Exploring and Measuring the Teaching and Development of Earth Systems Thinking Skills in Undergraduate Geoscience Courses

by

Nicholas Andrew Soltis

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
May 2, 2020

Earth System Thinking, Biogeochemistry, Instrument Development, Structural Equation Modeling

Approved by

Karen S. McNeal, Chair, Associate Professor Department of Geosciences
Christine G. Schnittka, Professor Department of Curriculum and Teaching
Stephanie L. Shepherd, Assistant Professor, Department of Geosciences
Joni M. Lakin, Associate Professor Department of Educational Foundations, Leadership, and Technology

ABSTRACT

The understanding of the development of Earth systems thinking (EST) is a major challenge in the field of geoscience education. Major questions exist relating to systems thinking teaching practices, student conceptions of complex Earth systems, and the assessment of systems thinking skills in the context of the Earth system. This research integrates these three strands through three individual studies representing an aspect of each strand. The first study aims to quantitatively build on this work done on EST teaching practices by employing structural equation modeling to understand the current state of EST-teaching as shown by the 2016 iteration of the National Geoscience Faculty Survey (n=2615). Exploratory and confirmatory factor analyses were conducted on survey items to understand and develop three models, one for EST-teaching practices, one for course changes, and one for active-learning teaching practices. Analyses revealed that reported EST-teaching practices relate to four EST frameworks proposed in the literature. The three models explored in this study were used to build a full structural model, where it was hypothesized that active-learning teaching practices would predict EST-course changes and EST-teaching. However, the model revealed that EST-course changes mediate, or bring about, the relationship between active-learning teaching practices and EST-teaching. This implies the need for continued efforts to provide professional development opportunities in both active learning teaching practices and EST, as active-learning practices are not sufficient to implicitly teach EST skills. Results also revealed that the teaching approaches that emphasize modeling and complexity sciences had the weakest relationship to the broader EST-teaching practices, suggesting a need for more professional development opportunities as they relate to systems modeling, quantitative reasoning, and complexity sciences in the context of the Earth sciences.

The second study explores student conceptions of how the spheres of the Earth system are linked through the biogeochemical cycles that move matter and energy through the various parts of the Earth system. This study aims to fill a gap in the literature by examining how undergraduate students perceive fluxes and reservoirs of important elements within the Earth system: namely carbon, nitrogen, and phosphorus. Through interviews and concept drawings undergraduate students' conceptions and alternate conceptions about the Earth System and biogeochemical cycles were collected. These data were analyzed to provide a more complete understanding of what students know, do not know, and what they think they know about both the Earth System and biogeochemical cycles. This study revealed that undergraduate students across disciplines tend to hold a "bio-centric" view of the carbon cycle and have more limited conceptions in terms of detail and breadth of the nitrogen and phosphorous cycles. Additionally, conceptual drawings revealed a notable absence of the hydrosphere in students' mental models of all three cycles. Students who took more STEM courses and were in more interdisciplinary fields (i.e. geology, science education) tended to have more nuanced (though not necessarily complete) conceptions of these cycles. Implications for this study involve the improvement of teaching biogeochemical concepts across disciplines, but also inform our knowledge about using these cycles in the context of systems thinking. This work also provides a baseline for future work on developing learning progressions for biogeochemical cycles and complex Earth systems and in assessing systems thinking abilities through student knowledge of biogeochemical cycles.

The third study documents the development process of the Earth systems thinking concept inventory (EST CI). Evidence of validity and reliability was accrued using elements of both classical test theory (CTT) and item response theory (IRT). By using these two approaches to validity and reliability in a complementary fashion, we were able to take an iterative approach to provide robust evidence for both validity and reliability. Additionally, the instrument is semi-customizable as language regarding feedbacks can be shifted between using ‘positive’ and ‘negative’ or ‘reinforcing’ or balancing’ terminology, with the later resulting in better reliability among a largely novice audience.

ACKNOWLEDGMENTS

I've been fortunate during my time at Auburn to be supported by incredible mentors and friends. I would like to thank my advisor, Dr. Karen McNeal, for always supporting me, believing in me, and pushing me to always be my best. I could not ask for a better mentor and friend to guide me through my Ph.D. It's been a wild ride going from North Carolina State to Auburn, but Karen always made sure that I was taken care of. I would also like to thank my amazing committee members: Dr. Chris Schnittka for making me a better teacher, Dr. Joni Lakin for her support in all things statistics, and Dr. Stephanie Shepherd for taking me under her wing my semester at Auburn. In addition, I am grateful to Dr. Cissy Ballen for being my University reader and an incredible colleague and friend in biology education research.

I would also like to thank the Department of Geosciences for the four years of support and for giving me the opportunity to teach and grow as a teacher during my time in the department. I truly appreciate all of the faculty who have supported me and believed in me through my time here. I would especially like to acknowledge Dr. Chuck Savrda, Dr. Ann Ojeda, Dr. Stephanie Rogers, Dr. Laura Bilenker, Dr. Brennen van Alderwerelt, and Dr. Martin Medina, for their friendship, support, and mentorship. I also want to acknowledge Tony Hall, Ashleigh Rudd, and Amy Goode for always keeping my technology running, myself functioning, and myself paid during my time at graduate school. Others who were instrumental through my various adventures in higher education include Dr. Astrid Schnetzer and Dr. Lonnie Leithold at NC State; Dr. Renee Clary, Dr. John Rogers, and Dr. Sarah Lalk at Mississippi State; and Dr. Linda Rogness, Dr. Jean Morris, Dr. Shelley Amstutz-Szalay, Dr. Laura Hilton, and Dr. Dave Rodland at Muskingum.

I would also like to thank my fellow graduate students who became a second family during graduate school and helped me maintain my sanity. Thank you to my lab mates past and present Dr. Lindsay Maudlin, Rachel Atkins, Steph Courtney, Eli Johnson, Akilah Alwan, Haven Cashwell, and Tyler Smith for always being there for me and a constant source of support. I would also like to thank countless other students in the geosciences who have become dear friends over the years and were there for all the ups and downs. The list is too long to name everyone and I would hate to accidentally forget anyone, but I am eternally grateful.

Lastly, I want to thank my family. My parents Gerry and Betty Soltis have always been my number one supporters and have always been by my side, even when I wanted to give up my job and everything I built to take a massive pay cut and go back to graduate school. I owe my sister, Maggie, and brother-in-law Jacob thanks for letting me live with them during my stint at NC State and for their unwavering support. I also thank Maggie for making me pick up productive habits like Pokémon Go during my time completing my doctorate. I wouldn’t be where I am today with the love and support of my family, and thus I dedicate this dissertation to them.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

EST	Earth Systems Thinking
SEM	Structural Equation Model
EFA	Exploratory Factor Analysis
CFA	Confirmatory Factor Analysis
CTT	Classical Test Theory
CI	Concept Inventory
EST CI	Earth Systems Thinking Concept Inventory
IRT	Item Response Theory

CHAPTER 1: INTRODUCTION

Beginning in the 1970s, several events fundamentally changed the way that geoscientists looked at the Earth. The first event was the now-iconic photo of Earth from space taken by the crew of *Apollo 17* in 1972. The second event was the emergence of the concept of complex systems from the field of mathematics (Ormand, 2011, Gleick, 1988, Manduca & Kastens, 2012). A view from space opened up new ways to collect data about the planet as a whole, introducing the new field of Earth system science (NASA, 1988). The application of complex systems—that is systems that are chaotic, fractal, self-organizing, and sensitive to initial conditions—opened new frontiers as to how geoscientists study the Earth from the micro to macro scale. This concept of complex Earth systems brings together two major concepts. First, the Earth viewed as a single interacting system with many contributing subsystems, and second, the characterization of Earth processes as resulting from complex systems behavior (Manduca & Kastens, 2012). This concept of complex Earth systems not only applies to geoscientists who research components or the whole of the Earth system but also how geoscience is taught.

Perspectives on Earth Systems Thinking

The geosciences encompass a broad range of disciplines and subdisciplines, all surrounding the study of the Earth. These disciplines include, but are not limited to, geology, meteorology, climatology, oceanography, environmental science, etc. Phenomenographic work by Stokes (2011) demonstrated that the most advanced and complex conceptions of the geosciences involve the conceptions of interacting systems and the relationship between Earth and society. Thus, increasing interest has been paid to the development of Earth systems thinking and teaching in geoscience courses. This Earth systems perspective allows for an integrated view of the Earth as interacting parts, rather than approaches to geoscience that involve presenting facts about the Earth, ocean, atmosphere, and life without highlighting the interactions between all components. This concept is highlighted by Ireton and colleagues' (1997) call to study the Earth as an integrated system, characterized by the dynamic interactions between the atmosphere, hydrosphere, solid Earth, biosphere, and anthroposphere (built or human components). This stance, sometimes known at the "anti-stovepipe stance," rejects the teaching of the geosciences from the perspective of independent, isolated disciplines ("stovepipes"). This Earth systems approach to education (Mayer, 1991; Ireton et al., 1997) has been particularly well documented in K-12 science education literature (NGSS Lead States, 2013; Orion and Libarkin, 2014; The College Board, 2016) and geoscience workforce expertise (U.S. Bureau of Labor Statistics, 2015). The new Next Generation Science Standards (2013) employed by multiple states emphasize the importance of an Earth systems perspective in K-12 education and reflect an important shift in science education and the intrinsic value of an Earth systems perspective.

Naturally, Earth systems thinking is predicated on the learner's systems thinking abilities (Orion and Libarkin, 2014). Systems thinking was a term originally used to indicate a holistic approach that considers dynamic interdependencies among part, or in terms, seeing a whole as a sum of its parts (Arnold & Wade, 2015). This definition has been expanded to include several emergent characteristics of systems thinking. These characteristics include the ability to identify the components of a system and processes within the system, the ability to identify relationships among the system's components, the ability to organize the systems' components and processes within a framework of relationships, the ability to make generalizations based on the system, the ability to identify dynamic relationships within the system, the ability to understand the hidden dimensions of a system, the ability to understand the cyclic nature of systems, and the ability to think temporally employing retrospection and prediction (Assaraf & Orion, 2005). Systems thinking and its development are not unique to the geosciences and have been explored by a wide range of fields, highlighting the transdisciplinary nature of systems thinking (Trujilo & Long, 2018). The idea of Earth systems thinking, however, goes beyond pure systems thinking and considers the role that humans play within the Earth system (Manduca & Kastens, 2012; Gosselin et al., 2013; Orion and Libarkin, 2014; InTeGrate Program, 2015; Orr et al., 2016; Kastens & Manduca, 2017). Stillings (2012) synthesized many ideas about EST, particularly as they relate to complexity, and outlined many future pathways for curriculum, instruction, and research on complex Earth systems.

Manduca and Kastens (2012), created a concept map for the domain of thinking and learning about complex Earth systems. In their concept map, the three main components we Earth systems, complex systems, and pedagogy. Making Earth systems thinking unique from systems thinking in other disciplines is the fact that concepts of complex systems are juxtaposed on content about a complex system, the Earth system. This presents many challenges when considering teaching and learning about the Earth system. Stillings (2012) offers a psychological perspective on thinking about complex systems. In particular, thinking in systems is not a way of thinking that comes easily to humans nor is it a traditional way of thinking. Herbert (2006) laid out three particular cognitive challenges to student understanding of complex Earth systems. These include conceptualization of natural Earth environments as systems, understanding the complex characteristics of systems, and the application of conceptual models of complex Earth systems to support environmental problem-solving.

As mentioned previously, humans have some degree of fundamental difficulty in understanding complex systems. Despite this documented innate difficulty, an Earth systems perspective allows for a much more accurate and holistic understanding of geosciences.

Additionally, beyond that, the innate documented challenge of developing systems thinking skills makes the development of these skills all the more important. Additionally, an Earth systems perspective can make these skills salient to students from a variety of backgrounds. This is due to the fact that an Earth systems approach presents geoscience content in the context of people's daily lives rather than in an isolated discipline. Many of these natural cycles that connect components of the Earth system (biogeochemical cycles, food chains, the rock cycle) directly affect humans (Orion & Libarkin, 2014). Additionally, the holistic framework of an Earth systems approach serves as an effective learning tool for the development of environmental insights, as well as an important lens for examining issues of sustainability such as global climate change, loss of biodiversity, and ocean acidification. These complex phenomena simply cannot be properly conceived through the lens of any single discipline, and truly require an Earth systems perspective to address.

Earth Systems Frameworks and Current Practices

A literature review was conducted by Scherer and colleagues (2017) to identify the current state of the study of learning and teaching the Earth system in geoscience education research. This work built on the pathways laid out by Stillings (2012) and identified four conceptual frameworks (capitalized for clarity) relating to complex systems within the geoscience education research literature: Earth Systems Perspective, Earth Systems Thinking Skills, Complexity Sciences, and Authentic Complex Earth and Environmental Systems. By exploring each framework, we gain insight into ways instructors and researchers are approaching the nebulous and broad field of Earth system science. The Earth Systems Perspective Framework focuses on the interactions of the four major spheres of the Earth system and their complex interconnections. This framework is concerned with the interdisciplinary nature of the Earth system and limits systems thinking to conceptualize the Earth system as a whole. This framework often also includes aspects of human interactions and environmental decision-making (Davies, 2006).

The second framework, Earth Systems Thinking Skills, emphasizes systems thinking skills, particularly as they relate to cyclic and dynamic thinking. Work by Assaraf and Orion (2005) on applying systems thinking to the transformation of matter in Earth cycles—e.g. the water cycle—exemplify this framework. Work on identifying feedback loops as well as the understanding of underlying causes of processes are part of this framework. Thus, this framework can be differentiated from the Earth systems perspective by its emphasis on the inclusion of specific systems thinking skills and abilities. While systems thinking is highly emphasized in this framework, heavy use of computer modeling or consideration of complex systems or chaos theory is not included.

The framework of Complexity Sciences largely pulls from the theoretical tradition of the interdisciplinary study of complex systems. This framework is embodied by a wide array of studies that have considered complexity science from the lens of systems dynamics (Shepardson et al., 2014), complex systems theory (including mathematical approaches) (Fichter et al., 2010), and Gaia theory (explaining the environmental conditions of Earth in terms of biological forcing) (Haigh, 2002, 2014). Structure-behavior-function analysis, which emphasizes thinking about how a system works and its function rather than focusing on the components, is also included in this framework and is evidenced in work by Hmelo-Silver and colleagues (2014). Computer modeling work often falls within this framework.

The final framework of Authentic Complex Earth and Environmental Systems pulls its systems ideas from the study of real-world environmental or ecological activities. This framework often makes intentional connections to human activities and environmental decision-making. In this framework, the systems thinking emphasis is on a real-world environmental system or phenomenon, thus incorporating complexity and looking beyond any one process or single component (Herbert, 2006). This framework, more so than the other three, is highly contextualized. Examples of work done in this framework include studies on student reasoning on real-world systems like coastal eutrophication (Sell et al., 2006; McNeal et al., 2008), ecosystem dynamics (Grotzer et al. 2013), soil microbial activity (Appel et al., 2014), and water in socio-ecological systems (Gunckel et al. 2012).

These four frameworks, each with strengths and limitations, represent four ways in which educators and researchers are working to understand and develop Earth systems thinking skills and in what contexts. These frameworks also offer a more focused way to consider Earth systems thinking instructional practices than the often-nebulous umbrella of complex systems, and provide us a path forward as we think about addressing the challenges presented by teaching in an Earth systems context. Additionally, these frameworks offer us insights into different teaching practices as they relate to Earth system science.

Challenges in Teaching Earth System Science

Professional thinking in the geosciences has been traditionally seen to be based on three abilities: the ability to think temporally, the ability to think spatially, and the ability to observe in the field; however, the construct of the Earth system now demands the ability to think about and grapple with complex systems (Kastens & Manduca, 2012; Stillings, 2012). As previously discussed, this means that research and teaching in the geosciences is increasingly influenced by the concept of Earth system science. Not only does this perspective introduce the role of multiple processes in geoscientific phenomena, but additional complexity is derived from the intricate, nonlinear interactions among processes (Stillings, 2012). This means that the complexity of Earth systems typically cannot be reduced through experimental control and replication, meaning evidence is largely observational and requires the measurement of multiple variables so that causal relationships can be teased apart through the use of reasoning and logic (Cleland, 2001, 2002). This clearly is contrary to how many undergraduate students view the nature of science and presents significant challenges to teaching and learning.

The complex nature of the Earth system makes it challenging to teach and to understand students' learning and conceptual change. Herbert (2006) identified three challenges regarding the reasoning behind complex thinking. The first is the fact that many Earth processes occur at scales—both spatial and temporal—beyond human experience (Dodick and Orion, 2003; Giorgi, 1997.) The second challenge is the difficulty in developing accurate conceptual models of complex systems in which several variables are controlling the behavior of the system (Berger, 1998). The last of Herbert's challenges is the tendency of individuals to disregard beyond average data as noise, meaning that individuals try to oversimplify systems as being near or at equilibrium, ignoring important information or data as just "noise." This is problematic as many systems are not at or near equilibrium (Goldenfeld and Kadanoff, 1999). Considering this, Herbert emphasizes the importance of inquiry-based and other active learning strategies as necessary in helping students understand complex Earth systems.

From the learning sciences, we know that memory is an important component of recalling learned material for it to be applied. Learned material has been shown to be more meaningful, better understood, and more richly interconnected with other contents of memory is more likely to be recalled (Anderson, 2009). Thus, material that is taught in a geoscience course that is free from interconnections to transdisciplinary context is unlikely to be recalled as anything but an interconnected series of facts (Stillings, 2012). This is particularly meaningful, as it implies that systems thinking skills must be specifically taught and are not going to just be "picked up" as more information is learned. Additionally, information must be presented in the context of the system in which it resides. This is coupled with the ineffectiveness of approaches taken in some introductory classes where students are taught facts about science or taught how to solve stereotyped quantitative problems, yet are then expected to learn to use these facts out of context for scientific thinking in other courses (Trowbridge & McDermott, 1981; Crouch & Mazur, 2001). Thus, we need to rethink how we more explicitly teach Earth system science and Earth systems thinking skills.

