instructor’s knowledge, beliefs, and attitudes (Personal Domain) are influenced by information or stimuli from external sources (External Domain, e.g. workshops, feedback from colleagues, etc.), by things that happen in the teaching environment (Domain of Practice), and by the successes or failures of the students as measured by course outcomes (Domain of Consequence; Figure 1). Unlike other models that require changes in beliefs to drive changes in practice, or vice versa, the Interconnected Model proposes that the interplay between practice, beliefs, outcomes and professional development all play significant roles in guiding instructional change and professional growth (Clarke & Hollingsworth, 2002).
Figure 1. The Interconnected Model of Professional Growth from Clarke and Hollingsworth (2002).

This study sought to investigate the efficacy of using curriculum adoption as professional development through the lens of the Interconnected Model. Eight faculty participants adapted instructional resources created for introductory geoscience courses as part of the InTeGrate (Interdisciplinary Teaching about the Earth for a Sustainable Future) project. The InTeGrate curricula were developed to promote geoscience literacy within the context of socioscientific issues and incorporate an interdisciplinary and systems thinking approach (Kastens & Manduca, 2017). These materials incorporate research-validated teaching practices, are grounded in the geoscience literacy documents (NAGT, n.d.; Wysession et al., 2012), emphasize the connections
between geoscience and society, and utilize real-world, authentic geoscience data sets. The curricula used here are divided into six modules that each focus on one geoscience topic (e.g. Human’ Dependence on Earth’s Mineral Resources). The modules are customizable, and each is composed of six units equivalent to individual lectures or lessons. Each unit features several related activities, and the modules can be broken down and incorporated into a course at a variety of scales. All materials are designed around measurable learning objectives, a variety of formative and summative assessments and student-centered classroom activities, and are publically available via the InTeGrate website (http://serc.carleton.edu/integrate).

The participating instructors were tasked with adapting and incorporating 18 units (out of a possible 36) of InTeGrate teaching materials into their courses. Instructors were free to select the units that best fit their course and were not required to use all units from a module. While not the focus of this paper, the faculty also collected data on students’ attitudes toward the geosciences, performance on a pre-post geoscience literacy exam, and performance on short essays related to interdisciplinary and systems thinking. This paper focuses on examining the impact of the teaching material adoption process on the teaching beliefs and practices of the participating faculty. The following questions were of interest to this project:

1. What impact did adopting the InTeGrate curricular materials have on the instructional practices of geoscience faculty?

2. What impact did adopting the InTeGrate curricular materials have on instructor beliefs about teaching and learning?
Methods

The participating instructors were recruited in the spring of 2015 from a pool of applicants who responded to a call for proposals. The call was sent out through various email channels associated with the National Association of Geoscience Teachers, the Science Education Resource Center (SERC, serc.carleton.edu), and the InTeGrate project. The notice was also posted as a news item on SERC-related websites and disseminated through SERC and InTeGrate social media sites. Applicants were selected based on the quality of their application, the number of students that would be impacted, the type of courses they taught, and their potential for continued use of InTeGrate materials in the future. Experience in the use of student-centered teaching strategies was not a criterion in the selection process. Selected participants attended a three-day orientation meeting in the summer of 2015 to familiarize them with the InTeGrate teaching resources and the characteristics of the research project. The initial group of participants included 11 instructors, however, three chose not to continue with the project after the orientation meeting. The remaining eight instructors taught a variety of introductory geoscience courses - Earth Science (n=2), Physical Geology (3), Environmental Science (2) and Oceans, Atmosphere and Climate (1) – at eight different institutions (Table 1).

Table 1. Study Participants.

<table>
<thead>
<tr>
<th>Pseudonym</th>
<th>Post-secondary teaching experience (years)</th>
<th>Institution type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>21-25</td>
<td>Public, Master’s</td>
</tr>
<tr>
<td>Claire</td>
<td>≤ 5</td>
<td>Private, Master’s</td>
</tr>
<tr>
<td>Dennis</td>
<td>16-20</td>
<td>Public, Doctoral</td>
</tr>
<tr>
<td>Ellen</td>
<td>11-15</td>
<td>Public, Associate’s</td>
</tr>
<tr>
<td>Ian</td>
<td>6-10</td>
<td>Public, Doctoral</td>
</tr>
<tr>
<td>John</td>
<td>6-10</td>
<td>Public, Master’s</td>
</tr>
<tr>
<td>Owen</td>
<td>6-10</td>
<td>Public, Doctoral</td>
</tr>
<tr>
<td>Sarah</td>
<td>11-15</td>
<td>Private, Doctoral</td>
</tr>
</tbody>
</table>
The research project occurred in three phases over the subsequent three semesters. The first semester (Fall 2015) served as a control where participants taught their courses using the materials and methods that they had employed in prior versions of the class. During the subsequent spring semester, participants piloted a revised version of their course by incorporating InTeGrate materials and replacing many of their standard materials. The pilot semester allowed the instructors to become familiar with the InTeGrate materials and their use and to address any issues that might arise when using new materials for the first time. The final semester (Fall 2016) served as the treatment semester with instructors using the InTeGrate units in their course for the second time. The research project compared learning outcomes among students in the two fall classes (2015 vs. 2016).

Data on teaching practices were collected during the control and treatment semesters using both a survey instrument and direct classroom observations. The Teaching Practices Inventory (TPI, Wieman & Gilbert, 2014) was used to survey instructors before and after the control and treatment semesters. The TPI surveyed instructor use of a wide range of teaching practices across eight categories ranging from in-class activities to the nature of assignments and utilization of teaching assistants (Table 2). It includes a rubric which yields a TPI score ranging from 0 to 67. Not every item on the survey generates points, but practices with research evidence that they support learning score one point and practices with evidence supporting large or robust benefits score two or three points (Wieman & Gilbert, 2014).
Table 2. Teaching Practices Inventory (TPI) categories and points.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>Course Information provided to students</td>
<td>6</td>
</tr>
<tr>
<td>II</td>
<td>Supporting materials provided to students</td>
<td>7</td>
</tr>
<tr>
<td>III</td>
<td>In-class features and activities</td>
<td>15</td>
</tr>
<tr>
<td>IV</td>
<td>Assignments</td>
<td>6</td>
</tr>
<tr>
<td>V</td>
<td>Feedback and testing</td>
<td>13</td>
</tr>
<tr>
<td>VI</td>
<td>Other (pre/post-test use, etc.)</td>
<td>10</td>
</tr>
<tr>
<td>VII</td>
<td>Training and guidance of teaching assistants</td>
<td>6</td>
</tr>
<tr>
<td>VIII</td>
<td>Collaboration and sharing in teaching</td>
<td>4</td>
</tr>
</tbody>
</table>

Teaching observations were conducted using the Reformed Teaching Observation Protocol (RTOP, Sawada & Piburn, 2002) along with a supplemental rubric used to ensure greater reliability among observers (Budd, van der Hoeven Kraft, McConnell, & Vislova, 2013). The RTOP is a 25-item observational protocol used to characterize the level of reformed teaching during a class period. The items are grouped into five sub categories: (1) lesson design and implementation, (2) content propositional knowledge, (3) content procedural knowledge, (4) classroom culture communicative interactions, and (5) classroom culture student-teacher relationships. Each item is scored from 0 if it never occurred to 4 if was very descriptive of that lesson. The RTOP provides a total score from 0-100, with scores equal to or less than 30 representing teacher-centered instruction, 31-49 being considered transitional, and lessons scoring 50 or greater are considered student-centered (Budd et al., 2013).

RTOP observations were conducted by the PI and four other observers, all who were trained in the use of RTOP as part of the On the Cutting Edge sponsored Classroom Observation Project (Science Education Resource Center, 2014). Due to the geographic distribution of the participating instructors and the availability of trained observers, not all participants were observed during both the control and treatment semesters. Five participants were observed during the control semester while seven were observed during the treatment semester. While the
RTOP and TPI measure two different constructs, there are some similarities between the two instruments. The practices an instructor reports using in Category III of the TPI (In-class features and activities) would likely influence observational data from the RTOP’s categories of Content Procedural Knowledge and Classroom Culture Communicative Interactions. Instructors who report frequent use of student-centered practices on the TPI would likely score higher in these two RTOP categories which deal with student engagement in classroom activities and student-student interactions. Conversely, the TPI surveys a range of strategies that would not be captured by a classroom observation (e.g., learning objectives provided to students, the nature of work assigned outside of class). Similarly, the RTOP captures aspects of student-centered instruction (e.g., whether a lesson was directed by student ideas, the nature of teacher-student relationships, etc.) that would not be recorded by the TPI.

Qualitative interview data regarding the teaching and learning beliefs of the instructors were collected using the Teacher Beliefs Interview (TBI, Luft & Roehrig, 2007). The PI conducted four TBI interviews with each participant prior to each semester and at the end of the project. The TBI is a seven question, semi-structured interview designed to capture an instructor’s beliefs about teaching practice, student learning, and assessment (Table 3). All interviews were audio recorded, transcribed, and coded using the coding protocol provided by Luft and Roehrig (2007). The coding protocol utilizes thematic analysis whereby one of five category classifications is assigned to each question’s response. The five categories ranging from teacher-centered to student-centered are Traditional, Instructive, Transitional, Responsive, and Reformed. By assigning a numeric integer to these ordinal codes, a TBI score can be generated for each interview. Additionally, instructors were asked two supplementary questions during the final interview to reflect on their experience (i.e. Do you feel that the Integrate project has
impacted your teaching in any way? Did you use any new teaching practices while you were employing the Integrate materials?).

**Table 3. Teaching Beliefs Interview Questions.**

<table>
<thead>
<tr>
<th>#</th>
<th>Question</th>
<th>Question Focus</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>How do you maximize student learning in your classroom?</td>
<td>(Teaching Practice)</td>
</tr>
<tr>
<td>2</td>
<td>How do you describe your role as a teacher?</td>
<td>(Teaching Practice)</td>
</tr>
<tr>
<td>3</td>
<td>How do you know when your students understand?</td>
<td>(Assessment)</td>
</tr>
<tr>
<td>4</td>
<td>How do you decide what to teach or what not to teach?</td>
<td>(Teaching Practice)</td>
</tr>
<tr>
<td>5</td>
<td>How do you decide when to move on to a new topic in your class?</td>
<td>(Assessment)</td>
</tr>
<tr>
<td>6</td>
<td>How do your students best learn science?</td>
<td>(Student Learning)</td>
</tr>
<tr>
<td>7</td>
<td>How do you know when learning is occurring in your classroom?</td>
<td>(Student Learning)</td>
</tr>
</tbody>
</table>

Two procedural steps were used to ensure reliability in the coding process. First, to eliminate confirmation bias, no interviews were coded until the end of the study, and all interviews were anonymized and randomized to prevent coders from knowing the instructor’s identity and which of their four interviews was being coded. Second, to establish reliability of the PI as primary coder, co-coding was conducted with two additional raters. Both co-raters were independent of the research, and while one had experience coding TBIs, the other was new to the process. Co-coding was done four interviews at a time, with all three raters coding each interview. Using the numeric integer assigned to the five ordinal codes, each raters’ coding for each question of the four interviews was used to calculate a reliability measure in IBM SPSS 24 (Statistical Package for the Social Sciences Software) using a two-way mixed intra-class correlation coefficient. Co-coding of the first four interviews resulted in a single measures reliability value of 0.638, indicating substantial agreement (Landis & Koch, 1977). After co-coding the second group of four interviews, the single measures reliability was 0.81, indicating almost perfect agreement. This reliability value remained consistent after a third round of co-coding, at which point the remaining 20 interviews were independently coded by the PI.
Results – Teaching Practices

The instructors showed little to no change in their scores on the self-reported Teaching Practices Inventory (TPI) over the course of the project (Figure 2). There was a statistically insignificant mean increase from pre- to post-scores of 2.38 points ($t(7)=2.13$, $p=0.071$). The observations of instructors’ classroom practices indicate that they became more student-centered and earned significantly higher RTOP scores. Five of the instructors were observed prior to adopting the InTeGrate materials and the nine lessons taught in Fall 2015 had a mean RTOP score of 35.9 (Figure 3). In contrast, seven instructors were observed using the InTeGrate materials in Fall 2016 and their nine lessons had a mean score of 48.1 (Figure 3). This represents a statistically significant mean increase of 12.2 RTOP points ($t(8)=4.12$, $p=0.003$). Instructors scored a mean 3.69 points higher (out of 20) in the Classroom Culture: Communicative Interactions category when using the InTeGrate teaching materials. The other four categories all showed smaller mean increases of 2.03-2.29 points.
Figure 2. Teaching Practices Inventory scores for each instructor.

Figure 3. Reformed Teaching Observation Protocol scores from seven of the eight instructors. One instructor was not observed.
Teaching Beliefs

Figure 4 shows the TBI scores on all four interviews for each of the eight instructors. Six of the eight participants made at least a modest shift toward more student-centered beliefs (<6 TBI points) after the use of InTeGrate teaching materials, and two of those instructors made substantial shifts toward more student centered beliefs (≥6 TBI points, Alan and John). While one instructor, Claire, showed an initial sharp drop in TBI over the first two interviews, her scores showed small increases over the remainder of the project. She began with a high initial TBI of 27, dropped down to 18 on her second TBI and increased to 19 on her third interview and 21 on the final TBI. The other instructor to not show a directional change was Owen, who fluctuated between a score of 22 on his first and third interviews, and 18 and 19 on his second and fourth.

Figure 4. Teacher Beliefs Interview scores for all four interviews from each instructor.

Histograms displaying the proportion of all codes pre-InTeGrate (interviews one and two) and post-InTeGrate (interviews three and four) for each TBI question from all instructors.
are illustrated in Figure 5. The most substantial shift in beliefs toward a student-centered perspective occurred in response to the first two questions (How do you maximize student learning in your classroom? How do you describe your role as a teacher?). The responses to three questions showed a moderate shift toward more student-centered beliefs (How do you know when your students understand? How do your students learn science best? How do you know when learning is occurring?), while two questions saw no real change in the beliefs of the instructors (How do you decide what to teach and what not to teach? How do you decide when to move on to a new topic in your class?).