What we know about human reasoning abilities creates additional challenges to teaching systems thinking skills in the context of the Earth sciences. Humans naturally form inductive generalizations and construct causal explanations; however, these generalizations and explanations tend to be based only on confirmatory evidence. Additionally, reasoning about situations that involve thinking through multiple possible influences on singular outcome tends to be poor (Kuhn, 1991; Kuhn & Dean, 2004). Prior conceptions about a phenomenon are also very difficult to overcome (Novick and Nussbaum, 1978). These patterns of building knowledge are antithetical to systems thinking and must be confronted to effective teach systems thinking skills. This knowledge from learning science indicates that using scientific concepts in the sorts of activities in which scientists engage should be a feature of science learning and teaching throughout a student’s education. Additionally, this implies that student activities need to be carefully structured and sequenced to provide pathways that overcome the barriers to learning and thinking like a scientist and in turn thinking in systems (Clark and Linn, 2003).

Another challenge to overcome to build students’ Earth systems thinking skills is the development of metacognition. It has been suggested that students’ metacognitive skills may have a reciprocal causal relationship with their explicit views of the nature of science (Sandoval, 2005). Metacognition, or thinking about thinking, is an important skill demonstrated by expert geoscientists who distinguish between theories and evidence and are aware of multiple kinds of data. They are also aware of how emerging data can change current theories and thus are sensitive to the degree of confirmation. Expert geoscientists are also able to think of ideas and knowledge hierarchically and can effectively organize that information while still being cognizant of alternate explanations (Stillings, 2012). All of these characteristics of expert geoscientists—which are necessary for Earth systems thinking-- require metacognition. Thus, the development of metacognition in students is an important component of developing Earth systems thinking skills in geoscience courses. It also suggests a tight coupling between students' epistemological understanding of the nature of science with their metacognitive skills and by proxy a possible relationship between students' epistemology and Earth systems thinking abilities.

Connecting Learning Science to Teaching Earth System Science

The psychological challenges and barriers to developing systems thinking skills have specific implications for learning and instruction. Model-based reasoning, which may include reasoning based on models such as texts, mathematical expressions, visual representations, computer programs, and physical models), is certainly central to modern science (Magnani et al., 1999; Neressian, 2002). While models transcend any discipline, reasoning about models is especially essential to any sort of complex natural system. Model-based reasoning, when employed in geoscience classrooms, engages students in a number of levels, from remembering the structure of a model and understanding the model well enough to reason with it, to developing a reflective awareness of the importance of models for understanding complex systems, model-based reasoning can work to develop memory and recall, reasoning, and metacognitive skills needed for the development of systems thinking skills (Stillings, 2012). This is especially important because research has found that an individual’s stable internal representation of a model in terms of its coherence, depth, correctness, generality, and generativity is crucial to performance across learning situations (Anderson et al., 1996; Vosniadou, 2007). At the same time, students’ mental models tend to be incomplete or flawed relative to an expert standard (Raghavan et al., 1998). Systems models, in particular, have been identified as particularly difficult to understand across fields (Hmelo-Silver et al, 2007). In the geosciences, systems concepts are found in curriculum from the earliest levels (e.g. the water cycle), and learning goals throughout the curriculum require the mastery of fairly sophisticated systems concepts as well as the integration of information likely learned in a variety of disciplines (biology, chemistry, physics, etc.) (Stillings, 2012).

Model-based reasoning is not a quick fix for developing systems thinking skills; rather, learning goals associated with model-based reasoning can only be achieved in multiple steps with a great deal of scaffolding. The concept of learning progressions (Duncan & Hmelo-Silver, 2009) integrates several factors that explain the challenges of teaching model-based reasoning. First, systems concepts develop over time, only as students begin to integrate important features and concepts and start to apply them to their reasoning contexts. Second, for most learners, successful trajectories for learning depend on well thought out and sequenced instruction (scaffolding). Third, the interval from initial knowledge to desired cognitive and metacognitive outcomes typically outlasts traditional planning units (courses, semesters, school years, and even curriculums). Thus, the teaching of Earth system science requires a focus on learning progressions that integrate curriculum over extended periods as well as those that incorporate assessment in order to understand immediate states of knowledge and develop appropriate interventions (Stillings, 2012).

These connections between learning science and Earth system science emphasize key points in designing and implementing courses that emphasize Earth systems thinking as proposed by Stillings (2012). First, instruction in the Earth science should teach systems concepts in a multidimensional, integrated, and meaningful way that promotes persistence in memory. Second, instruction should promote reasoning with system models. Third, instruction should promote students' awareness of the nature of science and this should be used to guide their thinking and develop metacognitive skills (Sandoval, 2005). In application, this means that systems thinking should be explicitly taught in all parts of an undergraduate curriculum, and not assumed that these skills will be picked up passively or taught somewhere else. Assessment must be used extensively to monitor student learning and to design and structure interventions. Emphasis should also be placed on making distinctions between models and reality as well as on the nature of science within geoscience courses. Instructional technology should also be utilized to promote model-based reasoning and also to give students more experience working with and manipulating complex systems (Stillings, 2012).

Gaps in the Research and Future Directions

The complexity of Earth system science education and the challenges associated with the development of systems thinking means that there is much to learn about how to best develop systems thinking skills in undergraduate geoscience courses. While we know that students have difficulty with complex systems, there is limited research on student conceptions of particular complexity concepts in Earth systems, in particular, more work is needed on student conceptions of feedback loops, following the lead of Raia (2005, 2008) and Batzri and colleagues (2015). In particular, qualitative work is needed on these systems conceptions. Additionally, in the vein of Mohan and colleagues (2009) and Assaraf and Orion (2010), more work needs to be done on learning progressions in Earth system science classes. Some work has been done on specific concepts such as plate tectonics (McDonald et al., 2019), sea-level rise (Breslyn et al., 2016), water in socio-ecological systems (Gunckel, et al., 2012), and climate change (Chang et al., 2017). It would be especially interesting to look at broader learning progressions across entire undergraduate Earth system science curriculums. Additional work on the ever-changing world of learning technology and its applications to complex systems and in particular modeling is also needed.

The connection between metacognition and systems thinking skills in the Earth sciences is another area worthy of future research. Holder and colleagues (2017) have explored the role of problem-solving expertise in student learning of complex Earth systems. The use of instructional resources like InTeGrate and how they develop systems thinking skills while emphasizing metacognition has been explored by Gilbert and colleagues (2018). This overall area of research—metacognition and systems thinking—presents an exciting opportunity for additional collaboration between cognitive scientists and geoscience education researchers to explore this important relationship. Additionally, the relationship between science epistemology, metacognition, and potentially systems thinking skills warrants additional investigation. This line of research requires the understanding of how student understanding of the nature of science and the nature of geoscience plays in shaping metacognition and systems thinking ability as suggested by Grotzer and Lincoln (2007) and Raia (2005, 2008).

A serious need exists for large-scale studies to examine the wide range of systems thinking teaching practices in the context of geoscience courses. While many studies have been conducted examining teaching practices and systems thinking on a small scale, there has not been work done on a national or broader scale to analyze trends in systems thinking instructional practices. Large scale work can also allow the geoscience community to better understand what resources and professional development are needed to help instructors improve their teaching of Earth system science. Initial work with the National Geoscience Faculty Survey to understand systems modeling teaching practices has been undertaken by Lally and colleagues (2019) and to understand broader Earth systems teaching practices and their relation to active learning practices by Soltis and colleagues (2019).

Another major area of interest in the development of instruments to measure students' systems thinking abilities (Orion & Libarkin, 2014). It is unlikely that one instrument can capture all dimensions of the context of Earth systems thinking, especially after the literature analysis of Scherer and colleagues (2017). So rather, a host of different instruments will need to be developed and validation studies completed to better assess how students are developing systems thinking abilities. Arnold and Wade (2017) have developed a complete set of systems thinking skills that cut across multiple disciplines that has the potential to serve as an important starting point. Instruments for systems from fields like ecology (Jordan et al., 2014) and paradigms like ill-structured problems (Grohs et al, 2018) have been developed and may be of use. Also important is the development of instruments that may measure affect and can be used to understand the relationship between Earth system science instruction and student affect (Orion & Libarkin, 2014). As noted earlier, there is also a need for more qualitative work to understand undergraduate student conceptions of the Earth system and specific systems concepts, which is also necessary for instrument development.

Orion and Libarkin (2014) proposed several broad areas for future research in Earth system science education. The first area proposed is environmental insight. Systems thinking skills are crucial for developing environmental insight; however not much research has considered this relationship or the mechanism of how it may develop. Conceptual understanding is another major area of proposed research in Earth system science. While there has been significant research among a variety of populations for the presence or absence of understanding of specific core ideas in the geosciences, this research needs to shift to a more Earth systems approach. Meaning researchers need to move beyond exploring conceptual understandings of isolated parts of the Earth system, to asking questions about how well learners are able to understand and reason about interacting processes and systems. Another proposed area of research is in decision making, and the role that Earth system science education can play in that decision-making process, particularly as it relates to issues of sustainability. Expertise research is also proposed and is seen as especially important as the geoscience fields change more and more to reflect an Earth system paradigm. This involves studying expertise in traditional fields like geology and meteorology, but also in newer or emergent fields like biogeochemistry, climate change science, medical geology, etc. Lastly, novel technology is proposed as major area of research. As discussed earlier, novel technology has incredible potential to increase model-based reasoning and give students more opportunities to interact with and manipulate complex systems.

The Earth systems approach to geoscience education and the development of related Earth systems thinking skills is an exciting frontier in geoscience education research. The Earth systems approach has been expanding through the geoscience disciplines since the 1970's, there is still much work to be done to understand how to best teach Earth system science and to reform and improve traditional geoscience undergraduate programs. Much progress has been made in understanding the benefits of Earth systems instructional practices and the role of systems thinking in geoscience thinking and learning, but there is still much more to learn. As the geoscience disciplines continue to evolve and change, it is critical that as researchers we work to ensure that the way we teach geoscience contents evolves at the same rate. This allows us to not only help students become better geoscientists, but also more environmentally aware and sustainability minded generations. In the short term, we all benefit from improving our own systems thinking skills, but in the long term our planet and our society benefit by developing systems thinking skills and in turn a population that better understands the complex planet on which we live. Understanding Earth system science education truly is a grand challenge and a tremendous opportunity for geoscience education researchers.

References

Arnold, R.D., & Wade, J.P. (2015). A definition of systems thinking: A systems approach. *Procedia Computer Science*, *44,* 669-678.

Anderson, J. R. (2005). *Cognitive psychology and its implications*. Macmillan.

Anderson, J. R., Reder, L. M., & Simon, H. A. (1996). Situated learning and education. *Educational researcher*, *25*(4), 5-11.

Assaraf, O. B. Z., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of research in science teaching*, *42*(5), p. 518-560.

Assaraf, O. B. Z., & Orion, N. (2010). System thinking skills at the elementary school level. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *47*(5), 540-563.

Batzri, O., Assaraf, O. B. Z., Cohen, C., & Orion, N. (2015). Understanding the Earth systems: Expressions of dynamic and cyclic thinking among university students. *Journal of Science Education and Technology*, *24*(6), 761-775.

Breslyn, W., McGinnis, J. R., McDonald, R. C., & Hestness, E. (2016). Developing a learning progression for sea level rise, a major impact of climate change. *Journal of Research in Science Teaching*, *53*(10), 1471-1499.

Berger, R. (1998). Understanding science: Why causes are not enough. *Philosophy of Science*. v. 65, no. 2, p. 306–333.

Chang, C. H., Tan, I. G. C., Tan, J., Liaow, D., & Kwek, C. H. (2017). Learning progressions for climate change: How does it look like in Singapore's school geography?. Redesigning Pedagogy International Conference 2017, 31 May – 2 June 2017, Singapore

Clark, D., & Linn, M. C. (2003). Designing for knowledge integration: The impact of instructional time. *The Journal of the Learning Sciences*, *12*(4), 451-493.

Cleland, C. E. (2001). Historical science, experimental science, and the scientific method. *Geology*, *29*(11), 987-990.

Cleland, C. E. (2002). Methodological and epistemic differences between historical science and experimental science. *Philosophy of Science*, *69*(3), 447-451.

Crouch, C. H., & Mazur, E. (2001). Peer instruction: Ten years of experience and results. *American journal of physics*, *69*(9), 970-977.

Davies, C. P. (2006). Implementing earth systems science curriculum: Evaluating the integration of urban environments for an urban audience. *Journal of Geoscience Education*, 54(3), p. 364-373.

Dodick, J. & Orion, N. (2003). Measuring student understanding of geological time. *Science and Education*. v. 87, no. 5, p. 708–731.

Duncan, R. G., & Hmelo-Silver, C. E. (2009). Learning progressions: Aligning curriculum, instruction, and assessment. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(6), 606-609.

Fichter, L. S., Pyle, E. J., & Whitmeyer, S. J. (2010). Strategies and rubrics for teaching chaos and complex systems theories as elaborating, self-organizing, and fractionating evolutionary systems. *Journal of Geoscience Education*. *58*(2), p.65-85.

Giorgi, F.A.R.(1997). Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling. *Reviews of Geophysics*, v. 35, no. 4, p. 413.

Gilbert, L. A., Gross, D. S., & Kreutz, K. J. (2018). Developing undergraduate students' systems thinking skills with an InTeGrate module. *Journal of Geoscience Education*, 1-16.

Gleik, J. (1988). *Chaos*: Making a new science: New York, Penguin.

Goldenfeld, N. & Kadanoff, L.P. (1999). Simple lessons from complexity. *Science*, v. 284, no. 5411, p. 87–89.

Gosselin, D.C., Manduca, C., Bralower, T., & Mogk, D. (2013). Transforming the teaching of geoscience and sustainability. *Eos Trans AGU Publications*. Vol. 94, no. 25, pp. 221-222.

Grohs, J. R., Kirk, G. R., Soledad, M. M., & Knight, D. B. (2018). Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. Thinking Skills and Creativity, 28, 110-130.

Grotzer, T. A., Kamarainen, A. M., Tutwiler, M. S., Metcalf, S., & Dede, C. (2013). Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus. *BioScience*, *63*(4), p. 288-296.

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching*, *49*(7), p. 843-868.

Haigh, M. J. (2001). Constructing Gaia: Using journals to foster reflective learning. *Journal of Geography in Higher Education*. *25*(2), p. 167-189.

Haigh, M. (2014). Gaia:“thinking like a planet” as transformative learning. *Journal of Geography in Higher Education*. 38(1), p. 49-68.

Herbert, B. E. (2006). Student understanding of complex earth systems. *Special Papers Geologic Society of America*. 413, 95.

Hmelo-Silver, C, Eberback, C., & Jordan, R. (2014). Technology-supported inquiry for learning about aquatic ecosystems. Eurasia Journal of Mathematics, Science & Technology Education, 10, p. 405-413.

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *The Journal of the Learning Sciences*, *16*(3), 307-331.

Holder, L. N., Scherer, H. H., & Herbert, B. E. (2017). Student learning of complex Earth systems: a model to guide development of student expertise in problem-solving. *Journal of Geoscience Education*, *65*(4), 490-505.

Ireton, F. W., Manduca, C. A., & Mogk, D. W. (1997). Towards a coherent plan for undergraduate earth science education: a systems approach. *Journal of College Science Teaching*. 26(5), p. 304.

InTeGrate Program, (2015). InTe Grate pedagogy: Themes and strategies for teaching about geoscience and societal issues. Available at http://sec.carleton.edu/integrate/about/pedagogy_theme.html (accessed 18 April 2018).

Jordan, R. C., Sorensen, A. E., & Hmelo-Silver, C. (2014). A conceptual representation to support ecological systems learning. Natural Sciences Education, 43(1), 141-146.

Kastens, K. A., & Manduca, C. A. (Eds.). (2012). *Earth and mind II: A synthesis of research on thinking and learning in the geosciences* (Vol. 486). Geological Society of America.

Kastens, K. A and Manduca, C.A. (2017). Using systems thinking in the design, implementation and evaluation of complex educational innovations, with examples from the InTeGrate Project. *Journal of Geoscience Education*. Vol. 65, no. 3, August 2017, p. 219-230.

Kuhn, D. (1991). *The skills of argument*. Cambridge University Press.

Kuhn, D., & Dean, Jr, D. (2004). Metacognition: A bridge between cognitive psychology and educational practice. *Theory into practice*, *43*(4), 268-273.

Lally, D., Forbes C.T., McNeal, K., & Soltis, N. (2019). National Geoscience Faculty Survey 2016: Prevalence of systems thinking and scientific modeling learning opportunities. Journal of Geoscience Education, 67(2),174-191.

Magnani, L., Nersessian, N., & Thagard, P. (Eds.). (1999). *Model-based reasoning in scientific discovery*. Springer Science & Business Media.

Manduca, C. & Kastens, K. (2012). Mapping the domain of complex Earth systems in the geosciences. In Kastens, K., and Manduca C., eds. *Earth and mind II: A synthesis of research on thinking and learning in the geosciences*. Boulder, CO: Geological Society of America, p. 91-96

Mayer, V. J. (1991). Framework for Earth systems education. *Science Activities: Classroom Projects and Curriculum Ideas*. *28*(1), p. 8-9.

McDonald, S., Bateman, K., Gall, H., Tanis-Ozcelik, A., Webb, A., & Furman, T. (2019). Mapping the increasing sophistication of students' understandings of plate tectonics: A learning progressions approach. *Journal of Geoscience Education*, 1-14.

McNeal, K. S., Miller, H. R., & Herbert, B. E. (2008). The effect of using inquiry and multiple representations on introductory geology students' conceptual model development of coastal eutrophication. *Journal of Geoscience Education*. 56(3), p. 201-211.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *46*(6), 675-698.

National Aeronautics and Space Administration (NASA). (1988). Earth system science A closer view. Washington, D.C.: NASA Advisory Council, Earth System Sciences Committee, United States Aeronautics and Space Administration.

Nersessian, N. J. (2002). The cognitive basis of model-based reasoning in science. *The cognitive basis of science*, 133-153.

NGSS Lead States, 2013, Next generation science standards: For states, by states, Available at http://www.nextgenescience.org/next-generation-science-standards (accessed 18 April 2018).

Novick, S., & Nussbaum, J. (1978). Junior high school pupils' understanding of the particulate nature of matter: An interview study. *Science education*, *62*(3), 273-281.

Orion, N. & Libarkin, J. (2014). Earth system science education. *Handbook of research on science education*, 2, p. 481-496.

Ormand, C. (2011) What constitutes a complex system? In *On the cutting edge—Preofessional development for geoscience faculty: Developing student understanding of complex systems in the geosciences.*

Orr, C.H., Newman, A., & Manduca, C.A. (2016). InTeGrate: Interdisciplinary teaching of geoscience for a sustainable future. *In the Trenches*. Vol. 6, no. 3, pp. 3-4.

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

Raia, F. (2008). Causality in complex dynamic systems: A challenge in earth systems science education. *Journal of Geoscience Education*, *56*(1), 81-94.

Raghavan, K., Sartoris, M. L., & Glaser, R. (1998). Why does it go up? The impact of the MARS curriculum as revealed through changes in student explanations of a helium balloon. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *35*(5), 547-567.

Sandoval, W. A. (2005). Understanding students' practical epistemologies and their influence on learning through inquiry. *Science Education*, *89*(4), 634-656.

Scherer, H. H., Holder, L., & Herbert, B.. (2017). Student learning of complex Earth Systems: Conceptual frameworks of Earth Systems and instructional design. *Journal of Geoscience Education*. *65*(4), p. 473-489.

Sell, K. S., Herbert, B. E., Stuessy, C. L., & Schielack, J. (2006). Supporting student conceptual model development of complex earth systems through the use of multiple representations and inquiry. *Journal of Geoscience Education*, 54(3), p. 396-407.

Shepardson, D. P., Roychoudhury, A., Hirsch, A., Niyogi, D., & Top, S. M. (2014). When the atmosphere warms it rains and ice melts: Seventh grade students' conceptions of a climate system. *Environmental Education Research*, 20(3), p. 333-353.

Soltis, N.A., McNeal, K.S., Forbes, C., Lally, D. (2019). The relationship between active learning, course innovation and teaching Earth systems thinking: A structural equation modeling approach. *Geosphere*, 15(5), 1703-1721

Stillings, N. (2012). Complex systems in the geosciences and in geoscience learning. In Kastens, K., and Manduca C., eds. *Earth and mind II: A synthesis of research on thinking and learning in the geosciences*. Boulder, CO: Geological Society of America, p. 91-96.

Stokes, A. (2011), A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline, *Geological Society of America Special Papers*, 474, p. 23-35.

Trowbridge, D. E., & McDermott, L. C. (1981). Investigation of student understanding of the concept of acceleration in one dimension. *American journal of Physics*, *49*(3), 242-253.

The College Board. (2016). AP environmental science course overview. Available at https://apstudent.collegeboard.org/apcourse/ap-environmental-science (accessed 2/1/2019).

Trujillo, C. M. & Long, T. M., 2018, Document co-citation analysis to enhance transdisciplinary research, Science advances, 4(1), e1701130.

U.S. Bureau of Labor Statistics. (2015). *Occupational outlook handbook: Geoscientists.* Available at http://www.bls.gov/ooh/life-physical-and social-science/geoscientists.htm (accessed 18 April, 2018).

Vosniadou, S. (2007). The cognitive-situative divide and the problem of conceptual change. *Educational Psychologist*, *42*(1), 55-66.

CHAPTER 2: THE RELATIONSHIP BETWEEN ACTIVE LEARNING, COURSE INNOVATION, AND TEACHING EARTH SYSTEMS THINKING: A STRUCTURAL EQUATION MODELING APPROACH

Introduction

The teaching of geosciences takes many forms and covers a broad range of disciplines. Unifying all geoscience teaching is the understanding that the Earth itself is a system composed of interacting subsystems. To understand and improve the current state of post-secondary geoscience education and the impact of professional development aimed to improve undergraduate classroom teaching, it is necessary to understand how current geoscience instructors are teaching about the Earth system and incorporating aspects of systems thinking into their classrooms. Systems thinking was a term originally used to indicate a holistic approach through which to account for dynamic interdependencies among parts, or in other terms, seeing a whole as a sum of its parts (Arnold & Wade, 2015). This way of thinking is transferable to a variety of fields and disciplines, including the geosciences; however, this way of thinking is often challenging to undergraduate students (Stillings, 2012). One of the major challenges in teaching the geosciences is helping students to develop Earth systems thinking (EST) skills.

To address these needs, we conducted a study to investigate how geoscience instructors are engaged in teaching systems thinking in their classrooms and how teaching practices, particularly active-learning practices, are related to EST-teaching. We analyzed the results of the 2016 National Geoscience Faculty Survey to understand what EST-teaching currently looks like in post-secondary geoscience classrooms. Based on work by Scherer and colleagues (2017), we wanted to see if the frameworks described in the literature manifested themselves in how survey items about EST-teaching related to each other. We also examined the latent or underlying structure of participant responses to items related to changes or innovations that instructors were making to their courses, and if those changes were related to EST-teaching practices.

Additionally, this study aims to understand if instructors, who are more engaged in active-learning strategies, are more likely to engage in EST-teaching practices. Here, we hypothesized that instructors who engaged in active-learning practices would be more likely to engage in EST-teaching practices. This is based on the challenges associated with systems thinking (e.g., Herbert; 2006; Stillings, 2012), which are much more likely to be addressed by active learning than traditional lectures. Lastly, this study aims to understand if other factors relate to, influence, or bring about this hypothesized relationship. We hypothesized that active learning practices would influence both EST-teaching as well as changes, or innovations made to current courses. In summary, our research questions ask:

(1) What is the current state of EST-teaching in geoscience classrooms?

(2) How do instructor strategies for teaching EST relate to instructors' broader instructional practices?

(3) How do instructor strategies for teaching EST relate to recent changes instructors have made to their courses?

Perspectives on Earth Systems Thinking (EST)

The geosciences encompass a broad range of disciplines and subdisciplines, all surrounding the study of the Earth. Phenomenographic work by Stokes (2011) demonstrated that the most advanced and complex conceptions of the geosciences involve the conceptions of interacting systems and the relationship between Earth and society. Thus, increasing interest has been paid to the development of EST in geoscience courses. This Earth systems perspective allows for an integrated view of the Earth as interacting parts, different from approaches to geoscience that involve presenting facts about the Earth, ocean, atmosphere, and life without highlighting the interactions between all components (Ireton et al, 1997). This Earth systems approach to education (Mayer, 1991; Ireton et al., 1997) has been particularly well documented in K-12 science education literature (NGSS Lead States, 2013; Orion and Libarkin, 2014; The College Board, 2016) and geoscience workforce expertise (U.S. Bureau of Labor Statistics, 2015).

Naturally, EST is predicated on the learner's systems thinking abilities (Orion and Libarkin, 2014). Systems thinking and its development are not unique to the geosciences and have been explored by a wide range of fields, highlighting the transdisciplinary nature of systems thinking (Trujilo & Long, 2018). The idea of EST, however, goes beyond pure systems thinking and considers the role that humans play within the Earth system and the inherent complexities that come along with those interactions. (Manduca and Kastens, 2012; Gosselin et al., 2013; Orion and Libarkin, 2014; InTeGrate Program, 2015; Orr et al., 2016; Kastens & Manduca, 2017). Stillings (2012) synthesized many of ideas about EST, particularly as they relate to complexity and outlined many future pathways for curriculum, instruction, and research on complex Earth systems.

Earth Systems Frameworks

A literature review was conducted by Scherer and colleagues (2017) to determine the current state of the study of learning and teaching the Earth system in geoscience education research. This work built on foundations laid down by Stillings (2012) and identified four conceptual frameworks (capitalized for clarity) relating to complex systems within the geoscience education research literature: Earth Systems Perspective, Earth Systems Thinking Skills, Complexity Sciences, and Authentic Complex Earth and Environmental Systems (figure 1). The Earth Systems Perspective Framework focuses on the interactions of the four major spheres of the Earth system (lithosphere: solid Earth, biosphere: life, atmosphere: gaseous envelope surrounding the Earth, and hydrosphere: water and ice) and their complex interconnections. This framework is concerned with the interdisciplinary nature of the Earth system and limits systems thinking to conceptualize the Earth system as a whole and often includes aspects of human interactions and environmental decision-making (Davies, 2006).

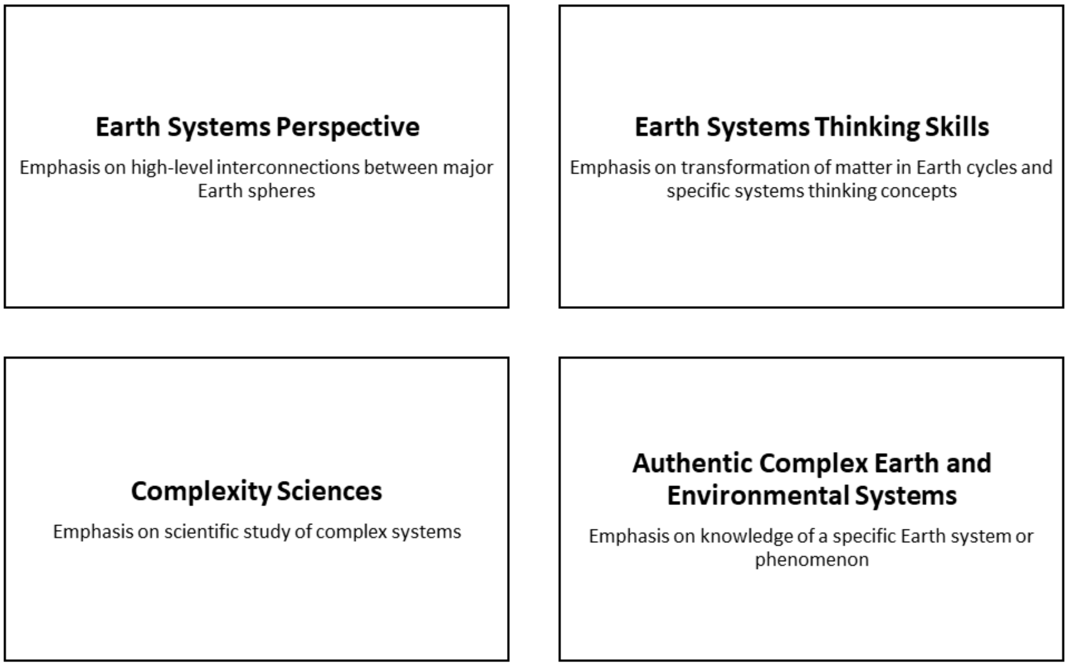


Figure 2.2 EST Frameworks (based on Scherer at al., 2017)

The second framework, Earth Systems Thinking Skills, emphasizes systems thinking skills, particularly as they relate to cyclic and dynamic thinking. Work by Assaraf and Orion (2005) on applying systems thinking to the transformation of matter in Earth cycles—e.g. the water cycle—exemplify this framework. This framework also includes feedback loop identification as well as the understanding of the underlying causes of processes. Thus, this Earth Systems Thinking Skills perspective can be differentiated from the Earth systems perspective by its emphasis on the inclusion of specific systems thinking skills and abilities. While systems thinking is highly emphasized in this framework, heavy use of computer modeling or consideration of complex systems or chaos theory is not included.

The framework of Complexity Sciences largely pulls from the theoretical tradition of the interdisciplinary study of complex systems. This framework is embodied by a wide array of studies that have considered complexity science from the lens of systems dynamics (Shepardson et al., 2014), complex systems theory (including mathematical approaches) (Fichter et al., 2010), and Gaia theory (explaining the environmental conditions of Earth in terms of biological forcing) (Haigh, 2002, 2014). Structure-behavior-function analysis, which emphasizes thinking about how a system works and its function rather than focusing on the components, is also included in this framework and is evidenced in work by Hmelo-Silver and colleagues (2014). Computer modeling work often falls within this framework.

The final framework of Authentic Complex Earth and Environmental Systems pulls its systems ideas from the study of real-world environmental or ecological activities. This framework often makes intentional connections to human activities and environmental decision-making. The systems thinking emphasis is on a real-world environmental system or phenomenon, thus incorporating complexity and looking beyond any one process or single component (Herbert, 2006). This framework, more so than the other three, is highly contextualized. Examples of work done in this area include studies on student reasoning on real-world systems like coastal eutrophication (Sell et al., 2006; McNeal et al., 2008), ecosystem dynamics (Grotzer et al. 2013; Sutter et al, 2018), soil microbial activity (Appel et al., 2014), and socio-hydrologic systems (Gunckel et al., 2012; Sabel et al., 2017; Forbes et al, 2018; Petitt & Forbes, 2019).

These four frameworks, each with strengths and limitations, represent four ways in which educators and researchers are employing EST. These frameworks also offer a more focused way to consider EST instructional practices than the often-nebulous umbrella of complex systems.

EST and Active Learning

The complex nature of Earth systems, regardless of framework, makes it challenging to teach and understand students' learning and conceptual change. Herbert (2006) identified three challenges regarding the reasoning behind complex thinking. The first is the fact that many Earth processes occur at scales—both spatial and temporal—beyond human experience (Dodick and Orion, 2003; Giorgi, 1997.) Secondly, it is difficult to develop accurate conceptual models of complex systems in which several variables are controlling the behavior of the system (Berger, 1998). Lastly, is the tendency of individuals to disregard beyond average data as noise, meaning that individuals try to oversimplify systems as being near or at equilibrium, ignoring important information or data as just "noise." This is problematic as many systems are not at or near equilibrium, and in the case of the Earth system, get farther from equilibrium due to human influence (Goldenfeld and Kadanoff, 1999). Considering this, Herbert emphasizes the importance of inquiry-based and other active learning strategies as necessary in helping students understand complex Earth systems.

Active learning is typically defined as any instructional method that engages students in the learning process (Bonwell and Eison, 1991). Frequently, we contrast active learning practices with more passive traditional lecture; thus, active learning is sometimes considered anything that is not traditional lecture (Prince, 2004). Work by Macdonald and colleagues (2004, 2005) has explored active learning teaching practices within the geoscience community through other iterations of the National Geoscience Faculty Survey. Additional work by Kastens and colleagues (2009) has explored how geoscientists think and learn and the role active-learning teaching practices can play in geoscience classrooms. Active learning, particularly problem-based learning, has been shown to positively affect student achievement, minimize misconceptions, and positively contribute to students' conceptual development (Tandogan & Orhan, 2007). Work by Holder and colleagues (2017) emphasized the importance of problem-solving and its relationship to student conceptualization of the Earth as a system. Problem-solving is a skill that cannot be learned passively and is a central component to understanding complex systems; thus, active learning is essential to helping students experience actual problem-solving tasks and develop this problem-solving expertise. Due to the complexity of the Earth system, we hypothesize that active learning practices may be key in helping students begin to make sense of the interrelationships among the Earth systems and their interdisciplinary nature.

Methods

This study aims to understand the current state of EST-teaching by postsecondary geoscience instructors based on recent survey results of the 2016 administration of the National Geoscience Faculty Survey. The National Geoscience Faculty Survey was initially developed in 2004 as part of On the Cutting Edge, a National Science Foundation (NSF)-funded professional development program for geoscience faculty sponsored by the National Association of Geoscience Teachers (NAGT). In our analysis of the survey results, we took a multipronged approach looking at distinct parts of the survey. We analyzed survey items relating to the teaching of EST, course change, and active learning practices using exploratory and confirmatory factor analytical procedures to find the latent structure of items relating to these overlying themes or constructs. This was used to develop models various teaching practices related to teaching EST-skills, course change, and active-learning practices. Through exploratory factor analysis, we identified items that grouped on overlying constructs. For clarity's sake, these constructs identified through factor analysis will be italicized (E.g. *EST-Teaching* is a construct made up of correlated survey items relating to it, as identified through exploratory factor analysis). We then developed a full structural equation model to understand how these constructs relate to each other.

Participants

The target audience was instructors teaching college-level geoscience courses. The survey was refined with multiple iterations (discussed below), and for this study, the 2016 results were used (Manduca et. al., 2017; MacDonald et. al., 2005; Lally et al, in press). The sampling frame for 2016 was comprised of seven lists of geoscience faculty: the American Geological Institute membership list (obtained with permission from AGI), the SERC Cutting Edge participant list, the Geosciences Two-year College list, the Texas Two-Year College list, the SAGE Two-Year College List, the SERC Cutting Edge Early Career List, and a list of atmospheric science faculty. After removing 2,116 duplicates and removing 81 names without email addresses, the total number of eligible individuals was 10,910. The survey was piloted in September 2016 with a sample of 200 individuals randomly selected from the survey sampling frame. A total of 33 individuals completed at least one question of the pilot survey. Based on the results of the pilot survey, a few minor changes were made to the final survey. As none of these changes were sufficient enough to alter the meaning or the order of the questions, the results of the completed 33 surveys were included in the data set.

The survey was conducted with the remaining sample of 10,710 individuals between October 19 and November 6, 2016. Individuals were contacted up to four times until they took the survey. 1,296 emails were returned as bad or invalid. 27.3% or 2,615 of the 9,596 eligible individuals completed the survey. Of these participants, 60.9% reported having a geology or geophysics disciplinary focus, 8.0% reported an oceanography or marine science disciplinary focus, 9.1% identified an atmospheric science or meteorology disciplinary focus, 8.9% reported a geoscience education or science education disciplinary focus, and 13.0% indicated some other disciplinary focus. 82.2% of respondents reported a Ph.D. or doctorate as their highest completed degree level, while 18.8% indicated that a master’s degree was their highest degree level. In terms of courses taught, 2,290 participants reported teaching undergraduate classes and 123 reported teaching graduate classes. Regarding undergraduate geoscience courses, 539 participants reported teaching introductory courses targeted towards a general audience, 570 reported teaching a major (non-introductory) course, and 1,053 reported teaching introductory courses geared primarily towards majors. This study looked at all instructors, regardless of faculty type, university type, education level, or type/ level of courses taught.

There was some slight response bias, as survey respondents were more likely to be tenured or tenure-track faculty rather than instructors, lecturers, adjunct faculty, or other faculty types (28% of contacted professors, associate professors, and assistant professors responded to the survey versus 21% of contacted instructors, lecturers, adjuncts, and others (Chi-square=33.38, df=1, *p*<.001)). Survey respondents were also less likely to teach at research and/or doctoral institutions and more likely to teach at master's, baccalaureate, two-year colleges, and other institution types (23% of contacted faculty from research and/or doctoral institutions responded to the survey, 28% of contacted faculty from the other institution types responded (Chi-square= 36.64, df=1, *p*<.001)).