**Figure 5.** Histograms showing the distribution of codes on the pre- and post-InTeGrate interviews for all seven TBI questions. Questions in the left column show substantial change from pre- to post-ITG, questions in the middle column show moderate change, and questions in the right column show no change.

**Qualitative Results by TBI Question**

How do you maximize student learning in your classroom? During the first two interviews, almost all instructors (81%) provided responses that coded as transitional. Most of
the instructors talked about creating a hands-on classroom with activities that involved the student, with a focus on the importance of the activities. For example, in his first interview, Alan said ‘I try to get them to be as hands on as possible where we’re doing activities, activities where maybe they’re making calculations, or looking at maps and interpreting maps, or they’re reading essays…’ Many would also mention the use of activities that involved electronic response systems (clickers) or quizzes.

During interviews three and four, 50% of the responses were coded as responsive (Figure 5). Many of the instructors stressed not just the importance of activities for maximizing learning, but also the interaction between students in the classroom during activities. During his final interview, Alan’s reply to the question was ‘I think one of the things is to try to get the students to interact with each other so they can sort of bounce ideas or what they understand off of each other and figure it out in that way.’

*How do you describe your role as a teacher?* While a majority of responses to this question coded as transitional both before and after InTeGrate adoption, 44% of the post-InTeGrate responses received a student-focused code of either responsive or reformed (Figure 5). Owen’s pre-InTeGrate responses contained elements that were traditional, focusing on information delivery ‘My role as a teacher, I think it is to be a conduit of knowledge. I need to make sure that I’m doing a good job, a complete job, in transferring the knowledge that we know.’ His responses on the post-InTeGrate interviews were more typically transitional focusing on the idea of student understanding and development of skills such as critical thinking, but he also gave a reformed coded response in his third interview stressing the importance of students’ backgrounds and interests and using that to connect with the content, ‘I try to make it relevant…I have some of those business majors, pre-nursing, so, I think, people that you want to be
interested in…But I can relate that as well to their majors…it will make them be more interested and aware of different consequences of an activity…’

*How do you know when your students understand?* The instructors showed a moderate positive shift on this question, with half of all the pre-InTeGrate responses being coded as instructive (Figure 5), focusing on correct responses on questions, quizzes, or exams. As Dennis said in his first interview ‘The two main ways of seeing that are with the course response system answers and with answers on the exam.’ On his third interview, he gave a more transitional response, where students are engaged and active as a sign of understanding ‘…the level of noise in the classroom, the students speaking with me and with them to understand what they’re getting and what they’re not getting.’

*How do your students learn science best?* Responses showed a moderate shift toward a student focused mindset with a 20% increase in the number of responsive codes (Figure 5). A transitional response was the most common answer on the early interviews as instructors expressed some form of students learn science best by doing it. As John said on his first interview ‘I feel like if they’re actually doing something, I mean if they’re actually seeing it and going out in the field and doing stuff like that.’ Post-InTeGrate more responses were coded as responsive, indicating that students should not only be doing something procedural, but that they should also be interpreting their observations and drawing their own conclusions about them.

*How do you know when learning is occurring in the classroom?* This question showed a moderate shift toward student-focused beliefs. Whereas half the responses coded as instructive on the pre-InTeGrate interviews, this shifted to half the responses as transitional in the later interviews (Figure 5). During interviews one and two, John expressed the instructive idea that he only knows based on their responses, ‘…from how they respond to questions, and then also once
I give back all their quizzes and exams, that’s when I really know if they learned it or not.’

During interviews three and four, John seemed to have moved toward more student-centered beliefs, expressing the importance of student initiated interactions about the topic as indicators of learning: ‘…when they start asking questions beyond what I specifically wanted them to know. That’s when I start to understand that they’re at least getting interested in it, and that interest – actually, to me, it shows that they’re actually learning something.’

*How do you decide what to teach and what not to teach?* The distribution of codes to this question remained relatively unchanged from control to treatment semesters (Figure 5). Half of all the responses coded as transitional, with instructors mentioning some decision based on what they think their students will be interested in. The other half are split between instructive codes where the decision is made based on their own interest or comfort, and responsive codes where the instructors are using student feedback to partially decide what they will teach.

*How do you decide when to move on to a new topic in your class?* This question showed little change throughout the project (Figure 5). Responses to this question received the whole range of codes, with almost 40% coding as responsive from the pre-InTeGrate interviews, a number that drops to around 20% in the post-InTeGrate interviews. Five of the authors remained very consistent in their beliefs on this question, whereas Claire and Owen went from expressing responsive beliefs on their first interview to traditional beliefs on their last. Alan, who gave a traditional response on this first interview, was the only instructor to show a shift in beliefs on this question. During his final interview, he talked about how he would use a final discussion or debriefing to see if students had understood the material and could apply it to their observations of local or regional phenomenon as his basis for moving on.
Discussion - Practices

While the dataset on practices from this project is limited, it reveals that the instructors as a group were teaching in a more student-centered way when using the InTeGrate materials. Using the three categories described by Budd et al. (2013), the 48.1 mean RTOP score of the instructors using the InTeGrate materials would be classified as transitional, just a few points below the boundary with student-centered practices. The group’s mean RTOP score during the control semester (35.9) was below the average scores reported for instructors in geoscience classrooms by both Budd et al. (2013) and Teasdale et al. (2017; \(\bar{x} = 41.5\) and \(\bar{x} = 39.6\) respectively). Based on observer notes, Teasdale et al. (2017) describe mean transitional classrooms (RTOP score = 36-41) as commonly utilizing ‘shout out’ response questioning, often with little wait time, some student interaction to make predictions or hypotheses, and instructors reviewing previously covered material but not assessing student knowledge of that material. These practices were commonly observed in the instructors during the control semester. In the treatment semester, the instructors were beginning to use more of the practices seen in student-centered classrooms such as students interacting with each other in groups, students asking questions and volunteering ideas, utilizing data, graphs, and maps, and instructors circulating the room to formatively assess student understanding (Teasdale et al., 2017).

This shift toward more student-centered teaching when using InTeGrate materials is encouraging but there is a discrepancy between this observation data and the self-report (TPI) which no significant change in the reported practices. Differences in what the two instruments measure may partially explain this inconsistency. While the RTOP is assessing the degree of student-centered practices during a class, the TPI also scores aspects of the course that occur outside the classroom (Table 2). Owen, for example, reported using more effective in-class
activities on his post-TPI, an outcome that is likely recorded by his much higher post-RTOP score, but he also reported that he no longer used articles from the scientific literature, assigned homework, and he used fewer short answer questions on exams. These tradeoffs may have kept his TPI score relatively similar while explaining the increase in his RTOP score. Similar trends were not observed for all instructors though, so this explanation doesn’t fully explain the relatively stable TPI scores and increased RTOP scores.

Beyond the inherent contrasts between the instruments, the inconsistency in scoring patterns may be explained by how effectively new teaching strategies were employed. Smith et al. (2014) suggested that the TPI is limited in its ability to distinguish between the quality of use of a given research-based instructional strategy. For example, an instructor may have reported using small group discussions on the TPI both before and after InTeGrate adoption, but any difference between the quality of discussion or amount of time students are given to achieve their goals would likely be discerned by application of the RTOP instrument in both versions of the class. The more highly structured learning activities embedded in the InTeGrate materials may have played a role in the improved quality of use hinted at in the RTOP observations.

The fact that the use of InTeGrate materials accounted for less than half of the total class time may also have contributed to the inconsistency in scoring patterns. The instructors in this study each adopted activities from 18 units of InTeGrate teaching materials, and each unit is designed for a 50-minute class. Consequently, these resources were present in a maximum of approximately 40% of lessons for each course. It is possible that teaching with InTeGrate materials for only part of the class time was insufficient for the instructors to perceive they had significantly changed their practices.
Beliefs

The shift toward more student-centered beliefs for six of the eight instructors involved in the project is indicative that participation represented a positive professional development experience. The greatest change seen in the instructors’ beliefs about teaching and learning came on TBI questions related to teaching practice (Figure 5). This is the area where one would expect to see the greatest shift based on involvement in a project where new, student-centered teaching materials are being adopted. The RTOP data provides evidence that adoption of the InTeGrate materials resulted in these instructors using more student-centered teaching methods. Moderate to no changes in beliefs were seen on TBI questions related to assessment and student learning. It is perhaps not surprising that no change was seen in the instructors’ beliefs in these areas (i.e. How do you decide what to teach? How do you decide when to move on to a new topic?) when they are being asked to teach using a specific suite of curricular resources within the context of a standard semester/term.

Claire was the only instructor whose beliefs appeared more teacher-centered after using the InTeGrate materials. This was a consequence of the fact that she started out with a single high TBI score on her first interview (28), while her subsequent interview scores trended upward but were considerably lower (18, 19, and 21). Claire had to teach a completely new course during the treatment semester as part of a college level redesign that eliminated her original course and added a new one in its place. Claire was also the least experienced instructor (Table 1), and based on evidence from beginning secondary science teachers (J. A. Luft & Zhang, 2014), her beliefs may have been more mutable as a result.
Belief changes through the Interconnected Model

If we examine the instructors’ experiences through the lens of the Interconnected Model of Professional Growth (Figure 1), there is evidence of interactions between multiple domains for many of the instructors. The willingness to participate in adopting InTeGrate materials represented engagement between the External Domain and the Domain of Practice and represents a willingness to change some aspects of their teaching. Evidence for subsequent change in the other domains is apparent from interview responses and is reflected in the observation data that indicates specific changes in teaching practices. For example, the InTeGrate materials required the application of teaching practices that the instructors may not have used before. When asked about new teaching practices that he used during the project, Alan reflected on the use of jigsaw activities and gallery walks:

‘And then, you know, it sort of shows an understanding when one of the groups is looking at something that maybe a bunch of things that other people have put down and then they’re sort of synthesizing that with what they’ve already thought. So, I think that’s been very effective, which is something I had not done before.’

Here Alan is reflecting on the outcome of these practices and seeing the value in these activities for student learning. This reflection has impacted Alan’s beliefs, as evidenced from his response to question seven on the final interview where he expressed the belief that he knows learning is occurring:

‘…when they’re working in groups, the willingness to sort of engage in discussion within the group and then even often times take it beyond, because with the different groups sometimes they finish the actual assignment at different times and then they’ll sort of take the discussion a little further.’

Many of the InTeGrate modules feature activities that utilize real world data that students analyze and interpret. John reflected on the use of these activities in his final interview ‘the plate tectonics one where we actually get real data from a volcanic activity…I always thought it was
almost too in-depth for the students, but when I tried it, it actually worked better than I thought it would.' This illustrates how John’s use of a new practice and reflection on the resulting outcome (Domain of Consequence) may have altered his belief (Personal Domain) on what students are capable of in his classroom. For Ian, who held more student-centered beliefs coming into the project, the use of activities involving real data was seen as prohibitive due to situational factors, ‘I definitely didn’t teach with real data nearly as much in my large intro class, simply because it’s a challenge to do that in a class where you have at least 60 or more than 100 people, and the learning curve is steep, and things can easily go off the rails.’ Yet, when reflecting on using data-rich resources he mentioned that having the activities and data pre-packaged were a great help, ‘I think that was really important for getting me to take the plunge to try it. I think that it improved the student experience a lot.’ For Ian, the External Domain had influenced the Domain of Practice, bringing his practice more in line with his beliefs. For both John and Ian, using these activities were positive experiences that changed how they viewed both the practices they were comfortable using and the student outcomes of such practices.

When asked on the final interview if involvement in the InTeGrate project had impacted their teaching, two common themes emerged from the instructor responses. One was the desire to use practices that would encourage more student activity and engagement, a theme mentioned by four of the instructors. For Ellen, she was ‘now more interested in hands-on activities, interactions, you know, moving the students around the classroom, not so concerned with memorizing.’ The other main theme mentioned by five of the instructors in response to this question was how they valued the relevancy that the InTeGrate materials brought to the geosciences in their classroom. Ian felt that ‘a lot of the InTeGrate activities really gave me a way to incorporate into my teaching these sorts of social, economic, and political aspects of the
application of science, and I think that that was something that’s really strengthened and improved the relevance of what I’m bringing to the student that’s not a geology major that hopefully they’ll carry with them outside of their college experience.’ Reflecting on the enactment of new practices and the resulting outcomes has resulted in changes to the instructors’ personal beliefs and attitudes about their teaching and student learning. These interactions between multiple domains of the Interconnected Model imply that this was a positive experience of professional growth for the instructors.

**Limitations**

This study is limited due to the non-random method by which the instructors who participated were recruited. The recruitment process used was also likely to attract participants who were already familiar with the InTeGrate project and teaching materials. Applicants would have likely been on one of the SERC, NAGT, or InTeGrate related email lists or visited a related website. By having this connection to the geoscience education community, applicants were potentially more open to or interested in changing their teaching practices. This may have had the consequence of instructors with more malleable beliefs being recruited for participation.

The study is also limited by size of the participant population (n=8) and the limited data collected on teaching practices. A larger observations dataset would have allowed for more robust conclusions to be made concerning changes in teaching practices and also provided a more detailed look into the relationship between self-report of practices versus direct observation. Obtaining a larger observational data set was difficult due to the geographic separation between the instructors and trained RTOP observers.
Conclusions

Implementation of InTeGrate teaching materials led to a shift toward more student-centered teaching practices and beliefs for nearly all the group of eight instructors involved in this project. This suggests that the yearlong experience of adopting well crafted, student-centered teaching materials can be an effective professional development experience for college faculty and can promote change toward more student-centered teaching beliefs and practices. Changes in the beliefs of the instructors may ensure that these instructors will persist in the use of student-centered teaching practices that align with their newfound beliefs.

There is now an even more diverse suite of eleven InTeGrate modules composed of 65 units or lessons (https://serc.carleton.edu/75235). These customizable curricular materials incorporate a variety of teaching strategies that fall under the banner of ‘student-centered instruction’ and can help provide a pathway to guide more instructors toward the use of effective teaching strategies that support student learning. The challenge lies in finding ways to get them into the hands and classrooms of college geoscience faculty to better facilitate the transformation of undergraduate STEM education

Acknowledgements

We would like to thank the faculty instructors for their willingness to participate in this project and allow us to use their data. We would also like to thank Elizabeth Nagy-Shadman for her work with the research team and our other colleagues within the InTeGrate project for their feedback and support.
Funding

This work was supported by the National Science Foundation award 1125331. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the National Science Foundation.
References


United States of America, 111, 8410–5.