Materials

On the Cutting Edge developed the National Geoscience Faculty Survey in 2004, 2009, and 2012 and On the Cutting Edge Leadership modified it in 2009 and 2012. NAGT conducted this national survey in 2004, 2009, 2012, and 2016 (Manduca et. al., 2017; MacDonald et. al., 2005). The instrument was initially developed in 2003 and was modified in 2009 based on the results of the 2004 administration. Revisions to the survey took place after each iteration, and revisions for the 2016 survey were developed by leadership from On the Cutting Edge, InTeGrate, SAGE 2YC, and NAGT with expertise from Greenseid Consulting Group, LLC. and Professional Data Analysts, Inc.

(https://serc.carleton.edu/dev/NAGTWorkshops/CE_geo_survey/index.html). The survey instruments for all administrations can be viewed from the On the Cutting Edge Evaluation Summary web page: https://serc.carleton.edu/NAGT Workshops/ about/evaluation.html. The survey broadly explores three questions: 1) How are faculty teaching undergraduate courses?, 2) How do faculty learn about the content and methods that they use in their teaching?, and 3) How do faculty share with their colleagues what they learn about teaching? The survey follows a similar structure across years and has three parts.

1. The first section consists of demographic questions about education and experience teaching, disciplinary focus, and position and teaching responsibilities
2. The second section asks respondents to self-report about specific courses they have taught in the past two years, the design of these courses, and the teaching methods, strategies, content, and assessment approaches they used in their implementation of the course. It is from this section that we conducted exploratory factor analyses to understand EST-teaching strategies as well as active learning teaching styles.
3. The third section asks questions about how participants learned content and methods as well as information about any changes that were made to a course in the past year. We derived information on incorporating EST-course change elements from this section.

The survey consisted of 209 questions with a median completion time of 14.4 minutes. Respondents answered questions about 1) disciplinary focus, teaching background, and institution; 2) introductory level course teaching strategies; 3) major and minor teaching; 4) learning new teaching methods, active learning strategies included, course changes; 5) communication within the geosciences community and their reasons for attending teaching workshops; 6) use of online resources, articles published, and conference presentations (MacDonald, et al., 2005). Survey questions included a variety of types of items with variant response options, including open response, yes/no, and frequency responses.

Structural equation modeling as a tool for assessing survey results

Survey results are an excellent tool to gather a wide variety of information from a wide swath of a desired cross-section. Survey data can often be vast, and there are many tools with which to analyze the data. One way to examine the current state of teaching in the geosciences is to utilize structural equation modeling to make sense of the latent or underlying structure of the survey results. This approach also makes apparent the complex relationships between responses to survey items and allows researchers to identify direct and indirect relationships between various items and the overlying constructs they represent. Structural equation modeling (SEM) encompasses a variety of techniques that allow researchers to model the relationships among both observed and latent (unobserved) variables. These unobserved variables are often referred to as constructs (Putuch & Stevens, 2016). This methodology takes a confirmatory, or hypothesis-testing, approach to understanding causal processes through a series of regression, or structural, equations. These structural relations are also modeled pictorially in order to provide a clearer conceptualization of the theory that is being studied (Byrne, 2016).

Though based on regression, SEM has the advantage of allowing simultaneous analysis of all the variables in a model instead of separately (Fornell, 1985; Chin, 1998). Due to the large number of participants in this survey (N=2,615), SEM can be used to understand the relationship between participant responses to various items and in this case to quantify and examine how EST-teaching practices are manifesting themselves in current practice by a broad swath of geoscience instructors. SEM also allows us to examine what other latent variables can be gleaned from survey responses and how those variables influence EST-teaching.

Statistical Analysis

The statistical software suite used to analyze the data was the Statistical Package for Social Science (SPSS) Statistic 23 and SPSS AMOS 23. The data were used to develop understanding surrounding four models: EST-teaching, teaching changes, active learning, and the relationship between the previous three models. The whole survey data were randomly split in half using SPSS to use one-half for exploratory factor analyses and the other for confirmatory factor analyses in order to establish cross-validation. The items that were used in the exploratory factor analyses were then analyzed using Little's MCAR Test, which found that missing data could be treated as missing completely at random (*p*=.255). Then an expectation-maximization was used to impute missing data. For the confirmatory factor analyses, a Full Information Maximum Likelihood (FIML) in AMOS was used to create unbiased estimates of missing data. For all exploratory factor analyses, a combination of Kaiser's criterion and a scree analysis was used to determine the number of factors. Because the factors were expected to be correlated, an oblique (direct oblimin) rotation was used in all cases. Criterion pattern loading of .30 or higher was used to determine which items were loading onto which factors for all exploratory factor analyses (Byrne, 2016).

Initially, we identified items related to EST. The survey included nine items written to address system thinking skills (Table 1), which we analyzed in a previous study (Lally et al., in press). However, based on the work of Scherer and colleagues (2017), it was clear that these only addressed the frameworks of Earth System Thinking Skills and Complexity Sciences. Therefore, we examined the survey for other items that may include items relating to the other frameworks (Earth System Perspective and Authentic and Complex Earth and Environmental Systems, specifically including human interactions with the Earth system and interdisciplinary thinking). Items related to quantitative reasoning and data analysis were also included in order to align with the Complexity Science Framework of EST (Scherer et al., 2017). After this inspection of the survey, the items in Table 2 were incorporated into the exploratory factor analysis. These 21 items were shown to have reasonable internal consistency with a Cronbach's alpha of .77. All items were included with the assumption that they would not be utilized if they did not load on a factor. For all exploratory factor analyses, a maximum likelihood analysis was used.

Table 2.1 Items considered for Systems Thinking in the survey. Participants were given the prompt 'are there elements in your course that enable your students to:' for which they could select yes (coded 1) or not selected (coded 0).

Item
Discuss a change that has multiple effects throughout a system
Analyze feedback loops
Make systems visible through causal maps
Explore systems behavior using computer models
Build predictive models
Discuss relationship between implications and predictions
Discuss complexity of scale and interactions
Distinguish outcomes of current processes from results of prior history
Describe a system in terms of its parts and relationships

Table 2.2 Additional items considered for the EST-Teaching Exploratory Factor Analysis. Participants were given the prompt ‘In your most recent [introductory/major] course, did your students:’ for which they could select yes (coded 1) or not selected (coded 0).

Item
Collect their own data and analyze them to solve a problem
Address a problem of national or global interest:
Work on a problem of interest to the local community
Address a problem that required bringing together geoscience knowledge with knowledge from another discipline
Work on a community-inspired research or service project
Address environmental justice issues:
Address uncertainty, non-uniqueness, and ambiguity when interpreting data
Recognize distinctions among data sources (e.g. direct, indirect, and proxy)
Describe quantitative evidence in support of an argument
Evaluate important assumptions in estimation, modeling, or data analysis
Access and integrate information from different sources
None of the above

After a latent structure was hypothesized using exploratory factor analysis, we imposed this structure on the second half of the data in a confirmatory factor analysis using AMOS. We examined this for model fit and significance of all paths ($p<.05$). In all confirmatory factor analyses, Unit Loading Identification (ULI) was used to ensure that the model was identified. Throughout fit analyses, the following fit indices were used: ratio of χ^2 to degrees of freedom, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). We also used the Akaike Information Criterion (AIC) to compare initial models with pruned models if pruning was necessary. Table 3 summarizes the various fit statistics employed.

Table 2.3 Summary of fit indices used in this study. Based on Putuch & Stevens (2016) and Byrne (2016).

Fit Index	Type of Index	What is measures	Rules of Fit
χ^2	Absolute Fit Index	Determines if the model is a good fit based on comparison to an "empty" model	Ratio of χ^2 to degrees of freedom should be less than 4
Comparative Fit Index (CFI)	Comparative fit with hypothetical model	Determines if the model fits the data by comparing data to the hypothesized model Adjusts for sample size and number of indicators	>0.90 = Acceptable >0.95 = Good
Root Mean Square Error of Approximation (RMSEA)	Parsimony-corrected	Determines if adding a parameter is "worth" the cost of parsimony by penalizing fit for models with more parameters	<0.05 to .006= good 0.06 to 0.08 = acceptable 0.08 to 0.10=mediocre >0.10= unacceptable Entire confidence interval is reported and should fall within range for fit to be good
Akaike Information Criterion (AIC)	Purely Comparative Fit Index	Determines if one model fits the data better than the other	No set scale For two models estimated from the same dataset, the lower value is preferred.

We used a similar procedure with items relating to changes in the course (Table 4). The survey did not include specific items targeting EST-teaching changes; therefore, we identified items related to course changes connected with Scherer and colleagues' (2017) EST frameworks. These items had a Cronbach's alpha of .54, indicating low internal consistency. These items were included in an exploratory factor analysis with the assumption that we would eliminate items that did not load from the confirmatory factor analysis. We analyzed the confirmatory factor analysis for the significance of all paths as well as the model fit statistics discussed earlier.

Table 2.4 Items considered for the EST-Teaching Change Exploratory Factor Analysis. Participants were given the prompt 'Which of the following content changes did you make in your [introductory/major] course in the past two years?' for which they could select yes (coded 1) or not selected (coded 0).

Item
Included recent geological events covered in the general media
Added new content area
Increased emphasis on environmental issues
Added content linking geoscience to societal issues
Added content drawn from another discipline
Increased emphasis on systems thinking
Increased focus on quantitative skills
Increased focus on communication skills

Previous work by Manduca and colleagues (2017) identified a 3-factor structure that characterized learning profiles as active learning, active lecture, and traditional lecture (Table 5). Since this has already been established in the literature, we did not conduct an exploratory factor analysis of these constructs. Rather, we imposed this structure on the 2016 data in this study in a confirmatory factor analysis. The traditional lecture was not used as only two inversely related items loaded on it. Additionally, it is not related to active learning, and thus not related to this study. We analyzed the confirmatory factor analysis for the significance of all paths as well as the fit statistics discussed previously. These items had a Cronbach's alpha of 0.49.

Table 2.5. Rotated (Orthogonal) factor matrix for items assessing teaching strategies during lecture (Manduca et al., 2017). Items that loaded on a factor are bolded.

	Factor		
	Active Learning	Active Lecture	Traditional Lecture
Traditional lecture (16.1)	-0.165	0.052	**-0.347**
Lecture with demonstration (16.2)	0.106	**0.257**	-0.151
Lecture in which questions posed by instructor are answered by individual students (16.3)	0.074	**0.492**	-0.043
Lecture in which questions are posed by instructor are answered simultaneously by the entire class (16.4)	0.085	**0.441**	0.138
Small group discussion or think-pair-share (16.5)	0.304	0.398	**0.431**
Whole group discussions (16.7)	**0.553**	0.138	0.183
In-class exercises (16.7)	**0.725**	0.182	0.153

Upon the completion of the confirmatory factor analyses, we developed a full structural model using the three previously mentioned measurement models. To understand model fit, an initial model was developed using active learning as a predictor for EST-Teaching. We assessed this model for the fit and significance of paths. We then added EST-Teaching Changes to the model to analyze its relationship to the other two latent variables. Other demographic information from the survey, such as teaching experience and percent time spent active learning, was also included in the structural model to evaluate the best fit. Throughout the process, we analyzed the model to see how to modify it to improve fit. Throughout fit analyses, we used the same fit indices as for the confirmatory factor analyses: ratio of χ^2 to degrees of freedom, Comparative Fit Index (CFI), and Root Mean Square Error of Approximation (RMSEA). We also Akaike Information Criterion (AIC) to compare initial models with pruned models.

Results

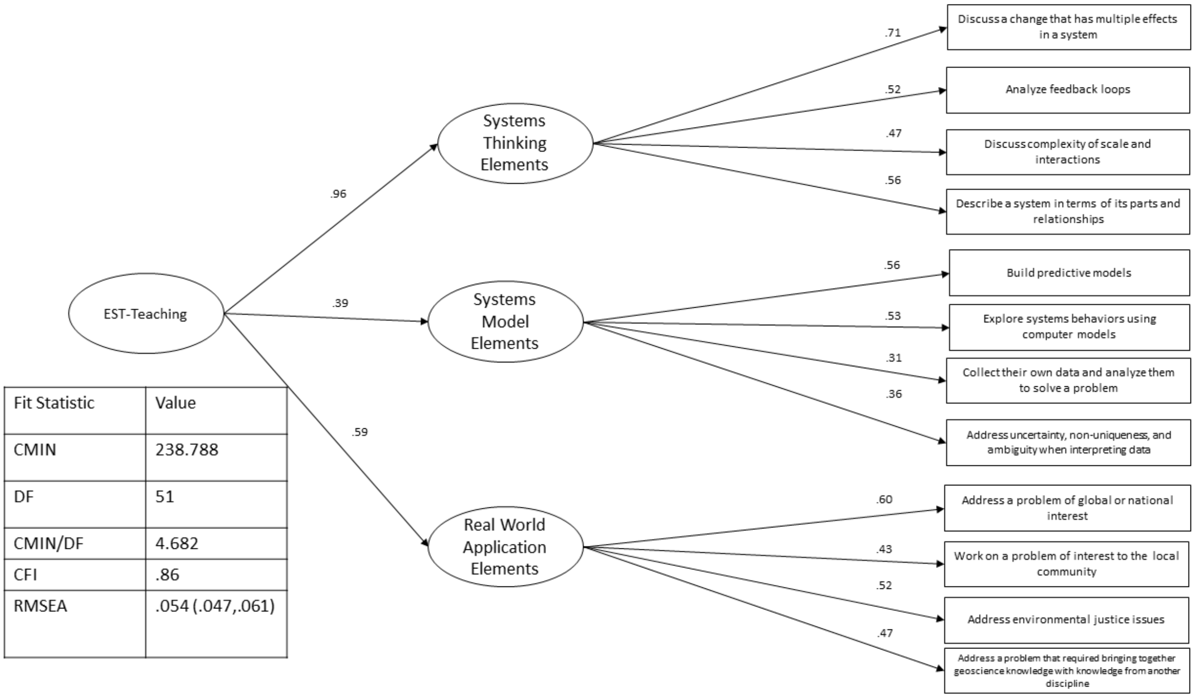
Model 1: EST-Teaching

An exploratory factor analysis was used to explore the factor structure of items relating to EST-teaching practices. This factor analysis was conducted on one-half of the data that was randomly selected, and a three-factor structure was explored (Table 6). Items loading onto Factor 1, which we named *Systems Thinking Elements*, represented classroom teaching practices that included: focusing a change that has multiple effects in a system, analyzing feedback loops, discussing complexity of scale and interactions, and describing a system in terms of its parts and relationships. Items loading onto Factor 2, which we name *Systems Model Elements*, represented classroom teaching practices that included: building predictive models, exploring systems behaviors using computer models, students collecting their own data and analyzing them to solve a problem, addressing uncertainly, non-uniqueness, and ambiguity when interpreting data. Items loading onto Factor 3, which we named *Real World Application Elements*, represented classroom teaching practices that included: addressing a problem of global or national interest, working on a problem of interest to the local community, addressing environmental justice issues, and addressing a problem that required bringing together geoscience knowledge with knowledge from another discipline. A four-factor structure was examined, but it included an unstable fourth factor, with only one item loading on it. In both structures, discussing relationships between implications and predictions failed to load and thus we excluded them. After eliminating the excluded items, these items had a Cronbach's Alpha of .66.

Table 2.6 Pattern Matrix for EST-Teaching Items that were factor analyzed (direct oblimin rotation, rotation converged in 8 iterations). Items that loaded on a factor are bolded. Participants were given the prompt "In the lecture portion of your [introductory/major] course, please indicate how frequently you used the following teaching strategies (1-never, 2-once or twice, 3-several times, 4- weekly, 5-nearly every class)

	Systems Thinking Elements	Systems Model Elements	Real-World Application Elements
Discuss a change that has multiple effects throughout a system	**0.686**	-0.081	0.074
Analyze feedback loops	**0.486**	-0.108	0.066
Make systems visible through causal maps	0.123	0.143	0.041
Explore systems behavior using computer models	0.044	**0.461**	-0.031
Build predictive models	0.013	**0.489**	-0.012
Discuss relationship between implications and predictions	0.282	0.210	0.094
Discuss complexity of scale and interactions	**0.447**	0.212	-0.047
Distinguish outcomes of current processes from a results of prior history	0.251	0.093	0.186
Describe a system in terms of its parts and relationships	**0.624**	-0.007	-0.083
Collect their own data and analyze them to solve a problem	-0.073	**0.450**	-0.045
Address a problem of national and global interest	0.075	-0.118	**0.624**
Work on a problem of interest to the local community	-0.078	0.195	**0.453**
Address a problem that required bringing together geoscience knowledge with knowledge from another discipline	0.116	0.095	**0.376**
Address environmental justice issues	-0.005	-0.135	**0.595**
Address uncertainty, non-uniqueness, and ambiguity when interpreting data	0.043	**0.367**	0.135

The second half of the data was used to perform a confirmatory factor analysis for cross validity before putting this measurement model into a full structural model (Figure 2). Since this measurement was used in a full structural model, rather than correlating the factors, it was represented as all making up a broader latent factor of *EST-Teaching*. The confirmatory factor analysis indicated that all loadings were significant. It also indicated acceptable fit with a χ^2 to degrees of freedom relationship of 4.682, a CFI of .86, and an RMSEA of .054 (.047, .061).



Fit Statistic	Value
CMIN	238.788
DF	51
CMIN/DF	4.682
CFI	.86
RMSEA	.054 (.047,.061)

Figure 2.2 Confirmatory factor analysis with standardized loadings and fit statistics for EST-Teaching. Survey items (on the right-hand side) are shown loading on their respective factors. Standardized unit loadings are shown over each path and all are significant. These factors together make up an overall factor of EST- Teaching. See Table 3 for a description of the fit statistics.

Interpretation

This model can be interpreted to mean that instructors engaging in EST- teaching practices are employing three main broad instructional techniques: incorporating systems thinking elements into their content, including modeling and quantitative systems approaches into their content, and bringing in real-world system elements into their content. Based on the lower factor loading, it is clear that the systems model elements are the most distinct from the other instructional practices.