Science Education Resource Center. (2014). Classroom Observation Project: Understanding and


CHAPTER 2: Can Teaching the Geosciences in the Context of Societal Issues Have an Impact on Student Geoscience Literacy Gains?

Introduction

Human society is facing a number of grand challenges, many of which are relevant to the geosciences and will require geoscientific solutions. These challenges include availability of mineral resources (Vidal, Rostom, François, & Giraud, 2017), energy needs and security (Thomas, Partridge, Harthorn, & Pidgeon, 2017), anthropogenic induced climate change (Mann et al., 2017), and the availability of clean water (Flörke, Schneider, & McDonald, 2018) among others. The Earth is a dynamic system, and interactions between the various spheres of the system link many of these challenges (Flörke et al., 2018; Holland et al., 2015). Additionally, our desires and needs for resources are often entangled in political policy (Ryan, 2017) and can have severe impacts for developing nations (Maystadt, De Luca, Sekeris, & Ulimwengu, 2014). It is vital that we provide the global citizenry with the geoscience literacy needed to properly evaluate, discuss, and address these challenges.

Ocean researchers were concerned that the National Science Education Standards (National Research Council, 1996) released more than twenty years ago contained minimal mention of topics related to the ocean. This led to an effort by a coordinated group of educators, scientists, policy makers, and federal agencies to draft a list of principles that students and the public should know in order to be literate about the oceans (Schoedinger, Tran, & Whitley, 2010). They produced a list of seven principles of ocean literacy (NOAA, 2006) and this effort inspired similar collaborations that resulted in the subsequent creation of documents on atmospheric science literacy (UCAR and CIRES, 2008), climate literacy (USGCRP, 2009), and
Earth science literacy (Wysession et al., 2012). An important focus of these literacy initiatives is dissemination that seeks to promote the incorporation of the literacy principles into formal educational settings.

For many undergraduate students, an introductory geoscience course may represent the last exposure they have to learning about the Earth system. It then becomes imperative that these courses provide students with the literacy needed to evaluate geoscience-related challenges and to make scientifically informed decisions regarding these issues. This task is made difficult by the fact the most students who enroll in introductory geoscience courses are non-science majors and select the course mainly to fulfill a general education requirement. As such, these students come into geoscience courses with low motivation (Gilbert et al., 2012). It becomes important then, that geoscience instructors are intentional in deciding how they will teach their courses in order to engage these students and provide them with the knowledge and skills to think scientifically.

InTeGrate (Interdisciplinary Teaching about the Earth for a Sustainable Future) was a multiyear National Science Foundation STEP Center project with a focus on geoscience literacy. The project had two main goals: 1) Increase the geoscience literacy of all undergraduate students so that they are better able to evaluate and make informed decisions on global sustainability issues; 2) Increase the number of geoscience majors, thereby building a future workforce capable of addressing the variety of geoscience grand challenges facing society. A strategy that the InTeGrate program has utilized to achieve these goals, is to create a suite of freely available teaching materials aimed at teaching the geosciences in the context of socioscientific issues and that incorporate an interdisciplinary and systems thinking approach (Kastens & Manduca, 2017). These materials are grounded in the geoscience literacy documents, incorporate research-
validated teaching practices, emphasize the connection between geoscience and society, and utilize real-world, authentic geoscience data (Pelch & McConnell, 2017). Some of these materials were designed for introductory geoscience courses and are divided into modules that each focus on one geoscience topic (e.g. Humans’ Dependence on Earth’s Mineral Resources) and contain three to six units that are each equivalent to one individual lecture period or class lesson. All of the materials are designed around measurable learning objectives, student-centered classroom activities, a variety of formative and summative assessments, and are publicly available via the InTeGrate website (http://serc.carleton.edu/intergrate).

The study described herein sought to investigate the efficacy of using the InTeGrate teaching materials on the geoscience literacy of undergraduate students at a variety of post-secondary institutions. Eight participating faculty members were tasked with incorporating approximately 18 units of InTeGrate teaching materials into their courses, and were free to select from up to 39 units that best fit their courses. Student geoscience literacy was assessed using a 16 question pre-post geoscience literacy exam. While not the focus of this paper, the faculty also collected data on students’ attitudes toward the geosciences, performance on a variety of module related assessments, and performance on two short essays related to interdisciplinary and systems thinking. The following research question guided this project:

1. Does the use of InTeGrate teaching materials lead to greater gains in the geoscience literacy of undergraduate students as measured using the Geoscience Literacy Exam?

2. Do the effects of InTeGrate material use on Geoscience Literacy Exam scores vary depending on the demographic factors of gender, ethnicity, race, or academic rank?
**Methods - Study Participants**

Geoscience instructors were recruited to participate in the spring of 2015 from a pool of applicants who responded to a call for proposals. The call was sent out through various email channels associated with the National Association of Geoscience Teachers, the Science Education Resource Center (SERC, serc.carleton.edu), and the InTeGrate project. The notice was also posted as a news item on SERC-related websites and disseminated through SERC and InTeGrate social media sites. Applicants were selected based on the quality of their application, the number of students that would be impacted, the type of courses they taught, and their potential for continued use of InTeGrate materials in the future. Experience in the use of student-centered teaching strategies was not a criterion in the selection process. Selected participants attended a three-day orientation meeting in the summer of 2015 to familiarize them with the InTeGrate teaching resources and the characteristics of the research project. The initial group of participants included 11 instructors, however, three chose not to continue with the project after the orientation meeting. The remaining eight instructors taught a variety of introductory geoscience courses at eight different institutions. Instructor pseudonyms as well as information about their teaching experience, institution, and courses are provided in Table 1.

**Table 1. Study Participants.**

<table>
<thead>
<tr>
<th>Participant Pseudonym</th>
<th>Teaching Experience (years)</th>
<th>Institution type</th>
<th>Course Taught</th>
<th>Course Size (# of students)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alan</td>
<td>21-25</td>
<td>Public, Master’s</td>
<td>Environmental Science</td>
<td>30-32</td>
</tr>
<tr>
<td>Claire</td>
<td>≤ 5</td>
<td>Private, Master’s</td>
<td>Physical Geology</td>
<td>12-24</td>
</tr>
<tr>
<td>Dennis</td>
<td>16-20</td>
<td>Public, Doctoral</td>
<td>Earth Science</td>
<td>88-96</td>
</tr>
<tr>
<td>Ellen</td>
<td>11-15</td>
<td>Public, Associate’s</td>
<td>Physical Geology</td>
<td>36-39</td>
</tr>
<tr>
<td>Ian</td>
<td>6-10</td>
<td>Public, Doctoral</td>
<td>Physical Geology</td>
<td>65-116</td>
</tr>
<tr>
<td>John</td>
<td>6-10</td>
<td>Public, Master’s</td>
<td>Earth Science</td>
<td>32</td>
</tr>
<tr>
<td>Owen</td>
<td>6-10</td>
<td>Public, Doctoral</td>
<td>Oceans, Atmosphere, and Climate</td>
<td>45-109</td>
</tr>
<tr>
<td>Sarah</td>
<td>11-15</td>
<td>Private, Doctoral</td>
<td>Environmental Science</td>
<td>20-30</td>
</tr>
</tbody>
</table>
At the time of the study, seven completed InTeGrate modules were available for use by the instructors. Each module contains six units of material (except Map Your Hazards which has three units), and each unit was designed as a single lesson for one 50 minute class period. It was suggested that the instructors adopt 18 total units of material, and they were free to select the units that best fit their course. A brief description of each module can be seen in Table 2. After piloting the InTeGrate materials in their spring 2016 courses, the number of units used by the instructors ranged from 14-19 units (Table 2). Consequently, the InTeGrate resources were present in approximately 40% of the lessons for each course.

Table 2. InTeGrate modules and units utilized by the participating instructors.

<table>
<thead>
<tr>
<th>Module Description</th>
<th>Climate of Change</th>
<th>A Growing Concern: Sustaining Soil</th>
<th>Living on the Edge</th>
<th>Human’s Dependence on Earth’s Mineral Resources</th>
<th>Natural Hazards and Risks: Hurricanes</th>
<th>Environmental Justice &amp; Freshwater Resources</th>
<th>Map Your Hazards</th>
<th>Total Units Used</th>
</tr>
</thead>
<tbody>
<tr>
<td>Examines short-term climate variability through analysis of authentic data. Explores past and present climate variability as well as human adaptation to climate variability.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>19</td>
</tr>
<tr>
<td>Knowledge of soil properties are used to compare natural landscapes to managed agricultural landscapes while using geospatial data and making decisions on future soil management.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>15</td>
</tr>
<tr>
<td>Geologic hazards and risks of plate boundaries are investigated using data related to geologic processes and historic events.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>18</td>
</tr>
<tr>
<td>Instruction on rocks and minerals is imbedded in lessons on mineral resource discovery, extraction, and implications of mineral resource use.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>14</td>
</tr>
<tr>
<td>Connections between Earth systems are explored through hurricanes. Historic events are used to evaluate risks and impacts on coastal development, planning, and hurricane response.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>17</td>
</tr>
<tr>
<td>Freshwater components of the hydrologic cycle and water science are explored through various cases of environmental justice dealing with access to clean freshwater.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>16</td>
</tr>
<tr>
<td>Place-based, interdisciplinary study of hazards and social vulnerability through student collected survey data.</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3 4 5 6</td>
<td>1 2 3</td>
<td>18</td>
</tr>
</tbody>
</table>

Study Design

The research project utilized a quasi-experimental, non-equivalent groups design and occurred over three subsequent semesters following the summer 2015 meeting. The first semester (fall 2015) served as a control semester, where the participating instructors taught their courses using the materials and methods that they had employed in prior versions of the class. In
the following spring 2016 semester, participants piloted a revised version of their course by incorporating InTeGrate materials and replacing many of their standard materials. The pilot semester allowed the instructors to become familiar with the InTeGrate materials and their use, while addressing any issue that might arise when using new materials for the first time. In an effort to ensure similar student populations, the project plan was to compare assessment results among students in the two fall semesters (control 2015 vs. treatment 2016). Fall 2016 served as the treatment semester, and instructors used the InTeGrate materials for the second time while making appropriate adjustments based on their experiences in the pilot semester.

Quantitative data was collected using a pre-post geoscience literacy assessment during each semester. Students also completed a pre-post attitudinal survey (InTeGrate Attitudinal Instrument, IAI) that provided student demographic information on gender, ethnicity, race, rank, and age. Each participating instructor administered a paper version of the quantitative pre-post assessment in their courses, and all assessment data was sent to the SERC office for data processing and digitization. The IAI survey was completed online by students at the beginning and end of each course.

Quantitative Instrument

Student geoscience literacy was assessed using a modified version of the Geoscience Literacy Exam (GLE). An original eight question version of the instrument was utilized to collect baseline literacy data from other InTeGrate projects. The original GLE contained two multiple choice questions targeting each of the four literacy areas (Earth, Ocean, Climate, and Atmosphere). Those eight questions were adopted for this project in addition to eight more questions, resulting in a sixteen question instrument (Appendix A). Five of these were additional
GLE questions that targeted Earth (4 questions) and climate literacy (1 question). The remaining three questions were taken from the Geoscience Concept Inventory (GCI, Libarkin and Anderson, 2006) and targeted concepts related to plate tectonics. The rationale for including additional Earth literacy and GCI questions was to improve alignment between the assessment and InTeGrate content used by the participating instructors, since the Living on the Edge and Earth’s Mineral Resources were the most widely used modules (Table 2).

All GLE questions are classified as either level 1 or level 2. Level 1 questions are single option multiple choice questions at the “remember” and “understand” level of Bloom’s taxonomy (Bloom, Engelhart, Furst, Hill, & Krathwohl, 1956). Level 2 questions are multiple mark type questions and classify at the “applying” and “analyzing” levels of Bloom’s taxonomy. Of the 16 GLE questions contained in the instrument, nine were level 1 and seven were level 2. Multiple mark items were scored using a partial credit algorithm which improves the reliability and discriminatory power of the items versus using dichotomous scoring where all options must be marked or unmarked correctly to get full credit. The formula used is method six as described by Hsu et al. (1984), and assumes that incorrect choices are the result of guessing. Partial credit is determined by subtracting the probability for incorrect choices from the proportion of correct choices (marked answers and unmarked distractors) made.

All data analysis was conducted using IBM Statistical Package for the Social Sciences (SPSS) 25. Only data form students with paired pre- and post-test GLE results were used in the analyses. Alan’s post-GLE data from his treatment semester was incomplete, so the results from his course were omitted from data analysis. Additionally, a visual boxplot inspection of average gain scores for the entire population revealed five outliers below the minimum. These five outliers had absolute gain scores that were greater than two and a half standard deviations below
the average gain. These data points were eliminated from analysis, as it is improbable that they would regress this much after completing a geoscience course and their post-test GLE results are reflective of a lack of effort and not their true ability.

Classical Test Theory was used to establish measures of reliability and discriminatory power of the instrument for the pre-test administration during the control semester. Test statistics for three item-analysis and two test-analysis measures are provided in Table 3, along with recommended values from Kline (1998) and Ding and Beichner (2009). The difficulty, discrimination and point biserial all represent item level statistics and fall within the desired values. The Cronbach’s alpha reliability value is below the desired value, but Adams and Wieman (2011) caution that a high internal consistency is not the best indicator of reliability for an assessment that is measuring multiple constructs or concepts. They suggest a test-retest stability coefficient as an alternative measure of reliability, where retesting is done using two similar student populations as opposed to administering the same test twice to a single group. To accomplish this, pre-test values from students in the control and treatment semesters were compared. Since these populations were of different sizes (n = 343 and 254), a Pearson correlation was determined by correlating the 16 difficulty and discrimination indices form the control semester with the values from the treatment semester.

**Table 3.** Test statistics from the pre-test administration during the control semester. Sample size n = 343 students. The desired values are those suggested by Ding and Beichner (2009). * = correlation of difficulty and discrimination indices (respectively) between semesters.