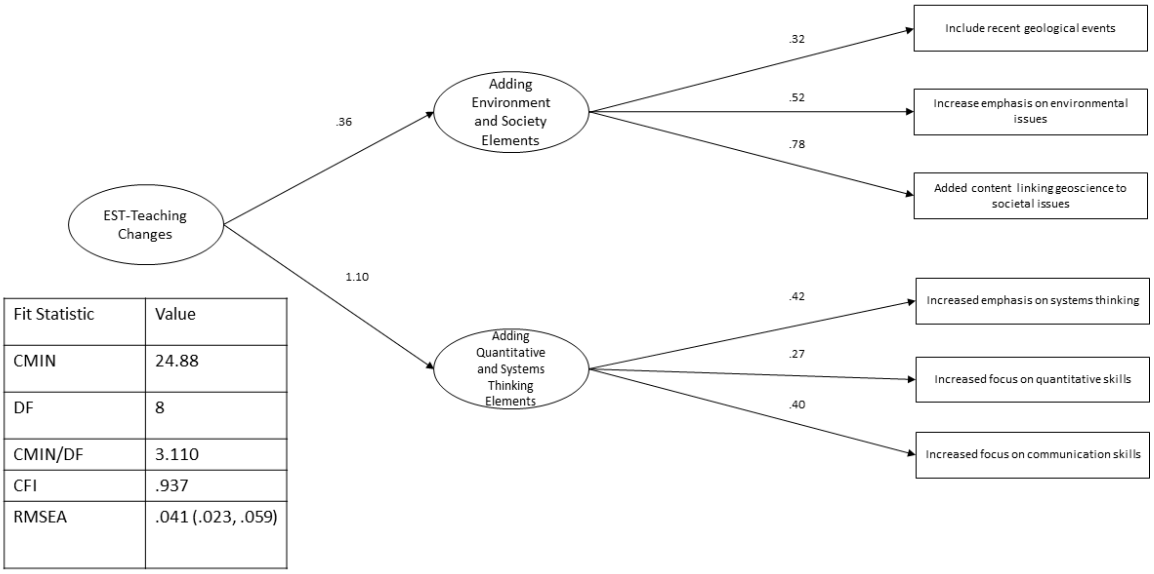
Model 2: EST-Teaching Changes

An exploratory factor analysis was used to explore the factor structure of the survey items relating to changes that instructors recently made to their courses. In this case, we explored a two-factor structure (Table 7). Items loading onto Factor 1, which we named *Adding Environment and Society Elements*, were including recent geological events, increasing emphasis on environmental issues, and adding content linking geoscience to societal issues. Items loading onto Factor 2, which we named as *Adding Quantitative and Systems Thinking Elements*, were increasing emphasis on systems thinking, increasing focus on quantitative skills, and increasing focus on communication skills. We explored a three-factor structure, but it featured an unstable third factor, with only one item loading on it, so a two-factor structure best fits the data. In both structures, several items failed to load including updating content with latest research findings, changing textbooks, and reorganizing topics covered. After excluding items that failed to load, these items had a Cronbach’s alpha of .55.

Table 2.7 Pattern matrix for EST-Teaching Changes exploratory factor analysis (direct oblimin rotation, rotation converged in 10 iterations). Items loading on a particular factor are bolded.

	Adding Environment and Society Elements	Adding Quantitative and Systems Thinking Elements
Updated content with latest research findings	.247	.086
Changed textbook	.071	.102
Reorganized the topics covered	.087	.014
Included recent geological events covered in the general media	**.529**	-.194
Added new content areas	.122	.116
Increased emphasis on environmental issues	**.534**	.123
Added content linking geoscience to societal issues	**.710**	-.020
Added content drawn from another discipline	.240	**.299**
Increased emphasis on systems thinking	.019	**.462**
Increased focus on quantitative skills	-.150	**.529**
Increased focus on communication skills	.140	**.386**

We used the second half of the data to perform a confirmatory factor analysis for cross-validity before putting this measurement model into a full structural model. Since we would be using this measurement model in a full structural model, rather than correlating the factors, they were set to make up a broader latent factor of *EST-Teaching Changes* (Figure 3). The confirmatory factor analysis indicated that all loadings were significant. It also indicated a good model fit with a χ^2 to degrees of freedom relationship of 3.110, and an RMSEA of .041 (.023,.059). A CFI of .937 indicated an acceptable fit.



Fit Statistic	Value
CMIN	24.88
DF	8
CMIN/DF	3.110
CFI	.937
RMSEA	.041 (.023, .059)

Figure 2.3 Confirmatory Factor Analysis for *EST-Teaching Changes* with standardized loadings and fit statistics. Survey items on the right are shown loaded onto their respective factors. Standardized unit loadings are shown over each path and all are significant. These latent factors are shown making up a hierarchical factor of Earth system thinking teaching changes See Table 3 for a description of the fit statistics.

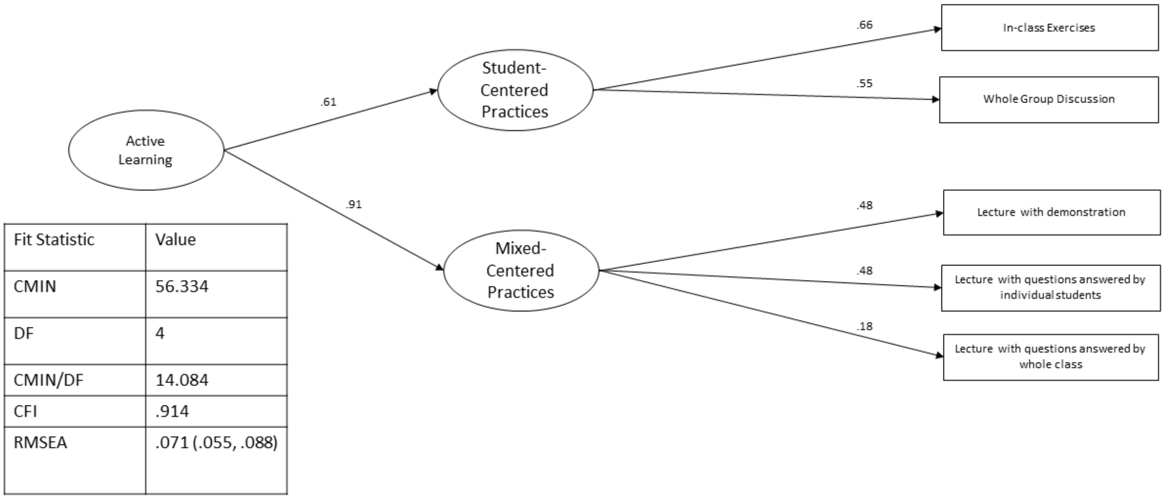
Interpretation

The results can be interpreted to mean that when it comes to making changes in course content, instructors are engaged in two main practices to incorporate more systems thinking: (1) adding elements relating to the environment and society and (2) adding explicit elements relating to using more quantitative data and adding explicit systems thinking elements.

Model 3: Active Learning

We used the full dataset to perform a confirmatory factor analysis for cross validity before putting this measurement model into a full structural model. This was based on an exploratory factor analysis previously completed by Manduca and colleagues (2017) with a past iteration of the survey. Based on this work, which identified the factors as active learning and active lecture, they were renamed in this study as *Student-Centered Practices* and *Mixed-*

Centered Practices (student and instructor-centered) based on the items themselves. Since we were using these in the full structural model, we hypothesized that these would make up a broader construct of *Active Learning* (Figure 4), which relates to the use of active learning teaching practices. The confirmatory factor analysis indicated that all loadings were significant. The ratio of χ^2 to degrees of freedom relationship of 14.084 did not indicate good fit; however, an RMSEA of .071 (.055,.088) and a CFI of .914 indicated acceptable fit. Thus, since it was based on a hypothesized structure from earlier work (Manduca et al,. 2017), it was included in the full structural model.



Fit Statistic	Value
CMIN	56.334
DF	4
CMIN/DF	14.084
CFI	.914
RMSEA	.071 (.055, .088)

Figure 2.4 Confirmatory factor analysis based on work by Manduca et al. (2017) with standardized factor loadings and fit statistics. Survey items are on the right and are shown to load on their respective factors. These latent factors underlie an overarching construct of active learning. Standardized unit loadings are shown over each path and all are significant. See Table 3 for a description of the fit statistics.

Interpretation

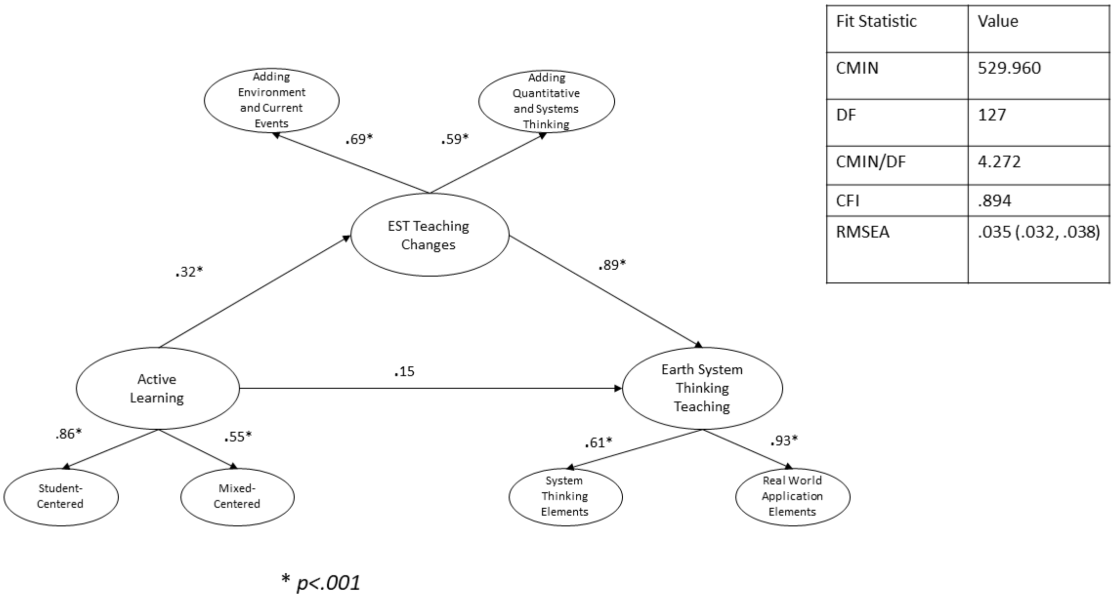
This model, based on the survey results, shows two main active learning approaches employed by geoscience instructors: those that are very student-centered, and those that compromise between student and instructor (or mixed) instruction.

Model 4: Full Structural Model

A full structural model (Figure 5) was developed using the aforementioned measurement models. It was hypothesized that *Active Learning* (model 3) teaching practices would predict both *EST-Teaching Changes* (model 2) as well as *EST-Teaching* (model 1). The full model (model 4), however, showed all paths to be significant (p<.001) except for *Active Learning* as a predictor of *EST-Teaching*. Thus, this model was mediated by *EST-teaching changes*. The initial model's fit was not ideal, with a $\chi 2$ to degrees of freedom ratio of 5.607, a CFI of .812 and an AIC of 1274.987. To address this, the weakest loading of *Systems Model Elements* as an underlying latent variable to *EST-Teaching* was pruned. This decision was made based on the theoretical basis laid out by Scherer et al (2017) of quantitative reasoning and computer modeling largely being an aspect of the Complexity Science Framework of EST and due to its relatively weaker loading. A comparison of fit between the two models can be seen in Table 8. Thus, this model may be more applicable to the other three frameworks or to EST-practices that do not include modeling or extensive use of quantitative data. The new model fit was improved with a $\chi 2$ to degrees of freedom ratio of 4.272, a CFI .894, and an AIC of 655.205. An RMSEA of 035 (.032, .038) also indicated a good fit. In this model, active learning predicted 9% of the variance in Earth system thinking teaching change (r^2=.09) and EST-Teaching Changes accounted for 91% of the variance in EST-teaching (r^2= .91).

Table 2.8 Fit statistics comparing the initial full structural model (Model 4) including the Systems Model construct and the pruned model.

	χ^2	DF	χ^2 / DF	CFI	RMSEA	AIC
Initial Model	1126.987	201	5.607	.812	.042 (.040. .044)	1274.987
Pruned Model	529.960	127	4.272	.894	.035 (.032. .038)	655.205



Fit Statistic	Value
CMIN	529.960
DF	127
CMIN/DF	4.272
CFI	.894
RMSEA	.035 (.032, .038)

Figure 2.5 Full structural equation model with mediation, standardized factor loadings and fit statistics after pruning. These latent variables were derived from the earlier measurement models. Active learning is shown to be a significant predictor of *EST-Teaching Changes* but not *EST-Teaching*. Standardized unit loadings are shown over each path, and paths with an asterisk are significant at the *p*<.001 level. See Table 3 for a description of the fit statistics.

In this model, the path between *Active Learning* and *EST-Teaching* is significant at *p=.02*; however, all other paths are significant at that alpha level of *p<.001*. This noticeably smaller factor loading suggests a mediation effect of *EST-Teaching Changes* on the relationship between *Active Learning* and *EST-Teaching*. A model that excludes *EST-Teaching Changes* and just models the relationship between *Active Learning* and *EST-Teaching* has a much higher factor loading of .45 (*p*<.001), indicating that *EST-Teaching Changes* is a variable mediating that relationship.

Interpretation

Adding the construct of *EST-Teaching Changes* to the model reveals that *Active Learning* is indirectly, rather than directly, related to *EST-Teaching*. This indicates that those instructors who are involved in practices that support active learning are more likely to engage in making changes to the curriculum that involve practices related to EST. In turn, it is the element of course change that is bringing about enhanced EST- teaching practices. It is also important to note that the *Active Learning* construct from work by Manduca et al. (2017), was a superior predictor when compared to self-reported percent time spent involved in active learning when that was incorporated into the model.

Discussion

EST reflects an advanced state of competency in the geosciences (Stokes, 2011) and is a core focus of both standards and outcomes for geoscience teaching and learning (NGSS Lead States, 2013; Orion and Libarkin, 2014; The College Board, 2016) and preparation for the geoscience workforce (U.S. Bureau of Labor Statistics, 2015). However, EST has been shown to be challenging for students given the inherent complexity of Earth systems involving both natural and human dimensions (Berger, 1998; Dodick and Orion, 2003; Giorgi, 1997; Goldenfeld and Kadanoff, 1999; Herbert; 2006; Stillings, 2012). Scherer and colleagues' (2017) literature review on EST in geoscience education foregrounded four conceptual frameworks supported by empirical research. The purpose of this study was to investigate these frameworks for EST and examine the current state of EST-teaching by geoscience faculty members using data from the 2016 National Geoscience Faculty Survey (Lally et al., in press; Manduca et. al., 2017; MacDonald et. al., 2005). Specifically, we sought to better understand how *EST-*

Teaching, innovative teaching (i.e., *Active Learning*; Bonwell and Eison, 1991), and course innovation (*EST-Teaching Changes*) are related to each other. Structural equation modeling including exploratory and confirmatory factor analyses allowed us to discover how survey items related to these constructs, and in turn how these constructs relate to each other.

Model 1: EST-Teaching

We found that three main teaching strategies make up the broader concept of *EST-teaching*: incorporating systems thinking elements, incorporating systems model elements, and incorporating real-world application elements. These general practices relate to Earth systems frameworks identified by Scherer and colleagues (2017) (table 9). The first practice of adding systems thinking elements corresponds to aspects of both the Earth Systems Perspective Framework and the Earth Systems Thinking Skills Perspective, which together involve interconnections between parts of a system and applying systems thinking vocabulary and concepts to a system. The practice of adding system model elements correspond very well to the Complexity Sciences Framework of EST, which includes a strong quantitative and modeling component. Finally, the practice of adding real-world application elements corresponds to the Authentic Complex Earth and Environmental Systems framework as well as to aspects of the Earth Systems Perspective framework. This provides an excellent quantitative analog to the qualitative work that was based on the existing literature on EST-teaching and learning. It also confirms that these frameworks that are grounded in the literature are also being expressed in the self –report of many practicing post-secondary geoscience educators. Based on the structure of the survey, it is not surprising that the Earth Systems Perspective Framework was not distinct, as there were no items that seem to represent it explicitly. It is also interesting to note that of these three general teaching practices identified from the survey, teaching practices involving incorporating systems model elements had the weakest relationship to other EST-instructional practices. This indicates that incorporating systems models into teaching is not as strongly correlated to other EST-teaching practices, meaning that these practices are either done in isolation or not done as frequently as other EST-teaching practices.

Table 2.9 Comparison of constructs identified in the EST-Teaching model and frameworks from Scherer et al. (2017)

Construct from Model	Corresponding Frameworks
Systems Thinking Elements	Earth System Thinking Skills, Earth System Perspective
Systems Model Elements	Complexity Sciences
Real-World Application Elements	Authentic Complex Earth and Environmental Systems, Earth System Perspective

Model 2: EST-Teaching Changes

When we examined items related to teaching changes, we initially considered a variety of changes to curricula to which survey respondents indicated they made to their reported courses in the past two years. Interestingly, only items that are related to EST loaded on broader constructs, while items like changing content sequence or textbook failed to load. This suggests that when making course changes, instructors are intentionally or unintentionally implementing a suite of changes that enhance the teaching of EST, which are not related to more simple changes like changing the sequence of a course or a textbook. This has positive implications as it means instructors are collectively making changes reflecting bringing in both more quantitative and society-based elements into their courses. These changes corresponded to the frameworks previously discussed, with one featuring a clearly quantitative and data-based component and another related to interdisciplinary and environmentally-based teaching (Scherer et al., 2017) (Table 10). Instructors who are adding more environment and society elements to their course are likely making changes in their courses that reflect the Earth Systems Perspective and the Authentic Complex Earth and Environmental Systems framework. Those who are adding quantitative and systems thinking elements are incorporating components of Earth Systems Thinking Skills and Complexity Sciences frameworks. It was surprising that increased focus on communication skills was part of the broader construct of adding quantitative and systems thinking elements, but this may be due to the relationship between communication and problem solving (Holder et al., 2017). This may also indicate that as instructors are adding quantitative and systems thinking elements, perhaps in the context of problem-solving, they are also increasing their emphasis on communication skills.