<table>
<thead>
<tr>
<th>Test Statistic</th>
<th>GLE Values</th>
<th>Desired Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty Index</td>
<td>Average of 0.53</td>
<td>0.30-0.90</td>
</tr>
<tr>
<td>Discrimination Index</td>
<td>Average of 0.35</td>
<td>≥0.30</td>
</tr>
<tr>
<td>Point Biserial Coefficient</td>
<td>Average of 0.37</td>
<td>≥0.20</td>
</tr>
<tr>
<td>Reliability Index (Cronbach’s Alpha)</td>
<td>0.61</td>
<td>≥0.70</td>
</tr>
<tr>
<td>Test-Retest Stability (Pearson)</td>
<td>0.98 &amp; 0.82*</td>
<td>≥0.70</td>
</tr>
</tbody>
</table>
Results

Normalized change scores were calculated for each student with paired pre- and post-test data and used in the majority of comparison analyses. Normalized change accounts for post-test scores that are lower than pre-, by normalizing for the amount of points that could have been lost. It provides an advantage over the single formula normalized gain (post-pre/100-pre) in comparing the change in performance between groups (Marx & Cummings, 2007). Independent t-tests were run to compare pre-test, post-test, and normalized change between all students in both the control and treatment semesters. Values reported represent mean ± the standard deviation. Students in the control semester (n=327) had higher scores on the pre-GLE (11.83±3.24) than did students in the treatment semester (n=254) (11.19±3.24), a statistically significant difference of 0.64 points (t(579)=2.35, p=0.019, d=0.2). Students in the control semester also had higher post-GLE scores (13.34±3.27) than students in the treatment semester (12.83±3.55), but this was a non-significant difference of 0.51 points (t(579)=1.8, p=0.073, d=0.15). Students made significant gains from pre- to post-GLE in both the control (1.51±2.94 points, p<0.001, d=0.46) and treatment semesters (1.64±2.78 points, p<0.001, d=0.48).

Normalized change scores were non-significantly different (t(579)=-0.42, p=0.678) between students in the control (0.13±0.235) and treatment (0.138±0.231) semesters.

Factorial analyses of variance (ANOVA) were conducted to examine if the effects of InTeGrate materials on GLE normalized change scores vary depending on the instructor or on demographic characteristics of the classes. Table 4 provides a summary of the ANOVA results for both main effects and interaction effects. A significant main effect indicates that an independent variable (e.g. instructor) influences the dependent variable (GLE normalized gain) while ignoring the other independent variable (e.g. phase = control vs. treatment). A significant
interaction effect indicates that the effect of one independent variable on the dependent variable depends on the level of the other independent variable. In the case of a significant interaction effect, simple main effects were analyzed with a post hoc test to explore all pairwise comparisons among the independent variables.

Table 4. Summary of ANOVA results on GLE normalized change scores. Asterisks indicate significant main effects (*) and significant interaction effects (**) at the 95% confidence level.

<table>
<thead>
<tr>
<th>Source</th>
<th>Type III SS</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>( \eta^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phase</td>
<td>0.024</td>
<td>1</td>
<td>0.473</td>
<td>0.492</td>
<td>0.001</td>
</tr>
<tr>
<td>Instructor</td>
<td>2.347</td>
<td>6</td>
<td>7.845</td>
<td>&lt;0.001*</td>
<td>0.077</td>
</tr>
<tr>
<td>Phase * Instructor</td>
<td>1.056</td>
<td>6</td>
<td>3.529</td>
<td>0.002**</td>
<td>0.036</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>31.572</td>
<td>580</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>0.007</td>
<td>1</td>
<td>0.132</td>
<td>0.716</td>
<td>0.0</td>
</tr>
<tr>
<td>Gender</td>
<td>0.38</td>
<td>1</td>
<td>7.245</td>
<td>0.007*</td>
<td>0.015</td>
</tr>
<tr>
<td>Phase * Gender</td>
<td>0.225</td>
<td>1</td>
<td>4.292</td>
<td>0.039**</td>
<td>0.009</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>25.397</td>
<td>475</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>0.133</td>
<td>1</td>
<td>2.568</td>
<td>0.11</td>
<td>0.006</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.359</td>
<td>1</td>
<td>6.916</td>
<td>0.009*</td>
<td>0.015</td>
</tr>
<tr>
<td>Phase * Ethnicity</td>
<td>0.133</td>
<td>1</td>
<td>2.555</td>
<td>0.111</td>
<td>0.006</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>23.415</td>
<td>445</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>0.136</td>
<td>1</td>
<td>2.562</td>
<td>0.11</td>
<td>0.007</td>
</tr>
<tr>
<td>Race</td>
<td>0.512</td>
<td>3</td>
<td>3.227</td>
<td>0.023*</td>
<td>0.024</td>
</tr>
<tr>
<td>Phase * Race</td>
<td>0.279</td>
<td>3</td>
<td>1.759</td>
<td>0.155</td>
<td>0.013</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>21.448</td>
<td>398</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phase</td>
<td>0.066</td>
<td>1</td>
<td>1.244</td>
<td>0.265</td>
<td>0.003</td>
</tr>
<tr>
<td>Rank</td>
<td>0.465</td>
<td>3</td>
<td>2.909</td>
<td>0.034*</td>
<td>0.018</td>
</tr>
<tr>
<td>Phase * Rank</td>
<td>0.006</td>
<td>3</td>
<td>0.037</td>
<td>0.99</td>
<td>0.0</td>
</tr>
<tr>
<td>Corrected Total</td>
<td>25.269</td>
<td>471</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

There was a statistically significant interaction between InTeGrate use (control vs. treatment) and instructor on normalized change score. An analysis of simple main effects for instructor revealed that there was a significant difference in normalized change scores from control to treatment for students in Owen’s (F(1,567)=9.79, p=0.002, partial \( \eta^2 \)=0.017) and Sarah’s (F(1,567)=8.35, p=0.004, partial \( \eta^2 \)=0.015) courses. Using only data from within Sarah
and Owen’s courses, independent t-tests were run to compare normalized change scores between the control and treatment semesters. These tests reveal significant differences with a moderate effect size in Owen’s course \((t(149)=−3.41, p=0.001, d=0.56)\) and a large effect size in Sarah’s course \((t(41)=−2.577, p=0.014, d=0.80)\). No other instructors saw significant GLE gains in their courses between the control and treatment semesters.

There was a statistically significant interaction effect between gender and InTeGrate use (Table 4, Figure 1). An analysis of simple main effects for gender was performed, with all pairwise comparisons being run for each simple main effect with \(p\)-values Bonferroni-adjusted within each simple main effect. Females had normalized change scores 0.052 points higher in the treatment semester than in the control semester, a slightly nonsignificant difference \((p=0.052)\). Males had normalized change scores 0.037 points lower in the treatment semester than in the control semester, which was also non-significant \((p=0.274)\). Males had normalized change scores 0.103 points higher than females in the control semester, a statistically significant difference \((F(1,472)=11.28, p=0.001, \eta_p^2=0.23)\). The difference of 0.013 points in the Treatment semester was not significant \((p=0.66)\). Examining the main effect of gender reveals that males have normalized change scores 0.058 points higher than females, a statistically significant difference. This main effect does not qualify the interaction effect since males score higher in both the control and treatment semester, but the difference was less marked during the InTeGrate semester.
Figure 1. Line plot showing mean normalized change scores for male and female students from control to treatment. Population sizes: Control Male (n=88), Control Female (n=155), Treatment Male (n=98), Treatment Female (n=135).

The interaction effect between ethnicity and InTeGrate use was not statistically significant (Table 4). However, the profile plot (Figure 2) is indicative of an interaction effect, and non-significance does not always exclude the existence of an interaction effect in the population (Fox, 2008). For this reason, Pairwise comparisons were run and p-values were Bonferroni-adjusted within each simple main effect. It was found that Non-Hispanics had significantly higher (0.108) normalized change scores than Hispanics in the control semester (F(1,442)=8.23, p=0.004, partial $\eta^2=0.018$), but not in the treatment semester (0.026 points, p=0.446). No significant difference in normalized change score was found between the control and treatment semesters for Hispanic (p=0.068) or Non-Hispanic students (p=0.997). Examining the main effect of ethnicity reveals that Non-Hispanics have normalized change scores 0.067 points higher than Hispanics, a statistically significant difference (Table 4). This main effect
does not qualify the interaction effect since Non-Hispanics score higher in both the control and treatment semester, but the difference was less marked during the InTeGrate semester.

![Mean of Normalized Change by Control or Treatment by Ethnicity](image)

**Figure 2.** Line plot showing mean normalized change scores for Hispanic and Non-Hispanic students from control to treatment. Population sizes: Control Non-Hispanic (n=181), Control Hispanic (n=46), Treatment Non-Hispanic (n=159), Treatment Hispanic (n=60).

The interaction effect between race and InTeGrate use was not statistically significant, but there was a statistically significant main effect for race (Table 4). While the profile plot (Figure 3) is indicative of an interaction effect for American Indian students, the sample sizes during the control and treatment (n=8 and 4 respectively) were very small leading to an extremely wide 95% confidence interval. Pairwise comparisons were run to investigate the main effect of race where p-values are Bonferroni-adjusted, and the only significant difference found was that white students had normalized change scores 0.106 points higher than Asian American students, a statistically significant difference (p=0.035).
Figure 3. Line plot showing mean normalized change score by student race from control to treatment semester. Population sizes: Control = White (n=133), Black (n=51), American Indian (n=8), Asian American (n=20); Treatment = White (n=120), Black (n=41), American Indian (n=4), Asian American (n=22).

The interaction effect between academic rank and InTeGrate use was not statistically significant, but there was a statistically significant main effect for academic rank (Table 4, Figure 4). Pairwise comparisons with p-values Bonferroni-adjusted revealed no significant differences, but post hoc Tukey analysis revealed that freshman had normalized change scores 0.099 points higher than seniors, a statistically significant difference (p=0.043). No other significant difference between academic ranks were present.
The use of InTeGrate teaching materials did not lead to significantly increased gains in geoscience literacy as measured by the GLE for the entire student population in the study. Students in both the control and treatments semesters made significant gains from pre to post-GLE, and while the gains seen were modest (<2 points), similar results are common when using such pre-post instruments (J. Libarkin & Anderson, 2005). One possible explanation for these low gains could be lack of alignment between the instrument and content covered in the courses and InTeGrate modules. Nine of the 16 questions included on the constructed GLE covered Earth literacy or were geology questions from the Geoscience Concept Inventory. This focus on Earth literacy questions was in response to the number of participating instructors opting to use
the Living on the Edge (tectonics) and Earth’s Mineral Resources modules. The remaining 7 questions dealt with concepts related to climate, the atmosphere, and the oceans, which may have been topics that many of the courses discussed less frequently. A lack of direct instruction on assessed concepts could explain the modest gains from pre- to post-GLE in the study.

Two of the instructors, Owen and Sarah, did see greater literacy gains from pre- to post-course during the treatment semester when using the InTeGrate teaching materials. While it is convenient to attribute the gains seen by these instructors to the InTeGrate materials themselves, the variety of factors involved in a project of this scope make it difficult to give a definitive reason for why these two instructors saw positive changes in GLE gains while the others faced neutral results. The use of the InTeGrate teaching materials could have had a significant impact on how Owen and Sarah taught their courses, moving them to a more student-centered instructional style that increased student learning. While teaching observation data from the project was limited, both Owen and Sarah were observed twice using the Reformed Teaching Observation Protocol (RTOP) and received more student-centered scores during the treatment semester (Czajka and McConnell, in review). While encouraging instructors to adopt more effective teaching strategies is a benefit of the InTeGrate teaching materials, it is difficult to disentangle the impact of teaching strategies versus content on the learning gains recorded using the GLE instrument.

Various student or course factors could also be having an impact on the positive gains seen by Owen and Sarah. They were the only two instructors who had a majority of upper class students (juniors and seniors) in their courses, both during the control and treatment semesters. This observation is interesting in light of the fact that across the entire population, freshman saw significantly higher normalized change scores than seniors on the GLE. Perhaps the high
concentration of upper class students in these courses responded favorably to the InTeGrate materials and the active learning strategies that accompany their use. Sarah and Owen were also the only two instructors to use the entire climate of change module. This is the only module that the two had in common, so it is difficult to attribute this to the gains seen on the GLE during the treatment semester by their students. Alan’s choice of modules and units most closely matched Sarah’s (table 2), but unfortunately his treatment post-test GLE data was incomplete and couldn’t be analyzed to further investigate the idea of certain modules contributing to GLE improvements.

Demographic Factors

While seeing a main effect on GLE normalized change scores due to factors of gender, ethnicity, and race is concerning, this result is not entirely surprising. Studies have shown that gendered and minority performance differences exist across the STEM disciplines (Alexander, Chen, & Grumbach, 2009; Matz et al., 2017; Nguyen & Ryan, 2008). It is encouraging that there was a significant interaction between InTeGrate use and gender on GLE normalized change scores that indicates the InTeGrate materials may have been beneficial in reducing a gendered performance gap. During the control semester there was significant gender gap with males making greater gains on the GLE from pre- to post-test. This significant difference in gains between genders is not present in the treatment semester. It should be noted that even though GLE normalized gain scores for females improved between the control and treatment, near significantly, average change for males was lower, albeit non-significantly.

Matz et al. (2017) demonstrated that while female students underperform compared to males in lecture courses across a variety of disciplines, the difference is less prevalent in lab courses. This is likely due to the benefit that collaborative learning has for female students
(Stump, Hilpert, Husman, Chung, & Kim, 2011). A similar factor is probably explaining the gendered results seen in this study, with the active learning activities embedded within the InTeGrate materials contributing to a more collaborative learning environment in the courses. Hispanic students were also able to close a performance gap that existed in the control semester, and this may have also been related to increased active learning utilized in the treatment semester. Such strategies have been shown to reduce achievement gaps faced by underrepresented minorities and first generation students in STEM (Haak, HilleRisLambers, Pitre, & Freeman, 2011). While these results are encouraging, the impact of the InTeGrate materials on these students may be lower than possible due to the dosage used in this study. The 14-18 units used by the participating instructors likely represents around 30-40% of the lessons in their course. If the remainder of the course material was being presented in a more traditional teaching style, it could have partially offset the learning gains seen by these students in the treatment semester.