Table 2.10 Comparison of constructs identified in the EST-Teaching Changes model and frameworks from Scherer et al. (2017)

Construct from Model	Corresponding Frameworks
Adding Environment and Society Elements	Earth Systems Perspective, Authentic Complex Earth and Environmental Systems
Adding Quantitative and Systems Thinking Elements	Earth System Thinking Skills, Complexity Sciences

Model 3: Active Learning

The confirmatory factor analysis on teaching style based on earlier work with past iterations of the survey by Manduca et al. (2017) was confirmed in this study and applied to our current work. While Manduca et al. used the exploratory factor analysis to create teaching profiles (e.g. active lecture, active learning), this study interpreted them as two aspects of active learning, one being highly student-centered and the other mixing instructor and student-centered practices. The *Mixed-Centered* or active lecture construct included lectures with demonstrations or lectures mixed with individual and whole-class questions. The *Student-Centered* construct or active learning construct from Manduca et al. included in-class exercises and whole group discussions. As these did not include traditional lecture, in this study these were deemed as part of the broader construct of *Active Learning* (Bonwell and Eison, 1991). Manduca and colleagues noted that the underlying structure is not what would necessarily be hypothesized (think-pair-share/ small group discussion were not statistically related to the other items) and thus may reflect something about the survey and how participants are responding to it or how they are engaging in those practices. While the full range of active learning strategies may not be truly captured in the survey, the combination of items listed above served as a much better predictor of *EST-Teaching Changes* and *EST-Teaching* than the self-reported percentage of time engaged in active learning, when inserted into the full model (model 4) which was also collected in the survey. This indicates that responses to those survey items better-captured participant behavior than a participant's self-reported time spent engaging in active-learning teaching practices. So while the construct of *Active Learning* as found in this study might not be the ideal measure of actual active-learning practices, it does relate well to this particular model based on the significance of paths and model fit.

Model 4: Full Structural Model

The full structural model revealed that active-learning teaching practices did not necessarily predict *EST-Teaching*, as hypothesized; rather, *EST-Teaching Changes* made by instructors mediated, or controlled, the relationship. Thus, instructors who engage in active learning practices are more likely to make changes to their curriculum that in turn is related to the construct of *EST-Teaching*, meaning that instructors who are making changes are more likely to be engaging in EST-teaching practices. This means that engaging in active-learning practices alone is not sufficient to predict an instructor's engagement in EST-teaching practices. This has implications for the importance of training instructors in both active learning practices and EST.

Geoscience faculty need to have the opportunity to learn about the challenges associated with systems thinking (Herbert, 2006' Stillings, 2012) so that they can make changes to their course. As suggested by Holder and colleagues (2017), faculty are then likely better able to use active learning strategies, like group problem solving, to enhance EST-teaching.

Part of building this model involved putting together the previous three models, thereby making the model more complex. The initial *EST-Teaching* measurement model (model 1) did have significant paths; however, when it was placed in model 4 the model fit was not acceptable. This indicates that though the relationships between the survey items are significant, this model in this circumstance did not fit the data well. To prune the model and improve the fit, we chose to eliminate the construct of *Systems Modeling* from the broader *EST-Teaching*, which had the weakest loading, meaning it was the most distinct of the EST-teaching practices. Upon doing that, the fit statistics improved, meaning that the model now better fits the data. This tells us that teaching using systems modeling is a much more distinct teaching practice than adding systems thinking elements or real-world applications. Additionally, because it did not fit well in the model, this indicates that *Active Learning* and *EST-Teaching Changes* may not currently influence the likelihood of instructors engaging in systems modeling practices or implementing aspects of the associated Complexity Science Framework of EST (Scherer et al., 2017). However, the model does seem to correspond with the frameworks of Earth Systems Perspective, Earth Systems Thinking Skills, and Authentic and Complex Environmental Systems.

The distinctness of the systems modeling in the full model demonstrates a greater need for resources on quantitative reasoning as it relates to systems as well as system modeling resources in the vein of work by Shepardson et al., (2014), Ficheter et al. (2010) and Hemelo-Silver (2014). It appears that it is the quantitative reasoning skills as well as modeling skills that make this construct unique. Additionally, when we considered course changes, the relationship between adding an increased focus on quantitative skills was much weaker than that to add more emphasis on systems thinking to overall course change. This suggests that while instructors are adding systems thinking elements to their courses, they are not complimenting that by adding the quantitative component which may be critical to understanding complex systems during the classroom deployment of system modeling approaches.

The fact that systems modeling is the most distinct of the EST-teaching constructs indicates that it is not employed as frequently in relation to the other practices of including systems thinking elements or real-world elements. It also indicates that using active learning practices and making EST-teaching changes to courses is not as predictive of including systems modeling elements as it is to the other two EST-teaching practices. Thus, the community must continue and expand its discussion on how to best incorporate quantitative reasoning, modeling, and complexity sciences into resources for practicing Earth science educators. Survey respondents included a large number of instructors who taught introductory courses, so it is particularly important to pay special attention to introductory geoscience courses in this discussion. InTeGrate, a geoscience National Science Foundation-funded project, continues to work to address this problem and has assessed faculty and student weaknesses teaching and learning about systems thinking (InTeGrate Program, 2015).

Limitations and Future Research

Limitations of this study include the self-reported nature of the data, which may not be reflective of what instructors actually do. The voluntary nature of the survey, like with all surveys, is a natural limitation. These factors may have resulted in participants who are skewed toward those with an interest in teaching and learning in the geosciences. As with any survey, there may be issues with the instrument itself or fatigue associated with the length. Some item groupings did not have good internal consistency, as evidenced by lower Cronbach Alpha values. This may indicate that the results of some item groupings may not be as reliable as they were intended. There also is some response bias, as survey respondents were more likely to be tenured or tenure-track faculty rather than instructors, lecturers, adjunct faculty, or other faculty types.

Survey respondents were also less likely to teach at research and/or doctoral institutions and more likely to teach at master's, baccalaureate, two-year colleges, and other institution types. Additionally, individuals with disciplinary focuses in oceanography (9.3%) and atmospheric science (9.5%) were far less represented than those in geology or associated fields (81.2%). Thus, while the sample size is large, it is likely not truly representative of all geoscience instructors. It is also notable that the majority of instructors indicated teaching introductory courses, primarily to majors, while only about a quarter of respondents indicated that they taught upper-level geoscience courses. This means that the sample is slightly biased towards EST-teaching practices in introductory courses. It is also important to note that the models developed in this study are just models, and while they fit the data, they may not necessarily be true. As Rasch (1960) and Tukey (1963) note, no model is perfect, but it is the insight that they give researchers that is valuable. It is of similar value to note that the *Teaching Changes* items specifically asked instructors about changes made within the past two years. It is possible that some instructors may have made changes to their courses related to EST before the period indicated by the survey prompt. In this case, we may have lost the data of a small group of participants who may have already made these changes.

In terms of future directions for research, an obvious step is to use these measurements and structural models on future iterations of these surveys to ensure that they hold up to scrutiny.

While this work confirms many of the qualitative findings found in the literature by Scherer and colleagues (2017), it is important that this work be taken into the field and observed both qualitative and quantitatively. The qualitative approach calls for work examining EST-Teaching and its relationship to teaching approach in the classroom as well as aims to understand instructors' conceptions and understanding of EST. Quantitative work calls for the development of instruments that can help researchers measure learner development of systems thinking (in general or within each of the four frameworks) to complement and expand on existing instruments (Jordan et al., 2014; Grohs et al., 2018). This step will be essential in measuring the effectiveness of teaching EST, both by the nature of courses and contexts that may be the most effective and specific teaching strategies used. Ben-Zvi Assaraf and Orion (2004) noted the difficulty in conducting EST-research due to the need to evaluate the strength and weaknesses of each research tool concerning what skills each tool is actually measuring. Thus, it is difficult to assess EST skills across many courses or instructors without a more streamlined instrument. There is a need for more classroom-based studies in which EST-teaching practices are observed and coupled with student learning outcomes to complement and build upon previous studies based on self-report.

Additional structural equation modeling could also be completed to understand if the model is variant across a variety of groups (discipline, years teaching, training, etc.). Work to understand differences in EST-teaching practices between introductory and upper-level courses would also be worthwhile. Additionally, future work should focus on the survey to understand why individuals that make teaching changes to their courses seem to be making these changes towards more of an Earth systems approach, and if this is a result of nation-wide influencers such as InTeGrate, NGSS, and Quantitative Reasoning across the Curriculum. This is a crucial step in influencing professional development around EST-teaching and learning. Structural equation modeling is an excellent tool in analyzing large datasets, and this survey certainly contains many more constructs waiting to be explored.

Conclusion

This study, which used the National Geoscience Faculty Survey, found that the current state of Earth systems teaching in American colleges and universities is largely consistent with the qualitative Earth systems frameworks proposed by Scherer and colleagues (2017). Current innovation in geoscience teaching also tends to revolve around major frameworks of EST and is predictive of individuals incorporating EST elements into their courses. These results also suggest that the Complexity Science Framework of EST (involving computer modeling and quantitative data) is much more distinct than the other three frameworks. This means that instructors are frequently not engaging in teaching practices that involve modeling or complexity sciences in conjunction with other EST-teaching practices. Active learning teaching practices do share a relationship with EST-teaching practice; however, it is mediated through course changes instructors are making. Thus, individuals who engage in active learning are more likely to engage in changes to their curriculum that incorporate more EST-elements rather than just naturally including more EST-teaching practices in their courses. This also indicates that active-learning practices alone are not sufficient to bring about EST-teaching practices. This study gives us a snapshot of the current state of EST-teaching in higher education and suggests some interesting relationships between active learning, course changes, and EST instruction and implications for additional professional development opportunities.

This study also draws attention to the importance of geoscience educators receiving resources and training in EST, as active learning practices themselves are not enough to ensure that instructors are explicitly teaching EST. Thus, continued professional development in not only active learning practices but also in systems thinking and its teaching is essential. This is particularly true in the sense of pedagogical content knowledge and Earth systems content knowledge; which points to the significance of work being done by NAGT, On the Cutting Edge professional development programs, InTeGrate workshops, and others in bringing professional development to a wide range of geoscience educators. Professional development and training in the complexity sciences, quantitative skills, and systems modeling is especially essential, as in our model the construct of *Systems Modeling* is the most distinct and least related to the other constructs that make up *EST-Teaching*. As evidenced by the large number of participants who reported teaching introductory courses, it is particularly important to consider and research how we enhance systems modeling practices in these courses. As instructors continue to implement EST-teaching strategies in their courses, researchers must begin to take these findings from the literature and survey and begin exploring them in practice.

Acknowledgments

This work was supported by the National Science Foundation's Division of Undergraduate Education (DUE) under award #s 0127310, 0127141, 0127257, 0127018, 0618482, 0618725, 0618533, 1022680, 1022776, 1022844, 1022910, 1125331, 1525593, 1524605, 1524623, and 1524800. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

We acknowledge and thank the following individuals for other contributions that made this study possible:

- Raymond Y. Chu, Julius Dollison, and Roman Czujko of the Statistical Research Center of the American Institute of Physics helped develop the 2004 and 2009 survey instruments, administer these surveys, and did the initial analysis of the results.

- Diane Ebert-May and colleagues in biology provided an unpublished copy of a similar survey developed for biology from which the 2004 leadership team benefited.
- Staff including Nick Claudy and Christopher Keane from the American Geological Institute worked through permissions to provide the initial set of geoscience faculty email addresses.
- John McLaughlin, the *On the Cutting Edge* external evaluator, for contributions to the development of the 2004 and 2009 survey instruments.
- Experts from Professional Data Analysts, Inc., including Michael Luxenberg, Becky Lien, Eric Graalum, and Mao Thao for work on the analysis of the 2009 survey and development and analysis of the 2012 and 2016 survey.
- Lija Greenseid, Greenseid Consulting Group, LLC who facilitated survey design and implementation and contributed to interpretation of data analysis (2012 and 2016).
- *On the Cutting Edge* PIs: R. Heather Macdonald, Cathryn A. Manduca, David W. Mogk, Barbara J. Tewksbury, Rachel Beane, David McConnell, Katryn Wiese, and Michael Wysession
- Joni Lakin, Kim Kastens, Rachel Beane, Kathleen Quadorkus Fisher, and Professional Data Analysts, Inc. for reviews and suggestions that strengthened this article.
- The geoscience survey working group and assistance from Greenseid Consulting Group, LLC. and Professional Data Analysts for support and guidance in working with the survey data.
- Leadership from National Association of Geoscience Teachers (NAGT), *On the Cutting Edge,* InTeGrate, and SAGE 2YC for their work in administering the 2016 geoscience faculty survey.

- The editors and reviewers of *Geosphere* for their constructive and thoughtful feedback to strengthen this manuscript.

References

Arnold, R.D., and Wade, J.P. 2015, A definition of systems thinking: A systems approach. *Procedia Computer Science*, *44*, 669-678.

Assaraf, O. B. Z., and Orion, N., 2005,. Development of system thinking skills in the context of earth system education, Journal of research in science teaching, *42*(5), p. 518-560.

Berger, R, 1998, Understanding science: Why causes are not enough, Philosophy of Science, v. 65, no. 2, p. 306–333.

Bonwell, C.C., and Eison, J.A, 1991, "Active learning: Creating excitement in the classroom," ASHEERIC Higher Education Report No.1, George Washington University, Washington, DC.

Byrne, B. M., 2016 Structural equation modeling with Amos: Basic concepts: applications, and programming, (6th ed.), New York, NY: Routledge, 437 p.

Chin, W., 1998, The partial least squares approach for structural equation modeling. In G. A. Marcoulides (Ed.), Modern methods for business research, Mahwah, NJ: Lawrence Erlbaum Associates, p. 295-336.

Davies, C. P.,2006, Implementing earth systems science curriculum: Evaluating the integration of urban environments for an urban audience, Journal of Geoscience Education, 54(3), p. 364-373.

Dodick, J., and Orion, N., 2003, Measuring student understanding of geolog ical time. *Science and Education*, v. 87, no. 5, p. 708–731,

Fichter, L. S., Pyle, E. J., and Whitmeyer, S. J., 2010, Strategies and rubrics for teaching chaos and complex systems theories as elaborating, self-organizing, and fractionating evolutionary systems, Journal of Geoscience Education, *58*(2), p.65-85.

Forbes, C.T., Brozovic, N., Franz, T., Lally, D., and Petitt, D., 2018, Water in Society: An interdisciplinary course to support undergraduate students' water literacy, Journal of College Science Teaching, 48(1), 36-42.

Fornell, C.,1985, A second generation of multivariate analysis: Classification of methods and implications for marketing research.

Giorgi, F.A.R., 1997, Representation of heterogeneity effects in earth system modeling: Experience from land surface modeling, Reviews of Geophysics, v. 35, no. 4, p. 413.

Goldenfeld, N., and Kadanoff, L.P, 1999, Simple lessons from complexity, Science, v. 284, no. 5411, p. 87–89.

Gosselin, D.C., Manduca, C., Bralower, T., and Mogk, D., 2013, Transforming the teaching of geoscience and sustainability, Eos Trans AGU Publications, Vol. 94, no. 25, pp. 221-222.

Grohs, J. R., Kirk, G. R., Soledad, M. M., and Knight, D. B., 2018, Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems, Thinking Skills and Creativity, 28, 110-130

Grotzer, T. A., Kamarainen, A. M., Tutwiler, M. S., Metcalf, S., and Dede, C, 2013,Learning to reason about ecosystems dynamics over time: The challenges of an event-based causal focus, *BioScience*, *63*(4), p. 288-296.

Gunckel, K. L., Covitt, B. A., Salinas, I., and Anderson, C. W., 2012, A learning progression for water in socio-ecological systems, Journal of Research in Science Teaching, *49*(7), p. 843-868.

Haigh, M. J, 2001, Constructing Gaia: Using journals to foster reflective learning, Journal of Geography in Higher Education, *25*(2), p. 167-189.

Haigh, M., 2014, Gaia:“thinking like a planet” as transformative learning, Journal of Geography in Higher Education, 38(1), p. 49-68.

Herbert, B. E., 2006, Student understanding of complex earth systems., Special Papers Geologic Society of America, 413, 95.

Hmelo-Silver, C, Eberback, C., and Jordan, R., 2014, Technology-supported inquiry for learning about aquatic ecosystems. Eurasia Journal of Mathematics, Science & Technology Education, 10, p. 405-413

Holder, L. N., Scherer, H. H., and Herbert, B. E., 2017, Student learning of complex Earth systems: A model to guide development of student expertise in problem-solving, Journal of Geoscience Education, 65(4), p. 490-505.

InTeGrate Program, 2015, InTeGrate pedagogy: Themes and strategies for teaching about geoscience and societal issues, Available at http://sec.carleton.edu/integrate/about/pedagogy_theme.html (accessed 18 April 2018).

Ireton, F. W., Manduca, C. A., and Mogk, D. W.,1997, Towards a coherent plan for undergraduate earth science education: a systems approach, Journal of College Science Teaching, 26(5), p. 304.

Jordan, R. C., Sorensen, A. E., and Hmelo-Silver, C.,2014, A conceptual representation to support ecological systems learning, Natural Sciences Education, 43(1), 141-146.

Lally, D., Forbes, C.T., McNeal, K., and Soltis, N., in press, National Geoscience Faculty Survey 2016: Prevalence of systems thinking and scientific modeling learning opportunities, Journal of Geoscience Education.

Kastens, K., Manduca, C.A., Cervato, C., Frodeman, R., Goodwin, C., Lieben, L.S., Mogk, D.W., Spangler, T.C., Stillings, N.A., and Titus, S., 2009, How geoscientists think and learn, *Eos Trans, AGU*, Vol. 90, no. 31, p. 265.

Kastens, K. A., and Manduca, C.A., 2017, Using systems thinking in the design, implementation and evaluation of complex educational innovations, with examples from the InTeGrate Project, *Journal of Geoscience Education*, Vol. 65, no. 3, August 2017, p. 219-230.

Macdonald, R.H., Manduca, C.A., Mogk, D.W., and Tewksbury, B.J., 2004, On the cutting edge: Improving learning by enhancing teaching in the geosciences, Invention and Impact: Building Excellence in Undergraduate Science, Technology, Engineering, and Mathematics (STEM) Education, Vol. 381.