There was no impact of InTeGrate use on GLE gains for students of differing races. There was a main effect for race indicating that Asian American students were making gains that were significantly less than white students regardless of InTeGrate use. It would be worth further investigating this result, perhaps with attitudinal or motivational survey data, to find out why Asian American students are facing this performance difference.

**Limitations**

The quasi-experimental nature of the study could be viewed as a limitation, but this study design is prevalent in education. It is rarely possible to achieve random assignment of students into courses, and this becomes even more difficult when control and treatment courses do not
occur in the same semester as in this study. Additionally, a piloting of the compiled GLE instrument would have helped to improve its reliability and discrimination, although the test statistics returned during the control semester are all near at above desired values. Based on the reported statistics, we remain confident in our interpretations of the results from the instrument in this study. A better effort to align the instrument with course and InTeGrate module content may have made it more effective at detecting student learning gains. However, this would have been a difficult task with eight instructors teaching a variety of geoscience courses from Physical Geology to Environmental Science, and an instrument too heavily weighted with questions aligned with the InTeGrate Modules would have unfairly tilted results toward the treatment semester.

Having eight participating instructors from institutions in varying locations also presented some challenges and limitations. Having the faculty spread across the country made it difficult to obtain sufficient teaching observations and control for teaching practices. While some observations were made as part of the study (Czajka and McConnell, in review), most of the instructors were only observed one to three times in total. This limited data set made it difficult to identify whether changes could be attributable to differences in teaching practice or materials used. Additionally, we had little control over how the instructors incorporated the GLE into their course.

**Conclusions**

The InTeGrate teaching materials were designed to provide instructors with a suite of customizable modules and units that teach the geosciences in the context of societal issue while employing research-based instructional strategies. Through the project of creating these materials
and making them available to geoscience instructors, it was hoped that they would support the improvement of geoscience literacy among undergraduate students and further broaden participation in the geosciences by attracting new undergraduate majors. This study recruited eight instructors from a variety of institutions to test the idea of whether the InTeGrate teaching materials were an effective way to improve the geoscience literacy of a diversity of students. Despite only two of the instructors seeing significantly higher learning gains on a geoscience literacy exam when using InTeGrate materials, the use of InTeGrate did have a positive effect on some important demographic performance gaps. Female and Hispanics students faced significant performance gaps on gains made on the literacy exam in the control semester. While effect sizes were small, InTeGrate use in the treatment semester did appear to reduce these gaps. The instructors replaced roughly 40% of their course curricula with InTeGrate, but using a greater proportion may have led to larger effects on literacy gains for these populations.

The InTeGrate teaching materials are a valuable resources for geoscience instructors who are interested in modifying how they teach or the context within which they teach about the Earth. Seven complete modules were available for the instructors of this study to use, but there is now an even more diverse suite of over 30 InTeGrate modules and courses freely available for use (https://serc.carleton.edu/61544). These customizable curricular materials incorporate a variety of research-based teaching strategies that can guide instructors in the use of teaching practices that can support and enhance student learning. The content of these modules are embedded within socioscientific issues that can further engage and motivate a diversity of students to become more invested in being geoscience literate and to evaluate their roles as global citizens. There is a value and need to attract a diversity of people and views to the geosciences in light of the geoscience-related grand challenges we face, and teaching our
undergraduate students in a manner promoted by the InTeGrate materials may represent one way of achieving that goal.

**Funding**

This work was supported by the National Science Foundation award 1125331. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors, and do not necessarily reflect those of the National Science Foundation.
References


CHAPTER 3: An Exploratory Study Examining Geology Students’ Conceptions Related to Geologic Time and Rates

Introduction

“So much of accumulating importance at earthly scales – the results of geological erosion, evolutionary changes in lineages – is invisible by the measuring rod of a human life. […] It takes a particular kind of genius or deep understanding to transcend this most pervasive of all conceptual biases and to capture a phenomenon by grasping a proper scale beyond the measuring rods of our own world.”


In his essay “The Golden Rule: A proper scale for our environmental crises”, Gould argues for the importance of considering temporal scale when facing issues of environmental concern and guiding human stewardship of the Earth. Similarly, Cervato and Frodeman (2012) argue the importance of student understanding of geologic time as a guide for decision making about environmental, political, and economic issues related to the sustainability of our planet.

Themes of geologic time and rates permeate the Earth Science Literacy Principles (www.earthliteracy.org), a set of big ideas that experts feel citizens should know to make informed decisions on societal challenges requiring geoscientific understanding and solutions.

The importance of Earth’s past is emphasized in Big Idea 1.5, “Earth scientists use their understanding of the past to forecast Earth’s future,” and Big Idea 3.4 extends this concept, stating “Earth’s systems interact over a wide range of temporal and spatial scales. These scales range from microscopic to global in size and operate over fractions of a second to billions of years. These interactions among Earth’s systems have shaped Earth’s history and will determine Earth’s future.” However, the ability to think across geologic or deep timescales is difficult for many students. Students harbor numerous misconceptions related to geologic time and Earth
historical (Francek, 2013), which, as future decision makers and problem solvers, could present roadblocks in their ability to adequately consider and address issues of societal importance within the framework of Earth’s past.

Research investigating student understanding of geologic time and related misconceptions has been conducted with K-12 students (Dodick & Orion, 2003a, 2003b; Schoon, 1992), pre-service teachers (Gosselin & Macklem-hurst, 2002; Teed & Slattery, 2011), in-service teachers (Dahl, Anderson, & Libarkin, 2005; Trend, 2001), and undergraduate students in introductory courses (Delaughter and Stein, 1998; Libarkin et al., 2007; Zhu et al., 2012). While fewer students struggle with the relative order of events in Earth history (Trend, 2001), many carry inaccurate conceptions related to the timing (absolute age) and temporal duration of events (Libarkin et al., 2007; Catley and Novick, 2009). When students are asked to estimate the temporal magnitude of a scientific phenomenon, they often exhibit the compression effect (Longo & Lourenco, 2007) where very short times (e.g. the blink of an eye) are overestimated and very long times (e.g. mountain formation) are underestimated (Lee, Liu, Price, & Kendall, 2011). The temporal magnitude of extremely small and large events are compressed toward more relatable human timescales. This effect impacts student abilities to estimate rates and temporal durations of geologic processes (Cheek, 2013), and may be partially explained by an inability to work with and relate to large numbers (Cheek, 2012). Additionally, while subject matter knowledge may help students improve on tasks related to temporal scale, further coursework does not necessarily translate into better temporal scale skills (Cheek et al., 2017).

Suggestions for teaching geologic time include ideas such as creating timelines or compressing Earth’s history into a relatable frame of time such as one calendar year (Pyle, 2007; Truscott, Boyle, Burkill, Libarkin, & Lonsdale, 2006). The use of such analogical mapping
methods is a potentially useful tool in teaching geologic time and temporal scale, although the specific techniques that instructors use to employ them and the long term efficacy is unknown (Jee et al., 2010). When using analogies, instructors should clearly demonstrate how the components work and what any limitations of the analogy are in order to avoid student misconceptions and overconfidence (Brown & Salter, 2010; Jaeger & Wiley, 2015). Using analogic mapping is more effective when students progressively map longer timescales onto the same space (e.g. one meter), and hierarchically include the previously used timescale onto the next (Resnick & Shipley, 2012). Allowing students to make predictions about scale alignment or event timing and providing immediate feedback has also be shown to improve learning (Grissom, Czajka, & McConnell, 2015; Resnick, Davatzes, Newcombe, & Shipley, 2017).

Most of the studies above focus on introductory undergraduate students or younger, and report results of one time interventions or measurements. The purpose of this study was to expand on the body of research exploring student conceptions of geologic time, rates, and Earth history. Specifically, we were interested in extending research on these conceptions to undergraduate students who were majoring in geology, and to take measurements at multiple times and at various academic levels. As future geoscientists, these geology students will play a key role in addressing future geo-related global challenges, and should be equipped to view these challenges through the lens of Earth’s past. Undergraduate geology curricula typically introduce students to concepts of geologic time in introductory courses and expand on these themes in additional lower division Historical Geology or Earth history courses. Subsequently, geology majors may encounter themes related to geologic time and rates in upper division courses, especially those with a focus on geomorphology, stratigraphy, or paleontology. This study was guided by the following research questions:
1. How do student conceptions about geologic time, rates, and Earth history evolve during their academic experience as a geology major?

2. How do students think about problems related to geologic time and rates, and what misconceptions do they harbor that impede their ability to solve such problems?

**RESEARCH DESIGN AND METHODS**

*Study Population*

The study was conducted at a large four-year research intensive university in the southeastern US. Data were collected from five different geology courses; introductory Physical Geology, Historical Geology, Structural Geology, Geomorphology, and Geology Field Camp (Table 1). Students in the Physical Geology course were all non-geology students, and mostly non-STEM (Science, Technology, Engineering, and Mathematics) majors. Roughly half the Students in Historical Geology were geology majors or minors, six were other STEM majors, and about a third were non-STEM majors. The upper level Structural, Geomorphology, and Field Camp courses were composed almost entirely of geology majors, most of whom were juniors or seniors. Additionally, data were collected at the 2016 Geological Society of America annual meeting in Denver, Colorado from PhD holding faculty with expertise in a variety of geology fields. These participants represent an expert group in the study population. Data collection began in the spring 2016 semester and continued into the fall of 2017 (Table 1).
Table 1. Participant Demographic Characteristics. Fr = Freshman, So = Sophomore, Ju = Junior, Se = Senior, M = Male, F = Female.

<table>
<thead>
<tr>
<th>Course</th>
<th>Semester</th>
<th>n</th>
<th>Majors</th>
<th>Academic Level</th>
<th>Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Physical Geology</strong></td>
<td>Fall 2016</td>
<td>52</td>
<td>STEM (19), Non-STEM (33)</td>
<td>Fr (20), So (14), Ju (13), Se (5)</td>
<td>M (34) F (18)</td>
</tr>
<tr>
<td><strong>Historical Geology</strong></td>
<td>Spring 2016</td>
<td>48</td>
<td>Geology (19), Geology Minor (6), other STEM (6), Non-STEM (17)</td>
<td>Fr (5), So (7), Ju (18), Se (18)</td>
<td>M (37) F (11)</td>
</tr>
<tr>
<td><strong>Structural Geology</strong></td>
<td>Fall 2016</td>
<td>28</td>
<td>Geology (26), other STEM (1), Science Education (1)</td>
<td>So (1), Ju (4), Se (23)</td>
<td>M (19) F (9)</td>
</tr>
<tr>
<td><strong>Geomorphology</strong></td>
<td>Spring 2017</td>
<td>35</td>
<td>Geology (33), Geology Minor (1), Science Education (1)</td>
<td>So (5), Ju (13), Se (17)</td>
<td>M (24) F (11)</td>
</tr>
<tr>
<td><strong>Geology Field Camp</strong></td>
<td>Summer 2017</td>
<td>28</td>
<td>Geology (28)</td>
<td>Ju (2), Se (26)</td>
<td>M (21) F (7)</td>
</tr>
<tr>
<td><strong>Faculty Experts</strong></td>
<td>N/A</td>
<td>31</td>
<td>Expertise: Geomorphology (6), Sedimentology (4), Hydrology (3), Mineralogy/Petrology (3), Structural Geology (3), Paleontology (2), Economic Geology (2), Misc. (8)</td>
<td>N/A</td>
<td>Not Collected</td>
</tr>
</tbody>
</table>

Study Design

The exploratory study used a mixed-methods, pre-experimental design. While no control group or instructional intervention was used, data collected from introductory, non-majors in a Physical Geology course and from faculty experts provide novice and expert benchmarks to frame the findings from geology students. Quantitative data was collected using a pre-post assessment instrument, and qualitative data came from interviews conducted with a sample of senior geology students from the larger study population. Qualitative interviews were conducted at the end of the study to allow the quantitative findings to inform the focus of the interview questions. The lead author served as the primary investigator (PI) and administered the pre-post assessments in all courses and conducted all student interviews. Additionally, the PI served as
the lead instructor for the Physical and Historical Geology courses and as a teaching assistant for the Field Camp course.

Interviews were conducted with 11 senior geology majors to investigate student conceptions related to some of the more difficult concepts on the pre-post instrument. Participants were recruited via an email sent to all geology students who had completed the pre-post assessment at least once. Students who volunteered to be interviewed were compensated with a $10 gift card. The interviews utilized a semi-structured, think-aloud protocol, and lasted from 15-40 minutes. Each student answered 4-5 questions from the pre-post instrument during the interview. They were presented with a question, given a moment to read it, and then asked to explain their thought process and talk through their answering of the question. Interviews were recorded and transcribed for later data analysis.

Instrument Design

A quantitative instrument to assess student knowledge of geologic time, rates, and Earth history was created using items from other instruments or studies. Seven of twelve items from the Landscape Identification and Formation Test (LIFT, Jolley et al., 2013) were used, representing a range of timespans from short (impact crater) to long (mountains). These items ask students to first identify a landscape or feature and then answer a multiple choice question indicating how long it took to form. Thirteen questions targeting relative and absolute dating, Earth history, and the geologic timescale were adopted from a 20 question instrument developed using instructor feedback and student interviews (Rhajiak, 2009). Questions used in this instrument came from both the Geoscience Concept Inventory (Libarkin and Anderson, 2006) and from instructor-developed concept test questions available online (McConnell, Steer, Owens,
Finally, a constructed response question asking students to place tick marks representing events on a scaled line of Earth’s history was also added (see Figure 3, Grissom et al., 2015). The final instrument consisted of 21 items (Appendix B), with questions 1-7 being two part items (landscape Identification and formation rate), and was scored out of 34 total points. Table 2 shows the breakdown of the four subscales of the instrument and their respective point values. The instrument also collected demographic information from participants including gender, class rank, major, and geology courses taken.

Table 2. Subscales and concepts present on the pre-post assessment instrument.

<table>
<thead>
<tr>
<th>Subscale</th>
<th>Concepts Covered</th>
<th>Number of Questions</th>
<th>Points</th>
</tr>
</thead>
<tbody>
<tr>
<td>Landscape Identification</td>
<td>Alluvial Fan, Lava Flow, Impact Crater, Mountains, Volcano, Landslide, U-Shaped Valley</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Landscape Formation Rates</td>
<td>Formation rates of the above features</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Geologic Time Concepts</td>
<td>Earth History (3), Relative (3) &amp; Absolute Dating (4), Geologic Timescale (3)</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>Event Timeline</td>
<td>Origin of life, hard parts, origin of dinosaurs, extinction of the dinosaurs, and origin of humans</td>
<td>1</td>
<td>7</td>
</tr>
</tbody>
</table>

The items that were used to construct the assessment came from instruments that had used appropriate measures to establish evidence for the content validity of items (Libarkin and Anderson, 2006; Rhajiak, 2009; Jolley et al., 2013). The predictive nature of scores, with novice, introductory geology students scoring the lowest, faculty experts the highest, and geology students in between, provides general evidence of criterion validity for the assessment. Classical Test Theory, was used to establish measures of reliability and discriminatory power of the instrument from the post-test scores of students in Historical Geology, a course where the instrument is most suited. Values for three item-analysis measures and two test-analysis
measures are provided in Table 3, along with recommended values from Ding and Beichner (2009). The difficulty index is a measure of the proportion of students who answered each question correctly. The discrimination index is a measure of how well a question distinguishes between high performing and low performing students by comparing the number of correct respondents among performers in the top and bottom quartiles. The point biserial provides an item reliability by correlating individual item scores to total test scores. For test-analysis, a Cronbach’s alpha coefficient of reliability of 0.673 was obtained. While this value is slightly lower than the ≥0.70 recommended by Ding and Beichner (2009), Adams and Wieman (2011) caution that a high internal consistency is not often desired nor is it the best indicator of reliability for an assessment that is measuring more than one construct. Finally, a Ferguson’s delta was calculated to assess the discriminatory power of the entire instrument, and provides a measure of the distribution of scores.

Table 3. Test statistics calculated from the post-test administration in Historical Geology. Sample size n = 52 students. The desired values are those suggested by Ding and Beichner (2009).

<table>
<thead>
<tr>
<th>Test Statistics</th>
<th>Instrument Values</th>
<th>Desired Values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Difficulty Index</td>
<td>Average of 0.63</td>
<td>0.30-0.90</td>
</tr>
<tr>
<td>Discrimination Index</td>
<td>Average of 0.33</td>
<td>≥ 0.30</td>
</tr>
<tr>
<td>Point Biserial Coefficient</td>
<td>Average of 0.32</td>
<td>≥ 0.20</td>
</tr>
<tr>
<td>Reliability Index (Cronbach’s alpha)</td>
<td>0.673</td>
<td>≥ 0.70</td>
</tr>
<tr>
<td>Ferguson’s delta</td>
<td>0.873</td>
<td>≥ 0.90</td>
</tr>
</tbody>
</table>

The 13 geologic time questions and event timeline are concepts that are central to the Historical Geology course and were explicitly taught, while the concepts of landscape identification and formation rates were not directly discussed in the course. A review of course syllabi revealed that the concepts of landscape identification and formation rates were more directly covered in Geomorphology, especially during lessons on glacial geomorphology,
tectonic geomorphology & landscape evolution. While the Field Camp course was largely focused on geologic mapping, there was a four day stratigraphy project where students measured 300 meters of stratigraphic section, described the sedimentary facies, and created a stratigraphic column. Some of the skills utilized and learned during this activity align loosely with concepts assessed, such as questions related to relative dating and stratigraphic principles. Structural Geology was the course where students were least likely to have directly encountered the topics being assessed, as the course focused on concepts related to strain, stress, rheology, fractures, folds, fabrics, and shear zones. However, most of the students were taking multiple geology courses concurrently. During the fall of 2016, available geology courses included Structural Geology, Marine Sediment Transport, and Mineralogy and Petrology. Spring of 2017 offerings included Historical Geology, Sedimentary Petrology and Stratigraphy, Igneous and Metamorphic Petrology, and Geomorphology. In Sedimentary Petrology and Stratigraphy, students would also have encountered some of the concepts in Table 2, including landforms such as alluvial fans and references to relative dating and the geologic timescale.

**Data Analysis**

All quantitative data analysis was conducted using IBM Statistical Package for the Social Sciences (SPSS) 25. Only data from students with paired pre- and post-test results were used in the analysis. Qualitative data analysis of interviews was conducted using a thematic analysis process similar to that described by Braun and Clarke (2006). The process begins with multiple readings of the interviews to become familiar with the data, followed by the generation of initial codes. From these codes, themes are then identified, reviewed and defined. Think aloud
responses for each individual question were analyzed separately to identify alternative conceptions and themes relevant to the specific concept.

Results

Quantitative Results

Paired sample t-tests were run to determine if statistically significant gains were made from pre- to post-test for each course. Significant gains with a large effect size were seen in both Physical Geology \( (t(51)=10.397, p<0.0005, d=1.44) \) and Historical Geology \( (t(47)=8.83, p<0.0005, d=1.27) \). Students in the other three courses made no significant gains from pre- to post-test (Figure 1). Results for each of the individual subscales were also analyzed for each course using paired samples t-tests (Figure 1). The estimation of landscape formation rates was the only subscale that showed no significant improvement from pre- to post-test for students in both Physical Geology \( (t(51)=1.686, p=0.098) \) and Historical Geology \( (t(47)=1.517, p=0.136) \). However, there was a significant difference in formation rate pre-scores between students in Physical Geology and Historical Geology \( (t(98)=2.957, p=0.004) \).
Figure 1. Pre-Post test results from all courses and faculty experts who were only tested once. The score form each subscale of the assessment is shown within each bar. * = significant gains from pre- to post- on the subscale (p<0.05). ‡ = significant gains from pre- to post- on the entire assessment (p<0.0005). Subscales are scored out of 7 points, except for the Geo Time Concepts which is scored out of 13 points.

The final attempts on the test for all 62 geology students in the study were compiled to create a group of geology majors for comparison to introductory physical geology students (at post-test) and faculty experts. The geology majors were mostly seniors (n=40) and juniors (n=15) and more than two-thirds were male (n=43). A one-way ANOVA was conducted to determine if scores differed between physical geology students, geology majors, faculty experts. Scores were normally distributed for each group as assessed by Shapiro-Wilk test (p>0.05) and there was homogeneity of variances as assessed by Levene’s test (p=0.110). Scores were found to be significantly different between the groups, F(2,142) = 50.339, p<0.001 (Figure 2). A Tukey post hoc analysis revealed that all three groups were significantly different from each other (p<0.001). This procedure was repeated for all four subscales of the instrument across the three groups, and all subscales were found to be significantly different between the groups. Tukey post...
Hoc analyses revealed that the only non-significant difference was between Physical Geology students and geology majors on the event timeline (p=0.606).

Figure 2. Histograms showing the distribution of scores for Physical Geology students, geology majors, and faculty experts. Dashed lines represent the means of each group, and the differences between all means were significant (p<0.001).

Difficulty indices (percent who answered correctly) were calculated for each question to identify the specific conceptual areas with the greatest differences between geology students and faculty expert responses (Figure 3). The areas with the largest gap between faculty experts and geology students were the geologic time concepts related to absolute dating and the geologic timescale. The Earth History questions and landscape formation rates were the most difficult for both the geology students and the faculty experts. Students out performed experts on only one subset of questions, relative dating.
Within the study, there was a cohort group of four students who were pre-post tested all four semesters during Historical, Structure, Geomorphology, and Field Camp. Additionally, seven students were pre-post tested in three of the four semesters, and 25 students were assessed at least twice. Analyses were done to determine if scores or differences in the courses were being driven by a testing effect (practice effect) due to retesting using the same instrument. Half the students in the Structural Geology course (n=14) had taken the assessment in a prior semester as students of the Historical Geology course. There was no significant difference in pre-test scores between students who had already taken the assessment twice in Historical and those who were taking it for the first time in Structure ($t(26) = -0.318, \ p=0.753$). There was also no significant
difference in the number of geology classes taken by the two groups (t(26)=1.319, p=0.199).

**Qualitative Interview Results**

Due to the low performance of students on the formation rate questions, a sample of these items were utilized in the student think aloud interviews. These included the alluvial fan, and U-shaped valley. Additionally, the event timeline constructed response question was also used for think aloud interviews.

**Alluvial Fan**

The alluvial fan question used an image of a large fan in the Badwater Basin of Death Valley National Park. Estimating the formation rate of the alluvial fan was the most difficult question for geology students, with only 8% getting it correct, even though 79% could correctly identify the landform. Ten students were asked this question during a think aloud interview, with three giving a correct response. Coding of responses resulted in three major themes, and a variety of sub-themes (Table 4). When thinking about the formation rate of the alluvial fan, six students talked about the theme of catastrophism. Some would compare the alluvial fan to a landslide, “that reminds me of like a landslide in a way,” but knew it was not, “it doesn’t look like a slope failure.” For other students, they prescribed a very catastrophic process to the formation rate of the fan: “It’s either days or less, or years. It just seems like an alluvial fan to me. It’s kind of like a blowout, almost. It all gets channeled into those – a channel, I guess, and just blows out from the bottom.” Other students create a dichotomy, where the fan could either be the result of gradual or catastrophic processes. “It had to have taken at least a long time or a serious water-draining event.” The previously quoted student did estimate the formation rate correctly as 10k
to 100k years, and the only other student to answer correctly did not mention any catastrophic processes at all.

**Table 4.** Themes identified from each question during think aloud interviews and the number of students who were coded for each particular sub-theme.

<table>
<thead>
<tr>
<th>Question</th>
<th>Themes</th>
<th>Sub Themes</th>
<th># of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Alluvial Fan Formation Rate</strong></td>
<td>1 – Catastrophism</td>
<td>a) explains feature</td>
<td>2 of 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) feature is similar to a landslide</td>
<td>2 of 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c) either catastrophic or gradual</td>
<td>2 of 10</td>
</tr>
<tr>
<td></td>
<td>2 – Use of Road</td>
<td>a) as scale for size</td>
<td>1 of 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) to base or revise Identification</td>
<td>3 of 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>c) as age or activity indicator</td>
<td>4 of 10</td>
</tr>
<tr>
<td></td>
<td>3 - Size</td>
<td>a) to accurately estimate rate</td>
<td>3 of 10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) revise rate estimate to be longer</td>
<td>2 of 10</td>
</tr>
<tr>
<td><strong>U-Shaped Valley Formation Rate</strong></td>
<td>2 – Periodicity</td>
<td>a) Accurate conceptions of glacial periods</td>
<td>3 of 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) Confusion of ice ages and glacial periods</td>
<td>4 of 9</td>
</tr>
<tr>
<td></td>
<td>1 – Retreat</td>
<td>a) Part of glacial cycles</td>
<td>2 of 9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) As a carving mechanism</td>
<td>3 of 9</td>
</tr>
<tr>
<td><strong>Event Timeline Question</strong></td>
<td>1 – Numeric Dates</td>
<td>a) accurate conceptions and placement</td>
<td>7 of 11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>b) temporal errors without a date</td>
<td>9 of 11</td>
</tr>
<tr>
<td></td>
<td>2 – Analogic Thinking</td>
<td>a) for placing the of origin of humans</td>
<td>4 of 11</td>
</tr>
</tbody>
</table>

The other two themes that students talk about when estimating the formation rate of the alluvial fan are the presence of a road in the picture and the size of the fan as an indicator of time. For some, the size allowed them to accurately estimate the formation time of the fan, “I would guess more in the 10k to 100k years because that’s a very large feature.” The others used the size as justification to revise an initial estimate, but still underestimated the true time: “I initially said years, but I’m saying 100 to 1,000 years. I’m looking at the relief on it…there’s a staggering difference in height here from where it’s actually fed in this channel down to where the water level is.” Finally, most students also talked about the presence of a road in the photo, with one using it as reference for size. Some used it as a basis for their identification of the feature or to revise their identification. “That’s not a delta because that’s a road cutting through
it,” or “I can’t imagine the DOT would put something on something as unconsolidated as a
delta.” Others used the road to revise an estimation of age “Well, is that a road? Oh, then maybe
it’s longer than years, huh?” or to infer that the feature is not the result of continual processes
that are still occurring today, but was a past event that happened rapidly: “that looks like a road
to me. So it happened before that road was built…and it hasn’t destroyed it since.” The same
student talking about the rate said, “It’s either days or less, or years. It’s kind of like a
blowout…a rainy season or something like that where it just really dumps out.”

U-shaped Valley

Estimating the formation rate of the U-shaped valley was the second most difficult rate
question for geology students, with only 47% answering correctly. Nine students were asked this
question during interviews, with five giving a correct response. The coding revealed two main
themes in student thought, classified as periodicity and retreat. When thinking about the
formation of a U-shaped valley, students often talked about the idea of periodicity. Some
students had accurate conceptions of glacial and interglacial periods, and used this to infer the
correct 10k to 100k estimate of formation rate for the U-shaped valley. Other students confused
the concept of glacial periods with ice ages, resulting in an overestimation of the time for glacial
valleys to form: “I’m going to go with the amount of time between ice ages I feel was in the
millions of years.” The other common theme discussed by students was that of glacial retreat.
While two students incorporated retreat into their accurate conceptions of glacial periods, four
students described retreat as the mechanism responsible for glacial carving: “the snow and stuff
makes me want to say that this is a glacier-related thing, and the ice is breaking down the rock as
it is retreating.” The misconception that glacial retreat carves the valley would lead students to
underestimate the formation rate based on the idea that extreme temperatures could melt the

glacier quickly: “So I know glaciers can move within a time frame of hundreds of years or less

and pretty much be non-existent if the temperature is provided.”