Macdonald, R. H., Manduca, C. A., Mogk, D. W., and Tewksbury, B. J.,2005, Teaching methods in undergraduate geoscience courses: Results of the 2004 On the Cutting Edge survey of US faculty, Journal of Geoscience Education, 53(3), p. 237-252.

Manduca, C. A., Iverson, E. R., Luxenberg, M., Macdonald, R. H., McConnell, D. A., Mogk, D. W., and Tewksbury, B. J.,2017, Improving undergraduate STEM education: The efficacy of discipline-based professional development, Science Advances, 3(2), e1600193.

Manduca, C., and Kastens, K.,2012, Mapping the domain of complex Earth systems in the geosciences, In Kastens, K., and Manduca c., eds., Earth and mind II: A synthesis of research on thinking and learning in the geosciences, Boulder, CO: Geological Society of America, p. 91-96

Mayer, V. J., 1991, Framework for Earth systems education, *Science Activities: Classroom Projects and Curriculum Ideas*, *28*(1), p. 8-9.

McNeal, K. S., Miller, H. R., and Herbert, B. E., 2008, The effect of using inquiry and multiple representations on introductory geology students' conceptual model development of coastal eutrophication, Journal of Geoscience Education, 56(3), p. 201-211.

NGSS Lead States, 2013, Next generation science standards: For states, by states, Available at http://www.nextgenescience.org/next-generation-science-standards (accessed 18 April 2018).

Orion, N., and Libarkin, J., 2014, Earth system science education, Handbook of research on science education, 2, p. 481-496.

Orr, C.H., Newman, A., and Manduca, C.A., 2016, InTeGrate: Interdisciplinary teaching of geoscience for a sustainable future, In the Trenches, Vol. 6, no. 3, pp. 3-4.

Petitt, D.N. and Forbes, C.T.,2019, Values use of undergraduate students in socio-hydrological reasoning: A comparative study, Natural Sciences Education, 48(1), 1-12.

Pituch, K. A., and Stevens, J. P, 2016, Applied multivariate statistic for the social sciences (6th ed.), New York, NY: Routledge. 793 p.

Prince, M., 2004, Does active learning work? A review of the research, Journal of engineering education, 93(3), 223-231.

Rasch, G. ,1960, Probalistic models for some intelligence and attainment tests, Copenhagen: The Danish Institute for Educational Research, p. 37-38.

Sabel, J., Vo, T., Alred, A., Dauer, J., and Forbes, C.T.,2017, Undergraduate students' scientifically-informed decision-making about socio-hydrological issues, Journal of College Science Teaching, 46(6), 64-72.

Scherer, H. H., Holder, L., and Herbert, B., 2017, Student learning of complex Earth Systems: Conceptual frameworks of Earth Systems and instructional design, *Journal of Geoscience Education*, *65*(4), p. 473-489.

Sell, K. S., Herbert, B. E., Stuessy, C. L., and Schielack, J., 2006, Supporting student conceptual model development of complex earth systems through the use of multiple representations and inquiry, Journal of Geoscience Education, 54(3), p. 396-407.

Shepardson, D. P., Roychoudhury, A., Hirsch, A., Niyogi, D., and Top, S. M., 2014, When the atmosphere warms it rains and ice melts: Seventh grade students' conceptions of a climate system, Environmental Education Research, 20(3), p. 333-353.

Stillings, N., 2012, Complex systems in the geosciences and in geoscience learning, Earth and mind II: A synthesis of research on thinking and learning in the geosciences: Geological Society of America Special Paper, 486, p. 97-111.

Stokes, A., 2011, A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline, Geological Society of America Special Papers, 474, p. 23-35.

Sutter, A.M., Dauer, J.M., and Forbes, C.T.,2018, Construal level and value-belief norm theories: Implications for undergraduate decision-making on a prairie dog socio-scientific issue, International Journal of Science Education, 40(9), 1058-1075.

Tandogan, R. O., and Orhan, A.,2007, The Effects of Problem-Based Active Learning in Science Education on Students' Academic Achievement, Attitude and Concept Learning, Online Submission, *3*(1), 71-81.

Trujillo, C. M., and Long, T. M., 2018, Document co-citation analysis to enhance transdisciplinary research, Science advances, 4(1), e1701130.

Tukey, J.W., 1963, Mathematics 596- An introduction to the frequency analysis of time series. In D. r. Brillinger (Ed.), The collected works of John W. Tukey, Volume I: Time series, 1949-1964, London: Chapman & Hall (p. 504).

U.S. Bureau of Labor Statistics, 2015, Occupational outlook handbook: Geoscientists, Available at http://www.bls.gov/ooh/life-physical-and social-science/geoscientists.htm (accessed 18 April 2016).

CHAPTER 3: UNDERSTANDING UNDERGRADUATE STUDENT CONCEPTIONS ABOUT BIOGEOCHEMICAL CYCLES AND THE EARTH SYSTEM

Introduction

Biogeochemistry is broadly the study of the chemical, physical, geological, and biological processes and reactions that govern the composition of the Earth system. Biogeochemists largely study the cycling of major chemicals and elements like water, carbon, nitrogen, phosphorous, sulfur, and trace metals. This cycling involves all components of the Earth system (atmosphere, geosphere, biosphere, and hydrosphere) and is heavily impacted by humans (Schlesinger & Bernhardt, 2013). Pollution, in fact, can be defined as any alteration to an existing biogeochemical cycle caused by humans (Jacobson et al., 2000). Thus, biogeochemical cycles are fundamental constructs for studying Earth system science and global change. The fact that these cycles are complex and consist of multiple feedbacks and paths, in addition to being transcendent of any one discipline, also makes them an ideal tool to both teach and assess systems thinking skills in students. Additionally, many students learn these cycles only in middle school or in bits or pieces in various undergraduate classes, meaning that biogeochemistry is likely an area where students are under-informed.

Earth Systems Thinking

Systems thinking, as used in this study, is an understanding of the circular nature of what happens in the physical world, and the ways that individual systems and their components (sub-systems) relate to and interact with each other. It is an understanding that a system is an indivisible whole, and much more than the sum of its parts (Ackoff, 1973). In the geosciences, qualitative work by Stokes (2011) has demonstrated that the highest level of understanding about the physical world is the ability to think about the Earth as a dynamic system. However, previous work has demonstrated that students have conceptual difficulties with systems thinking and understanding the dynamic nature of systems, particularly when the matter in these systems is not observable or readily apparent (McNeal et al., 2014; Orion, 2002; Sibley et al., 2007). Students also have difficulty when the instructional context of the system is not directly relevant to students' past experiences (Wilson, 2006). Thus, a major area of research in the field of geoscience education is the development of Earth systems thinking. As biogeochemical cycles are the fundamental linking construct of the Earth system, it is logical that their understanding is a crucial tool in the development of Earth system thinking and a lens into the teaching and learning of complex systems.

Earth systems thinking skills are also essential for understanding issues of sustainability and the development of solutions for important environmental issues. A report by the National Research Council (2000) outlined eight major challenges facing humanity, one of them being human alterations to major biogeochemical cycles. Other challenges such as biological diversity and ecosystem functioning, climate variability, hydrologic forecasting, infectious disease and the environment, institutions and resource use, land-use dynamics, and reinvention of the use of materials are directly or indirectly related to biogeochemical cycling (NRC, 2000). These environmental issues involve complex Earth systems, which are defined as near-surface Earth systems that exhibit complex spatial characteristics and dynamics (Herbert, 2006). These complex systems come with fundamental challenges to student understanding, including the conceptualization of the natural Earth environment as composed of systems, the characterization and explanation of the complex nature of Earth systems, and the application of conceptual and scientific models of Earth systems to support problem-solving and the development of effective environmental policy (Herbert, 2006; Oreskes et al., 1994).

Many researchers have developed various models for examining and evaluating systems thinking skills among university and high school students. The structure, behavior, and function (SBF) pedagogical model (Hmelo-Silver & Pfeffer, 2004; Libarkin et al., 2005; Hmelo-Silver et al., 2007) focuses on understanding how connections between structures, behaviors, and functions allow systems to work. Ben-Zvi Assaraf and Orion (2005) used this approach in designing educational elements for curricula on biogeochemical cycles. In this same paper, the authors also proposed another model that could be used for hierarchically identifying and characterizing different levels of systems thinking. Qualitative work by Sibley et al. (2007) explored identifying mobile and changeable components within systems using box diagrams to assess students' thinking systems about the rock, water, and carbon cycles, while Raia (2005, 2008) looked at the existence of multiple causes for a single phenomenon in complex Earth systems. This work has inspired recent studies that have examined how dynamic and cyclic thinking are major components of Earth systems thinking (Batzri et al., 2015). More recent work has focused on developing successful interventions for teaching systems thinking skills in high school students (Hmelo-Silver et al., 2017; Jacobson et al., 2017; Lavi et al., 2019; Tripto et al., 2018). Additionally, Yoon and colleagues (2019) have engaged in work towards the development of a learning progression of complex systems understanding for high school students.

With the expansion of work on Earth systems thinking, synthesis of a variety of studies is a major challenge. Scherer and colleagues (2017) reviewed the Earth systems thinking literature and synthesized four major Earth system thinking conceptual frameworks: Earth system perspective, Earth system thinking skills, complexity science, and authentic complex Earth and environmental systems (Figure 1). The Earth system perspective emphasizes the interconnections between the major Earth spheres. Systems thinking abilities relate to conceptualizing the Earth system as a whole. The Earth system thinking skills conceptual framework emphasizes the transformation of matter in Earth cycles and the thinking abilities related to identifying and organizing system components. The complexity science conceptual framework emphasizes the scientific study of complex systems and systems thinking abilities related to recognizing complex system characteristics. Finally, the authentic complex Earth and environmental systems framework emphasizes the knowledge of a specific complex near-surface Earth system or phenomenon and systems thinking abilities related to reasoning about the specific system or phenomenon. Biogeochemical cycles, depending on their application in learning, can fall in any of these conceptual frameworks, and these frameworks can be invaluable tools in accessing student learning and knowledge. Though many exciting models have been developed for characterizing systems thinking, when understanding biogeochemical cycles, it is critical to understand the interdisciplinary science content that underpins these systems concepts. This study aims to gather qualitative information about how university undergraduate students conceptualize these cycles to inform future research on applying existing systems thinking models to these cycles as well as potential work on learning progressions for biogeochemistry and complex Earth systems.

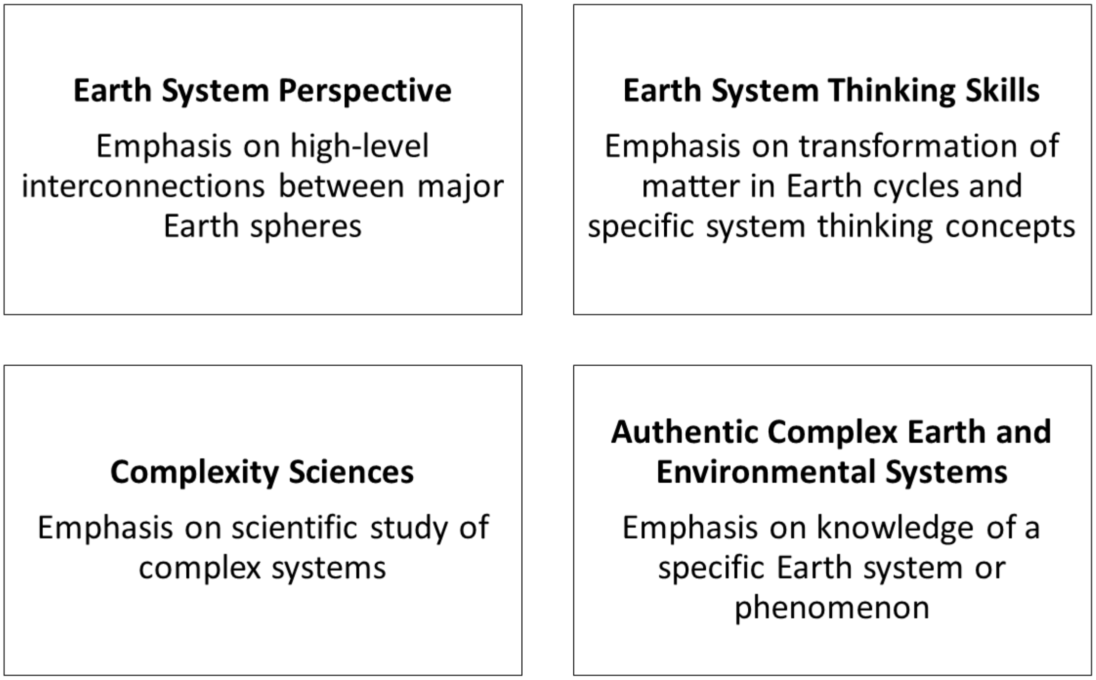


Figure 3.1 The four Earth system thinking frameworks as described by Scherer and colleagues (2017).

Mental Models

Mental models are internal mental representations people develop of complex systems. In contrast, conceptual models are the externally represented tools devised for the understanding or teaching of physical systems (Norman, 1983). A textbook illustration of the carbon cycle is an example of a conceptual model, whereas a student's illustration of the carbon cycle could be considered as a representative of a mental model. It is important to note, that a drawing may not completely capture a given student's mental model, and student self-reported data possess limitations (Stone et al., 2000). Students might think one way but may not express themselves that way for a variety of reasons: not wanting to appear wrong, being unaware of their thinking, or failing to make connections between various mental models. Additionally, not all student mental models are stable over time (Libarkin et al., 2003) and are not always coherent relative to other ideas held by the student (Delaughter et al., 1998, Mark et al., 1999). However, many different methods can be employed to try to make sense of student mental models and alternate conceptions (Libarkin & Kurdziel, 2001). This study aims to take a triangulated approach to examining students' mental models and to probe students' understanding to capture the most complete representation of their mental models as possible.

Student Drawings

Drawings are considered simple research instruments that enable easy comparisons. They also can be completed quickly, and often in a form that is more enjoyable than answering questions (Prokop & Faněoviěová, 2006). They also give researchers a glimpse into a participant's mind, which makes them especially useful in understanding mental models (Thomas & Silk, 1990). The act of creating drawings involves a variety of mental tasks such as recalling verbal and visual information, selecting appropriate information to use, and integrating those elements into a drawing (Van Meter & Garner, 2005). Thus, drawings of phenomena can provide researchers with a rich understanding of a participant's alternate conceptions and mental models (Ainsworth et al., 2011). Drawings are often employed in research with children, but they have also been successfully employed with college students in the field of geoscience and/ or biology education (Sell et al., 2006 Arthurs, 2011, Cardak, 2009, Herrera & Riggs, 2013, Köse, 2008). For example, drawings have been used to capture student conceptions and models of various biogeochemical phenomena. Köse (2009) used drawings from university preservice teachers to understand alternate conceptions relating to photosynthesis and respiration. In terms of elemental cycling, Cardak (2009) used drawings and interviews to understand alternate conceptions related to the water cycle. Sell et al. (2006) used undergraduate student drawings and written reports to understand their conceptions relating to coastal eutrophication and better probe student conceptions of complex Earth systems.

Past Work

Work surrounding student conceptions about biogeochemical cycling has largely been focused on carbon at both the K-12 and college levels. Work by Hogan and Fisherkeller (1996) involved understanding students' thinking about nutrient cycling within ecosystems. This study included interviews with eight urban 5th and 6th graders and focused largely on the biotic processes of nutrient cycling (photosynthesis, decomposition, etc.). Alternate conceptions relating to photosynthesis have been explored in several studies such as Lin and Hu (2003), Köse (2009), Hartley et al. (2011), and Parker et al. (2012), and have revealed several pervasive alternate conceptions relating to conservation of matter, energy flow, and matter transformations as well as student difficulties linking abiotic to biotic factors within these processes.

Work on learning progressions by Mohan and colleagues (2009) explored conceptions of carbon cycling in students from upper elementary grades through high school to develop a multi-year learning progression for carbon cycling in socio-ecological systems. This learning progression identified four levels of achievement. The first level viewed carbon sources as enablers of life processes and combustion rather than sources of matter transformed by those processes. At level two, students' traced matter in terms of materials changed by unknown or hidden mechanisms, and at level three these mechanisms were recognized as chemical processes. At the most advanced level, level four, students used chemical models to trace matter through hierarchically organized systems, connecting organisms and inanimate matter. This study noted that very few high school students reasoned consistently at level four, even though level four was consistent with national standards. It is important to note that due to the socio-ecological perspective taken by this work, the marine and geologic carbon cycle are largely left out, and thus are important components of the carbon cycle that are excluded from this learning progression. However, this work serves as a useful comparison in our current study to understand the conceptions of undergraduate students.

Other alternate conception work has explored the greenhouse effect, which is related to the biogeochemical cycling of water, carbon, and nitrogen. Quantitative work by Groves and Pugh (1999) found that elementary education majors held many misconceptions about global warming, while Askan and Çelikler (2015) used drawings to show that science teacher candidates in Turkey held many misconceptions about the greenhouse effect. Work by Libarkin and colleagues (2015) took this work a step farther by coding these drawings and conducting a factor analysis to reveal four archetypical models of the greenhouse effect that dominate student thinking. Harris and Gold (2018), used a similar principal components analysis on drawings before and after instruction to recommend that learning molecular behavior may help students develop more expert-like mental models.

McNeal and colleagues (2014) asked students to juxtapose arrows representing all of the important processes that move or change energy, water, or chemicals on an existing drawing and found that students drew more arrows after completing an *EarthLabs* (Ledley et al., 2012) climate change and Earth system module. Work by Sibley and colleagues (2007) used box diagrams to understand student conceptions about the rock, water, and carbon cycles. This work found that students tended to have more accurate mental models of the water cycle—likely due to the lack of chemical reactions taking place in the water cycle—than the carbon cycle or rock cycle.