Event Timeline

On the event timeline question, geology students performed about 12% below faculty

experts (Figure 3). This question allows students to apply temporal scaling skills by plotting five

events in Earth’s history on a 4.5” line. From student think aloud interviews, two primary themes

arose. The first was the importance of dates for students in their plotting of events on the
timeline. Students who had an accurate numeric date for an event were able to plot it accurately,

and the only student to plot all five events accurately during the interviews knew a date for all

four of the non-human origin events (figure 4a, correct timeline). In contrast, students who
cannot identify a date have trouble accurately plotting the events, with the most troublesome

being the origin of hard parts and the origin of dinosaurs. Figure 4b (inaccurate Timeline) shows

a student answer who only had a date for the origin of life. Most students failed to relate the

origin of hard parts to the Cambrian Explosion, and had trouble recalling the date signaling the

start of the Cambrian. This led to a variety of temporal scaling errors, such as “I know there was

a big difference between hard parts and dinosaurs appearing.” While there was around 300

million years separating these events, this only represents 5/16 of an inch on the timeline. This

student plotted them with a 1.25 inch gap, representing 1.25 billion years of separation. Another

student said “I feel there was a really, really large gap in between origin of life and first

organisms with hard parts.” While this statement accurately reflects the 3+ billion year gap, the

student plotted the events with a 1 inch gap representing only one billion years of separation.
Figure 4. Example timelines from student interviews, a) represents a student who used specific dates to create an accurate timeline and b) is a student with inaccurate conceptions about the timing of events.

The other theme that four students mentioned when answering this question was the use of analogy. Major events in Earth history are often taught using some form of analogy (e.g. Earth’s History in a calendar year). However, these analogies were generally only useful in aiding students in plotting the origin of *Homo sapiens*. Four students mentioned specific analogic examples when talking through the plotting of the origin of humans, including one encountered in popular science: “I know that, thank you Neil Degrasse Tyson, which if you were to put this in a calendar year, we appeared in the last second.” And while one student mentioned an analogic activity done in a Physical geology lab, “I know in the geology lab, we did that whole thing with the tape on the floor…I do know all of the stuff was kind of scrunched up towards the end,” they still had the origin of hard parts plotted 1.9 billion years from the present. Even though no student produced an accurate date for the origin of humans during the interviews, they all
recognized that we are a very recent geologic phenomenon and plotted the event accurately, even if they didn’t have an analogic reference.

**Discussion**

Student conceptions about geologic time, rates and Earth History show significant growth in introductory courses and then remain stable through the remainder of their academic experience as a geology major. Students made statistically significant gains from pre- to post-test during Physical Geology and Historical Geology courses. This result is not unexpected, as students are first introduced to many of the concepts covered on the assessment during Physical Geology and many of these concepts are the primary focus of Historical Geology. While students do not make significant gains from pre- to post-test on these concepts after taking upper level geology courses, they do maintain their level of knowledge as they proceed through the curriculum.

It is unlikely that student results were influenced by a testing (practice) effect from repeated exposure to the test. Students were never given feedback or scores after any administration of the test and were therefore unable to use this knowledge to inform subsequent testing. Testing in Structural Geology, where half the students had previously taken the assessment, revealed no difference in score between students who had taken the test previously and those who had not. Finally, with many students taking the assessment multiple times as the study progressed, a significant practice effect would likely result in score improvements over time. This was not the case, and significant improvements were not seen in upper level courses taken after Historical Geology or in students who were tested multiple times.
For students in Physical Geology and Historical Geology, significant gains were made on each subscale of the test with the exception of estimating landscape formation rates. However, we do see a significant difference on this scale between pre-test scores of students in Historical Geology and Physical Geology courses. It is possible that greater subject matter knowledge among the Historical Geology students is contributing to better rate estimation abilities. During Historical Geology and upper level geology courses, average student abilities do not improve on this task. Our findings on students’ abilities are also consistent with those of Jolley et al. (2013). Of the seven landscape features we utilized from the LIFT, the percentage of geology students who estimated the formation rates accurately in our administrations fell within the standard error of the percentages reported by Jolley et al. (2013). A larger proportion of geology students (15-25%) in our population were accurate in their rate estimations for the volcano and landslide items on the test.

It appears that there are specific misconceptions about geologic rates that may impede students’ abilities to continue to improve their scores. Our results confirm that students find estimation of intermediate time spans (i.e. alluvial fan and U-shaped valley) the most difficult. Jolley et al. (2013) proposed that students may be lumping events into short, intermediate, and long categories, and that instruction on intermediate time spans and events may have been lacking. It may also be that intermediate time spans, by nature, are more difficult for students to estimate due to the available range of error. If a landscape is identified as being formed rapidly, the probability of underestimation is low because of a basement effect. A similar ceiling effect would exist for long time spans and overestimation. However, with intermediate time spans, there is much more room for both over and underestimation errors. The intermediate time span features included in the LIFT (i.e. alluvial fan, hoodoos, river, U-shaped valley) are all in some way
related to depositional or erosional processes. It may be that the rates of these processes are less prevalent in instruction than those of short timespan process like faulting or long timespan processes such as mountain building.

Interviews provided further insight into why students may be inaccurate in their intermediate timespan estimations. With the alluvial fan, 60% of the students interviewed made some reference to catastrophism when discussing the timespan of its formation. Even if they knew it was not a landslide or result of a rapid event, the association to such an event may be causing them to underestimate its timespan. The presence of the road in the photograph also led students to inaccurate conclusions about the timespan. For example, the student quoted above who inferred that since a road was built at the base, it must have been a rapid depositional event in the past and that it was no longer depositionally active. Or another student who voiced similar thoughts, “It’s like a landslide and could happen in days or less…this kind of roadway or something makes me think that it hasn’t formed over when it originally formed,” implying there is no further threat of deposition that would block the roadway. In these cases, association with human created structures may be causing underestimation of timespans.

Students expressed two ideas that led them to inaccurate over- and underestimations when talking about U-shaped valley timespans. Two students conflated the idea of ice ages and glacial periods, applying the cyclic nature of glacial periods to the longer timespans of ice ages. This caused them to overestimate the formation rate of the U-shaped valley. Some students also expressed the misconception that glaciers carved the valley during retreat. This led them to think that extreme temperature changes leading to rapid glacial retreat would quickly carve a U-shaped valley, and they underestimated the timespan. This is a misconception about glacier erosion that is not reported in Francek’s (2013) compilation of over 500 geologic misconceptions, or in a
recent paper on student conceptions of glaciers and ice ages (Felzmann, 2017). Francek (2013) expressed surprise that more research has not been done in the area of glacial misconceptions considering their relevance to current climate issues, and notes the absence of peer reviewed research on the subject.

Many geology students also struggled with familiarity with the geologic time scale (Figure 3). Based on results from the assessment and from interviews, many students are unfamiliar with the names and relative ordering of time periods in the geologic timescale. Additionally, they don’t know some of the more prominent dates associated with the time scale, such as the start of the Cambrian or end of the Permian. This unfamiliarity with the geologic timescale also likely contributed to lower performance in some of the other topic areas as well such as the Earth History questions (e.g. identifying how long dinosaurs existed) and the event timeline question. Of the students interviewed on the event timeline question, those that knew the geologic time scale and associated dates performed much better. The only student to get all seven points during the interview responded to the probe of “Is this topic something you learned in a specific course?” with “Historical Geology, and honestly the repeated quizzes on the time scale.”

During interviews, some students mentioned learning the geologic timescale in Historical, but that in other classes it was only mentioned in passing to support other concepts. As one student said, “Outside of Historical, the timescale was mentioned, but it wasn’t ever reinforced. The assumption was that you totally grasped it by the end of that class, and it was just ingrained.”

The only area where students performed better than experts was on the topic of relative dating. This stems from the question asking which of three fossils would be the best choice to use as an index fossil based on four provided stratigraphic columns. Experts only answered this question correctly 61% of the time versus 80% for geology students. Based on the areas of
faculty expertise (Table 1), the surveyed faculty may not be in fields or teach courses that deal with index fossils. However, five of the six experts in fields where index fossils would likely be a part of their expertise (sedimentology/paleontology), did not supply the intended correct answer. It is therefore possible that this question lacks validity by not accurately targeting the concept of index fossils, and would benefit from further expert review.

Students performed equally with experts on the other two relative dating questions concerning stratigraphic principles (97% vs. 93%) and unconformity definition (65% vs. 66%).

An interesting phenomenon was encountered in the Physical and Structural Geology courses during pre-post testing. In Physical Geology, 88% of students were able to identify an impact crater on the pre-test, but only 73% were correct on the post-test. After having learned about volcanoes and calderas in the course, many students now chose to interpret the impact crater as a caldera. Their accuracy in estimation of the formation rate also dropped from 80% to 69% pre- to post- as a result of students now overestimating the formation time due to thinking the feature was a caldera. Similarly, in Structural Geology, the number of students who were able to correctly identify the lava flow dropped from 75% to 64% from pre- to post- test. On the post-test, a small handful of students were now identifying the feature as something related to structural geology, including joints, fractures, or an extensional landscape. In these instances, students may be falsely assuming that they are being assessed on concepts directly taught in a course, and overriding their correct prior knowledge to fit this assumption. Instructors utilizing pre-post testing to measure student learning in a course, especially when using instruments that may not fully align with course content, should be aware of this phenomenon.
Implication and Future Work

Our results demonstrate that students have a difficult time estimating formation rates of features that are formed on intermediate timespans, similar to Jolley et al. (2013). These similar findings indicate that this may be a skill that many geology students struggle with, and is not unique to a specific population. Intermediate timescale features are generally the result of erosional or depositional processes which may be discussed in some upper level geology courses but apparently have no effect on student perceptions of these features. Future work should investigate instructional interventions that could improve student abilities with intermediate timespans. Improvement here would also likely lead to better estimation abilities at short and long timespans. With the exception of one unpublished, but novel study into the use of time-lapse videos to teach rates (Schierl, 2014), we are unaware of any other studies investigating methods for teaching geologic rates of surface processes. Although Schierl did not utilize the LIFT, he did find that the use of before/after images and time-lapse videos were both effective at improving student abilities to estimate rates. Further testing of these strategies, as well as others that may involve the use of analog or digital models of surface process, would help identify effective strategies for instructors of both introductory and upper level courses.

Investigating the effect that knowledge of the geologic timescale has on student comprehension of concepts related to geologic time and rates would also be a potential avenue of further research. Many students in our study had poor knowledge of divisions in the geologic timescale, including names and major dates. For students solving the event timeline question during interviews, knowledge of the geologic timescale was more heavily relied upon than analogic thinking. Jee et al. (2010) wondered “As a student gains expertise in a topic, do the initially useful analogies remain stored in memory? Are they changed, discarded, or integrated
“with additional knowledge?” From our interview data, it seems that only the most impactful part of the analogy concerning the appearance of humans remains stored in memory, and the most proficient students are using divisions and dates from the geologic timescale to solve these problems. Further interviews that incorporate both introductory students and faculty experts would help illuminate how thinking varies across ranges of expertise, not only regarding the geologic timescale and events in Earth’s history, but especially with how experts deal with estimation across geologic timespans.

A difficulty faced by instructors of upper division courses is the diversity of subject knowledge that students bring to their courses. Not all students take their required coursework in the same order, which impacts the amount of subject matter and prerequisite knowledge they bring to a specific class. Upper division courses often rely on prerequisite knowledge with the assumptions that students are proficient in these areas. An example from our institution highlights how these assumptions can be problematic. The Historical Geology and Geomorphology courses are both on the class schedule for spring semester. Some students, especially those transferring from other institutions or degree programs, may take the two courses concurrently. Consequently, they may lack knowledge of content typically learned in Historical that would be beneficial in the Geomorphology course. Our results suggest that it would be worthwhile for instructors to reinforce needed concepts, such as the geologic timescale, or to assess student prior knowledge at the beginning of the course to know where reinforcement is needed. This strategy may help to enhance learning in higher level courses of more complex topics, and also help students to become more expert-like in their performance through repeated practice and reinforcement of learned concepts.
Limitations

The reliability and discrimination values of the constructed instrument were generally acceptable even though the two test-analysis values (alpha and delta) are slightly below those desired (Table 3). With a piloting of the instrument before beginning the research project, item analysis could have been conducted on weaker items to improve the overall reliability and discrimination of the instrument. However, the values obtained for both alpha and delta are close enough to the desired values that we remain confident in our interpretation of the study results.

While it is thought that content alignment between the assessment and Physical and Historical Geology courses is what led to the significant gains seen, there was no control for or measure of instructional style in the courses, a factor which could influence student learning gains. The PI, who also served as the instructor for the Physical and Historical Geology courses, is a geoscience education researcher, and may have used a more student-centered pedagogy than the instructors of other courses. There are also a variety of student variables which could have affected the findings. We were not able to control for the order in which students took geology courses. While students reported on which geology courses they had taken as part of the assessment, the order that courses were taken and where they were taken (for transfer students) varied. Furthermore, each course where the assessment was administered had a different population of students, each with a varied background of courses taken. Only four students made up the cohort that were tested in all four semesters. Having a larger cohort group would have made it easier to control for factors such as student abilities and backgrounds. Finally, caution should be used in broadly applying the findings of this study to other institutions. Institutions have differences in program level objectives, student populations, courses offered, and even differences in the content covered in courses similar to those in our study.
Conclusions

This study aimed to explore how geology students’ conceptions about geologic time and rates evolve during their course of study. An instrument was compiled that assessed students on landscape identification, landscape formation rates, and various geologic time concepts. The assessment was administered over the course of four semesters in various geoscience courses including Physical Geology, Historical Geology, Structural Geology, Geomorphology, and Field Camp. It was found that students made significant learning gains during Physical and Historical Geology, and maintain this level of knowledge through upper level geology courses, never reaching an expert level. The ability to estimate rates of landscape formation was the most troublesome challenge for students at all levels, especially for landscapes formed on intermediate timespans (e.g. alluvial fans, U-shaped valleys). Interviews with eleven senior geology students reveal misconceptions and thought processes that contribute to inaccuracies in timespan estimates.

Acknowledgements

We would like to thank all of the students who participated in this study, especially those who agreed to be interviewed. We would also like to thank the instructors of the geology courses who let us use their class time for pre-post testing of students. Finally, we would like to thank the faculty at the 2016 Geological Society of America annual meeting who provided valuable expert data by taking our assessment, as well as Julie Libarkin for allowing us to use her booth space at the meeting and administer the assessment.
References


University of British Columbia (Vancouver).


APPENDICES
**Appendix A: Sixteen Question Geoscience Literacy Exam (GLE)**

1. Natural hazards can be put in two major categories. Some natural hazards can be made worse by humans; others are largely independent of human activities. Select the natural hazard least likely to be affected by human activity.
   a. Forest fires
   b. Tsunami
   c. Landslides
   d. Coastal erosion

2. Which of the following geologic processes are mostly likely caused by the interactions between tectonic plates at their boundaries? **Select all that apply.**
   a. Earthquakes
   b. Continental Glaciation
   c. Floods
   d. Volcanic eruptions
   e. Mountains

3. Which of the following statements about the distribution of life in the oceans is most correct?
   a. Life is more abundant and diverse in some parts of the ocean than in others.
   b. Life is abundant and diverse throughout the ocean.
   c. Life is less abundant and diverse in the oceans than it is on land.

4. Which of the following ways do humans affect oceans? **Select all that apply.**
   a. Humans alter ocean ecosystems through fishing.
   b. Humans alter shorelines through development
   c. Humans mine mid-ocean ridges.
   d. Humans change overall ocean composition by desalination.
   e. Humans alter tidal cycles.

5. Which of the following processes primarily involves the atmosphere and the biosphere?
   a. The formation of limestone
   b. The photosynthetic cycle
   c. The hydrological cycle

6. Which of the following processes are sources of carbon to the atmosphere? **Select all that apply.**
   a. Plant decay
   b. Limestone formation
   c. Cattle ranching
   d. Fossil fuel use
7. There are several climate models used to research future change. Which climate modeling statement about 21st Century temperature change projections is most accurate?
   a. Climate model projections do not agree on future likely outcomes.
   b. **Climate model projections show similar trends for future outcomes.**
   c. Climate model projections show the same results for future outcomes.

8. The first reasonably accurate mercury thermometers were invented in 1724, almost 300 years ago. What kinds of processes and/or data are used by scientists to determine temperatures **more than** 10,000 years in the past? **Select all that apply.**
   a. Written records
   b. **Ice cores**
   c. Tree rings
   d. **Sedimentary layers**
   e. **Oxygen isotopes**

9. Which of the following is a good example of interaction between the geosphere and the hydrosphere?
   a. Water evaporates from the ocean into the atmosphere.
   b. A glacier carves a valley in the mountains.
   c. Plant roots break apart an outcrop of granite

10. Which of the following best describes what scientists mean when they use the word “earthquake”?
    a. When an earthquake occurs, visible cracks appear on the Earth's surface
    b. When an earthquake occurs, people can feel the Earth shake
    c. When an earthquake occurs, man-made structures are damaged
    d. **When an earthquake occurs, energy is released from inside the Earth**
    e. When an earthquake occurs, the gravitational pull of the Earth increases

11. Which of the following are considered common mechanism for weathering and erosion? **Select all that apply.**
    a. Wind
    b. Rain
    c. Earthquakes
    d. Volcanoes
    e. Rivers
12. The following maps show the position of the Earth’s continents and oceans. The dots on each map mark the locations where volcanic eruptions occur on land. Which map do you think most closely represents the places where these volcanoes are typically observed?

A. Mostly along the margins of the Pacific and Atlantic Oceans
B. Mostly along the margins of the Pacific Ocean
C. Mostly in warm climates
D. Mostly on continents
E. Mostly on islands

13. Which of the following alternative energy sources is renewable and does not release large amounts of carbon dioxide into the atmosphere? Select all that apply.
   a. Coal
   b. Geothermal
   c. Solar
   d. Natural Gas
14. Scientists often talk about the Earth’s tectonic plates and their role in mountain formation, volcanism, and earthquake occurrence. Which of the following figures most closely represents the location of the Earth’s tectonic plates?

a. A  
b. B  
c. C  
d. D

15. Select the factor that plays the most important role in determining where mineral resources such as oil and natural gas are discovered today.

a. Latitude  
b. Climate  
c. Population centers  
d. Ancient ecosystems

16. What are the likely long-term impacts of 21st Century climate change on the global availability of fresh water? Select all that apply.

a. More fresh water available in coastal areas as sea level rises  
b. Less fresh water available from mountain snow/glacier melts  
c. More fresh water in some locations as precipitation patterns change.  
d. Less fresh water in some locations as precipitation patterns change.
Appendix B: Geologic Time, Rates, and Earth History Assessment Instrument

Name: _______________________

Geology Pre-Test: The following pre-test will **not** be scored for a grade. You are not expected to know the answers to all of the questions, but please try your best and do not copy others’ work. The purpose of this test is to assess your level of knowledge on topics related to geology time and Earth history prior to this course.

**Part 1:** You will have 45 seconds to look at each photograph (on the projected slides) and write down what type of landscape/feature is shown in the image, as well as how long the depicted landscape/feature took to form (**NOT** how old it is). If you are unsure of the type of landscape, try to hypothesize the process or processes that created it. Please fill out your written answer for A) on the test booklet, and bubble in your response to B) on the scantron next to the corresponding number (i.e. for 1 B) bubble in your choice next to 1. On the scantron)

1. A) What type of feature is this? ____________________________

   B) How long did this feature take to form? Choose the **best** answer.
      a) Days or less
      b) Years
      c) 100 to 1,000 years
      d) 10,000 to 100,000 years
      e) 100,000 to 1,000,000 years

2. A) What type of landscape is this? ____________________________

   B) How long did this landscape take to form? Choose the **best** answer.
      a) Minutes or less
      b) Days
      c) Years
      d) 100 to 1,000 years
      e) 10,000 to 100,000 years

3. A) What type of feature is this? ____________________________

   B) How long did this feature take to form? Choose the **best** answer.
      a) Minutes or less
      b) 100 to 1,000 years
      c) 10,000 to 100,000 years
      d) 100,000 to 1,000,000 years
      e) More than 1,000,000 years
4. A) What type of landscape is this? ________________________________

B) How long did this landscape take to form? Choose the best answer.
   a) 1000s of years or less
   b) 10,000 to 100,000 years
   c) 100,000, to 1,000,000 years
   d) 10,000,000 to 100,000,000 years
   e) More than 1,000,000,000 years

5. A) What type of feature is this? ________________________________

B) How long did this feature take to form? Choose the best answer.
   a) Years or less
   b) 1,000 to 10,000 years
   c) 100,000 to 1,000,000 years
   d) 10,000,000 to 100,000,000 years
   e) More than 1,000,000,000 years

6. A) What type of feature is this? ________________________________

B) How long did this feature take to form? Choose the best answer.
   a) Minutes or less
   b) Days
   c) Months
   d) Years
   e) 100s of years or more

7. A) What type of feature is this? ________________________________

B) How long did this feature take to form? Choose the best answer.
   a) 100s of years or less
   b) 10,000 to 100,000 years
   c) 100,000 to 1,000,000 years
   d) 1,000,000 to 10,000,000 years
   e) More than 10,000,000 years
**Part 2:** You will have 15 minutes to complete this section. Please answer the questions below to the best of your abilities. Bubble in your answers for 8-20 on the scantron. Complete question 22 in the test booklet.

8. Approximately how many years back in time did the Earth form?
   a) 4 hundred years
   b) 4 hundred-thousand years
   c) 4 million years
   d) 4 billion years
   e) 4 trillion years

9. What is the **best** estimate of the age of F if A is 100 million years old and D is 70 million years old?
   a) 55 million years
   b) 85 million years
   c) 110 million years
   d) Same age as A
   e) Same age as D

10. Which method was **primarily** used to establish the Geologic Time Scale?
    a) Correlation of magnetic signatures in rocks
    b) Calculation of alpha decay of isotopes
    c) Calculation of beta decay of isotopes
    d) Correlation of rock types across vast distances
    e) Correlation of fossils in rock units across vast distances

11. An index fossil is a fossil that dates the strata in which it is found. Four outcrops of rock are examined in different locations. The rock types and the fossils they contain are illustrated in the diagram below. Which fossil would be the **best** choice to use as an index fossil for these rocks?
    a) Fossil 1
    b) Fossil 2
    c) Fossil 3
    d) There are no index fossils
    e) All fossils make equally good index fossils
12. The isotope Einsteinium-253 has a half-life of 20 days. If you began an experiment with an 80-gram sample of Einsteinium-253, how much would remain after 60 days?
   a) 10 grams  
   b) 20 grams  
   c) 40 grams  
   d) 60 grams  
   e) Not enough information provided

13. What do we call the feature left by a cycle involving deposition, then removal of previously deposited sediment by erosion, then a return to deposition?
   a) A cross-cutting relationship  
   b) An inclusion  
   c) A nonconformity  
   d) A turbidite sequence  
   e) An unconformity

14. In the figure below, Layer A is from the Ordovician, Layer C is from the Carboniferous. What geological time Period is most likely represented by Layer B?
   a) Lower Carboniferous  
   b) Upper Ordovician  
   c) Devonian  
   d) Permian  
   e) Paleozoic

15. How much time passed between the appearance and extinction of the dinosaurs?
   a) 35,000,000 years  
   b) 64,000,000 years  
   c) 100,000,000 years  
   d) 165,000,000 years  
   e) 200,000,000 years

16. What technique do scientists use today to determine when the Earth first formed? Choose the best answer.
   a) Calculation of the cooling rate of metal spheres  
   b) Comparison of different layers of rock  
   c) Analysis of uranium and lead found in rocks  
   d) Analysis of carbon found in rocks  
   e) Correlation of fossils in different layers of rock
17. An Archaeologist is studying layers of human settlement in an African cave. What method of radiometric dating should they use to determine the age of woody material found at an apparent fire pit in one of the layers?
   a) Potassium – Argon
   b) Argon – Argon
   c) Carbon – Oxygen
   d) Rubidium – Strontium
   e) Carbon – Nitrogen

18. Why is $^{238}\text{U}-^{206}\text{Pb}$ dating **not** the best method for determining the age of the woody debris from the previous question? Choose the **best** answer.
   a) Uranium is only found in Canada, and the cave is located in Africa
   b) Fire in the pit would cause the woody debris to reach closure temperature and all the daughter nuclides would be lost
   c) $^{238}\text{U}-^{206}\text{Pb}$ dating would only reveal the time of formation of the woody debris, not the actual time of deposition.
   d) $^{238}\text{U}-^{206}\text{Pb}$ dating is not accurate within the amount of time represented by the sample.
   e) Woody debris cannot be run through the analytical methods needed to determine $^{238}\text{U}-^{206}\text{Pb}$ ratios.

19. Which of the following labeled events would correlate to the **base** of the Mesozoic and the **base** of the Cenozoic?
   a) A, H
   b) I, J
   c) A, G
   d) H, J
   e) E, I

20. Place the following tectonic events into **chronological order** (oldest to youngest):

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>a) A, B, C, D</td>
<td>b) A, C, D, B</td>
<td>c) B, D, A, C</td>
<td>d) C, A, D, B</td>
</tr>
</tbody>
</table>
21. Using the timeline of Earth’s history below, place a tick mark on the line to represent where each of the following events occurs. Label your tick marks with the appropriate letter. (9 pts)

A – Dinosaurs Appear  
B – Origin of Life on Earth  
C – Humans (Homo sapiens) Appear  
D – First Organisms with Hard Parts (shells, skeletons, etc.) Appear  
E – Non-Avian Dinosaurs go Extinct

1 inch = 1 billion years

Demographic Questions:

What is your academic rank?

☐ Freshman  
☐ Sophomore  
☐ Junior  
☐ Senior

What is your major?

☐ Geology  
☐ Other with a Geology Minor  
☐ Other _________________________________________

Please select all of the Geology courses you have taken

☐ MEA 100 Earth System Science  
☐ MEA 101 Geology I: Physical (or equivalent intro geology course)  
☐ MEA 202 Geology II: Historical (or equivalent Historical geology course)  
☐ MEA 300 Environmental Geology  
☐ MEA 323 Earth System Chemistry  
☐ MEA 410 Introduction to Mineralogy and Petrology  
☐ MEA 411 Marine Sediment Transport  
☐ MEA 440 Igneous and Metamorphic Petrology  
☐ MEA 450 Introductory Sedimentary Petrology/Stratigraphy  
☐ MEA 451 Structural Geology  
☐ MEA 465 Geologic Field Camp  
☐ MEA 481 Geomorphology: Earth’s Dynamic Surface  
☐ MEA 485 Introduction to Hydrogeology
Have you done any undergraduate research projects related to the topics covered on this test?

☐ YES
☐ NO

If yes, please explain the research briefly:

**Demographic Questions for Faculty:**
1. What is the highest level of education you have completed?
   - ☐ Bachelor’s degree
   - ☐ Master’s degree
   - ☐ Doctoral degree
   - ☐ Other ________________________________

2. What is your rank at the institution you work?
   - ☐ Instructor/teaching professor
   - ☐ Assistant professor
   - ☐ Associate professor
   - ☐ Full professor
   - ☐ Other ________________________________

3. What is the highest level of degree awarded in your department?
   - ☐ Associates degree
   - ☐ Bachelor’s degree
   - ☐ Master’s degree
   - ☐ Doctoral degree

4. How many years of experience do you have (excluding graduate school)?
   - ☐ > 5
   - ☐ 5-10
   - ☐ 11-15
   - ☐ 16-20
   - ☐ > 20

5. In what geoscience field is your specific area of expertise? ________________________________

6a. Do you teach any courses that include presentations on any of the following concepts? Select all that apply.
   - ☐ Landscape/formation identification
   - ☐ Landscape formation rates
   - ☐ The geologic time scale
   - ☐ Relative and numeric dating
   - ☐ Timing of Events in Earth history
6b. If so, what types of courses are these?
- Introductory courses
- Upper level majors courses
- Graduate level courses

7. Please name the courses:

8. Do you conduct research in areas that involve any of the following concepts? Select all that apply
- Landscape/formation identification
- Landscape formation rates
- The geologic time scale
- Relative and numeric dating
- Timing of Events in Earth history
Images to be projected for #s 1-7:

#1

Copyright © Marli Miller, University of Oregon; Image Source: Earth Science World Image Bank
http://www.earthscienceworld.org/images

#2

Copyright and Source Unknown
#3

Copyright © B.P. Snowder; Image Source: Western Washington University Planetarium
http://www.wwu.edu/depts/skywise/a101_meteors.html

#4

Copyright © Travel Adventures; Image source:
Copyright © Luca Galuzzi; Image Source: Imaging Earth’s Surface: University of Vermont
http://www.uvm.edu/~geomorph/gallery