Most recently, You and colleagues (2017) took a more holistic look at the carbon cycle, moving beyond photosynthesis, respiration, and food chains. Rather, this research team examined interdisciplinary understanding as it related to the carbon cycle and developed an Interdisciplinary Science Assessment of Carbon Cycling (ISACC), which they validated using high school and college students. Their work demonstrated that college students had significantly higher levels of interdisciplinary thinking than high school students. This study employed both standard multiple-choice questions and some open-ended prompts which allowed for qualitative analysis. The authors called for further research to examine how students' relative levels of knowledge in various disciplines affect what disciplinary content they use when thinking about a question that requires an interdisciplinary approach. It is important to note that almost all of these studies focused on biologic and atmospheric portions of the carbon cycle, with few delving into marine or temporally long portions of any biogeochemical cycles.

This study aims to build on preexisting work on elemental cycling by taking a solely qualitative approach to fill in some gaps in previous studies and to lay a foundation for future work on both applying systems thinking models to biogeochemical cycles and developing learning progressions for complex systems based on these cycles. By using interviews, drawings, and questionnaires we sought a rich understanding of how students conceptualize these major biogeochemical cycles. In the future, we plan to use this knowledge to inform instruments that measure Earth system thinking skills as well as basic biogeochemistry knowledge. In addition, this work lays key foundations for applying systems models and developing learning progressions for complex systems based on these cycles. This study has two primary research questions: (1) How do university students conceptualize the biogeochemical cycling of carbon, nitrogen, and phosphorus? And (2) How does the Earth system manifest itself in these conceptions?

Methods

Content Analysis

Content analysis is a qualitative mode of systematic examination of materials, be they textual, musical, or pictorial. A key characteristic of content analysis is its systematic nature of breaking down the text or other artifact into single units of analysis (codes) which are then oriented into a system of categorical codes. Inter-coder or inter-rater reliability—a measurement of agreement between coders--is an important component of measuring content analysis against quality criteria, with kappa-coefficients of 0.70 typically being sufficient. Content analysis is useful in that it takes qualitative data and systematically analyzes it to create codes that allow for the incorporation of quantitative analytical procedures in a justified way (Mayrin, 2004).

Locating the Researchers

The lead author led the project, obtained university Institutional Review Board approval, conducted all the interviews, and led data analysis and writing. He is trained as a geoscientist, geoscience education researcher, and science teacher with advanced degrees in geosciences and science education and is currently working towards a Ph.D. in Earth System Science. In addition to a primary research interest in geoscience education, he has broad training across the geosciences, in particular biogeochemistry and geobiology. The second author is a professor of geosciences and serves as the first author's Ph.D. advisor and is currently primarily a geoscience education researcher but is also trained in biogeochemistry. The third author is a professor of science education, with a research focus in part on alternate conceptions of middle school youth. The second and third author's responsibilities included assistance in establishing trustworthiness and credibility through study design and inter-rater reliability; they also assisted with the development of the manuscript.

Participants

Participants were recruited from a variety of science courses at a large research university in the southeastern region of the U.S. during the spring semester of 2018. Participants were targeted by their field of study. For example, overview and introductory courses were targeted towards non-STEM majors, upper-level geology courses were targeted towards geology majors, etc. In this case of this study, non-STEM majors included those majors not housed in colleges of science, forestry, agriculture, or engineering. We also used appropriate on-campus listservs to recruit students. Once saturation was reached--meaning no new data was emerging-- recruitment in those classes was discontinued. Participation was voluntary and incentivized by offering participants a $20 Amazon gift card upon completion. In total, 53 students (29 males, 24 females) participated in the study. By credit hours, three were freshmen, 21 were sophomores, 13 were juniors, and 16 were seniors. Students were classified into broad fields based on their major.

Data Collection and Triangulation

All participants engaged in individual interview sessions, which took roughly 30-45 minutes and involved three major components. Participants first completed a semi-structured interview with questions related to where a given element is found (e.g. carbon, nitrogen, etc.), what its uses are, how it cycles, and what environmental issues may be associated with that element (Appendix A). The interview was semi-structured in the sense that the interviewer had set questions to ask, but there was also flexibility in the order, and questions could be added as needed to further probe the participant. During the interviews, participants were asked separately about carbon, nitrogen and phosphorous. Participants were reassured that this was not a test and that the study just focused on determining what they thought. Participants were encouraged to share anything they thought they knew, but they were also allowed to say that they did not know if they truly thought that they knew nothing about a particular topic.

After the interview portion, students were given a piece of paper and asked to sketch each of the carbon, nitrogen, and phosphorous cycles. If the participant was stuck, the researcher offered guidance based on information students stated during the interview portion. Students were asked to think aloud as they drew their cycles in order for the researcher to gain greater insight. Students were permitted to not draw a cycle if they truly did not have a conception of it and were unable to give any information during the previous interview portion. Once students drew the 3 cycles (or the ones they were familiar with), they were given a picture and asked to label elements in the picture that they thought had carbon, nitrogen, and phosphorous. This picture was the same illustration used in the McNeal et al. (2014) paper. Students were then asked to draw arrows showing how any of those elements could move. This was done to ensure that students were probed deeply about each element and to help confirm information drawn about their mental models from the interviews and the drawings.

Lastly, we asked students to take a brief computer questionnaire (Appendix A). This questionnaire collected demographic information (class rank, major, minor), past college courses taken, and information on when they recalled learning about the various biogeochemical cycles. The questionnaire closed with five pilot multiple choice questions on the biogeochemical cycling of carbon, phosphorus, and nitrogen. This information was used to give one final dimension to students' understanding of biogeochemical cycling and also added a small embedded quantitative portion to the data.

Data Analysis

Interviews were transcribed verbatim and imported to the qualitative software package, Dedoose. The first pass of the data consisted of open coding of the data and constant comparison analysis. Corbin and Strauss (2008) define constant comparison analysis as an inductive method that takes information from several data sources and compares one to another to find patterns. The next pass through the data involved categorical coding or grouping of the initial categories to find underlying structure and themes. The coding process was iterative and circular, allowing for the clarification and refinement of primary codes and themes to consolidate them. Interview data were initially analyzed by author 1, and then author 2 examined a subset of initial coding and agreement was reached concerning emergent codes and themes. Upon completion of coding, a random subset of specific codes were applied to randomly selected excerpts by author 2. For each "test," Cohen's Kappa was applied to ensure a value of at least 0.70 (Mayring, 2004). The excerpts were then discussed by both researchers to discuss any discrepancies and reanalyzed to ensure a 100% agreement.

The drawings were analyzed similarly, scanning and importing them into Dedoose and making an initial pass using open coding to describe what was in the drawings. Again, constant comparison analysis was used. Coding involved mainly focusing the reservoirs students drew in order to examine how the Earth system manifested itself in students' conceptions. Categorically we looked for the four major components of the Earth system: the atmosphere, biosphere, geosphere, and hydrosphere. We also coded each reservoir individually (i.e. ocean, plant, animal, atmosphere, shells, etc.). The transcript of the narration that students provided as they drew was also used to clarify reservoirs depicted as well as to code any additional information on fluxes, or the movement of material, shown in the drawings. The drawings were initially analyzed by author 1, then a subset was analyzed by author 3. The researchers verified the emergent codes and applied them through the subset until 100% agreement was reached.

The demographic data was collected and paired with the drawings and interview transcripts. This allowed for the analysis of the frequency of codes occurring by field, the score on the five pilot items, major, and class rank. The field of study was also of interest, so codes were then were also analyzed by these categories. Dedoose was used to normalize the data and show the percentage of each code and subcode by field, providing useful information in differences of students' conceptions based on their training. Information on courses taken and when students recall learning the information was also analyzed using constant comparison analysis.

Trustworthiness and Credibility

Triangulation is an important tool in ensuring trustworthiness and credibility in qualitative research (analogous to reliability and validity in quantitative research). We triangulated data collection to ensure that a rich dataset was built from interviews, drawings, and questionnaire data. When common themes emerged from multiple methods of data collection, this allowed for a cross-check of the accuracy of data generated from one method with data generated from another. This approach also enabled the effectiveness of each method to be assessed, producing student mental models that could be compared, and provided the most accurate reconstruction of student mental models possible. Additionally, tasks during the study were presented from broadest (least structure, e.g. open-ended interview questions and drawings) to narrowest (most structure; e.g. labeling a drawing, answering multiple-choice questions) in order to scaffold student responses and to truly try to build their complete mental models. The five pilot questions were examined by experts in biogeochemistry in order to ensure their quality and content validity. Throughout the process, the research team met to discuss and code subsets of the data to ensure inter-rater reliability and ensure the trustworthiness of the results.

Results

The Carbon Cycle

Figure 2 shows the most frequently occurring codes and the associated categorical codes that link them to the biogeochemical cycling of carbon. The numbers represent the percent of study participants who mentioned the given code at least once. Table 1 shows the codebook for the most frequently occurring codes and subcodes. In terms of reservoirs, or parts of the Earth system where carbon is stored, students most frequently discussed carbon being found in living things, followed by the atmosphere, and lastly in the ground (be it in rocks, minerals, soil, underground fossil fuels, or just the Earth itself). The largest reservoir of carbon, aside from kerogens in the deep Earth, is carbonate rocks (Knoll et. al., 2012); the atmosphere is a relatively small reservoir for carbon. In terms of fluxes, the most occurring codes were related to photosynthesis, respiration, and decay, which have been examined in detail in the literature (Hogan & Fisherkeller, 1996; Lin & Hu, 2003; Köse, 2009; Hartley et al. 2011; and Parker et al., 2012).

Table 3.1 The codebook for the most occurring carbon cycle subcodes

Primary Code	Sub-Code	Description	Example
Reservoirs	Atmosphere	Participant mentions carbon being a component of the atmosphere	“CO_2 is atmospheric...It's pretty abundant in the atmosphere, that gets taken in by people”
	Living Things	Participant mentions living things (plants or animals) containing carbon	“Obviously, as I said the human body has a lot of carbon. Then, DNA molecules have carbon too. I think that's crucial, genes and DNA.”
	Rocks and Soil	Participant mentions carbon being present in rocks, soil, or the solid Earth	“I think most of the carbon is just going to be on the ground, the soil, dirt”
Fluxes	Photosynthesis	Participant mentions photosynthesis moving carbon or plants taking in carbon dioxide	“CO_2 in the air is taken in by plants in photosynthesis, turned into sugars”
	Decay	Participant mentions decay or carbon being released when an organism dies	“In organic decay, carbon goes into the earth through, like I said, biomineralization. Maybe with decomposer microorganisms that might make it move through systems”
	Respiration	Participant mentions respiration or animals releasing carbon dioxide when they breathe	“Sugars can then be eaten by animals, break it down in respiration, and exhale it back as CO2”
Uses	Fossil Fuels/ Industry	Participant mentions carbon being used as a fuel or for power	“We're breaking down carbon that has been the fossils and oil, and all that is very carbon rich. When we take it and burn it, we're releasing all that carbon that's been stored underground into the air.”
	Living Things	Participant mentions living things requiring carbon or carbon is a component of essential cellular machinery	“I think it's one of the building blocks of organic molecules like I said. It's extremely important and crucial. It's in the simplest things as I mentioned before. Like the simplest molecules. So everything that's grown and developed probably started off as something like some simple carbon structure.”
	Sugars	Participant explicitly mentions sugar, glucose, or some other carbohydrate	“Energy and then structural. For planting, you have starch or glucose which are structural, or for energy.”
Environmental Impacts	Global Warming	Participant mentions global warming	“That's why global warming is such a thing because we're pumping so much carbon back into the atmosphere by burning coal, oil.”
	Destroys Ozone	Participate mentions carbon being harmful to the ozone layer	“Too much of anything could be bad with the ozone layer. I'm guessing how those elements react with oxygen could potentially be very harmful, just break down that protective layer that we have”
	Pollution and Smog	Participant mentions carbon being a component of pollution	“We're releasing excess carbon dioxide since the industrial revelation, polluting our air, causing smog. Acid rain has been a problem”

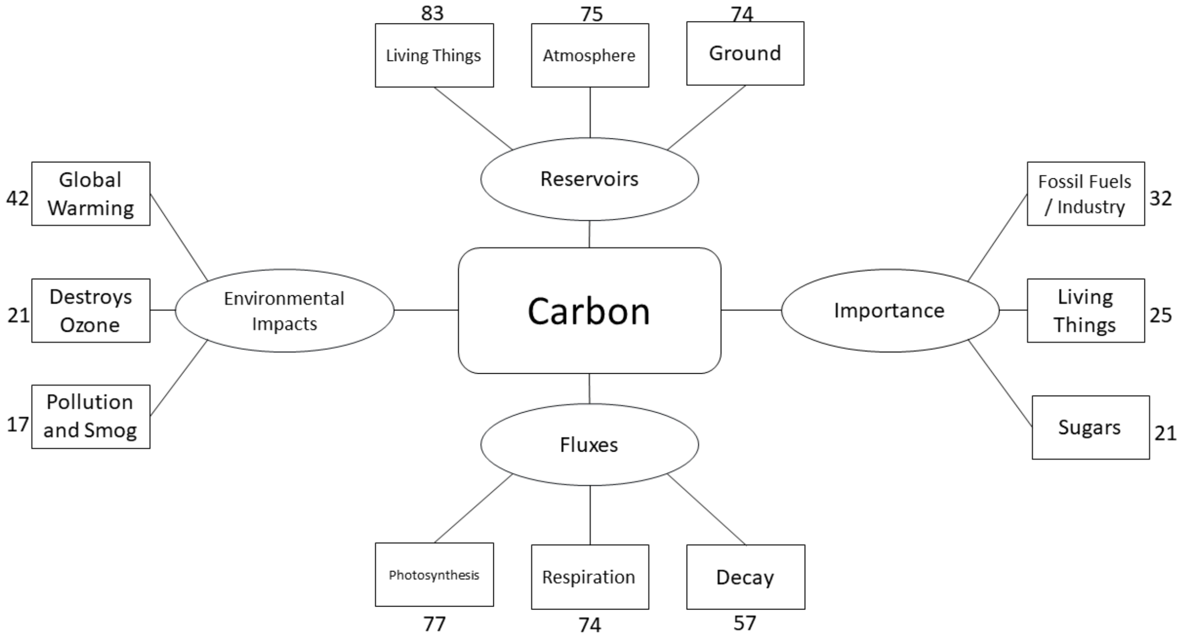


Figure 3.2 Map of most frequently occurring codes as they relate to the carbon cycle. The codes are grouped into major themes and the numbers represent the percent of participants who mentioned the code at least once during their participation.

Interestingly, students most identified fossil fuels as a major importance of carbon and carbon compounds, followed by the use of carbon by living things and specifically, its role in sugars. Students were also able to connect the burning of fossil fuels to global warming. However, an alternate conception emerged, as almost half of those students described ozone depletion as the mechanism for global warming. One student revealed:

" Things like carbon dioxide, carbon monoxide can, when they are released to, I don't know which part of the atmosphere, certain part of the atmosphere, degrade the atmosphere. I know ozone is a big thing, which is O3, creating gaps in our atmosphere, which then allows for greater heat influx and efflux, which causes things like global warming." – Senior, Biochemistry/ PreMed

While carbon compounds are associated with the depletion of ozone, it is not the carbon, but rather chlorine that reacts and destroys ozone in a positive feedback loop. Additionally, this does not cause global warming but a decrease in the absorption of ultraviolet radiation in the atmosphere (Schlesinger & Bernhardt, 2013). Other students who did not implicate the decrease in ozone with global warming, showed an equally fuzzy understanding. One describing global warming as: "Basically, the idea that it is weakening our outer atmosphere and causing a global climate change." So while students largely (almost half) recognize the role of carbon and global warming, they also have alternate conceptions or no conception of the actual mechanism.

Students also demonstrated some confusion as to what actually happened during photosynthesis and respiration. Many students mentioned photosynthesis and respiration in their interviews and drawings, but they did not always understand the role of carbon:

"I don't know if they use it for energy. I can picture the graphic of photosynthesis. I don't know how exactly how they use it or for what purpose. I don't know."-Sophomore, Computer Science

"Doesn't the flower just take the nutrients from light? I guess. I don't know if it's first, carbon, and light. I don't really know." –Sophomore, Natural Resources Management

"Like I said, carbon makes up all organic matter. We're organic matter. Simply the process of tissue and organ usage produces the carbon as a byproduct." –Senior, Medical Laboratory Studies

Of the students who mentioned respiration during their interview or drawing think aloud, 23% reported not knowing what happened to carbon during the process. While discussing photosynthesis, 20% did not know what happened to carbon during the process, but only 5% thought the energy was being produced during photosynthesis. Twenty-four percent of students also claimed that carbon was taken in through the roots. Fifty-one percent correctly identified that the carbon that was taken in during photosynthesis was used to make sugars. Thus, participants across disciplines did not consistently understand photosynthesis and respiration and the transformations involving carbon.

Interviews and drawings also revealed that students often do not include the geosphere or the slow carbon cycle in their mental models of carbon cycling, and if they do, it is often in the form of soil. The role of shells (8% of participants), limestone, or other carbonate rocks (9% of participants) seldom came up in interviews, even though they are major components of the carbon cycle. Only 9% of students discussed the role of weathering and erosion in carbon cycling, and not a single student (including geology majors) spoke about the chemical weathering of silicate rocks, which is the largest long-term control of atmospheric carbon (Schlesinger & Bernhardt, 2013). Additionally, only 29% of students correctly responded to the pilot question concerning the relationship of climate to chemical weathering only. Drawings fared worse in terms of the slow carbon cycle, with only one participant including shells in their illustration and five participants including rocks in their pictures. Overall, only 43% of participants depicted some element of geosphere (soils and fossil fuels being the most depicted aspects) in their drawings.

The lithosphere, however, fared better than the hydrosphere, which only appeared in some form in 22% of drawings. Of those drawings, 67% included the ocean while 41% included rain. Almost every incidence of including rain was related to trying to place carbon into the water cycle as part of the cycling of carbon, with only one drawing depicting rain as a source of chemical weathering. The atmosphere and biosphere were well represented in student drawings, with 94% of students depicting some elements of the biosphere (typically plants and animals) and 85% depicting carbon moving through the atmosphere. Only one student was unable to draw a representation of the carbon cycle.

The Nitrogen Cycle

Figure 3 shows the most frequently occurring codes and the associated categorical codes that link them to the biogeochemical cycling of nitrogen. The numbers represent the percent of study participants who brought up the given code at least once. Table 2 shows the codebook from which these most frequently occurring codes arose. In terms of reservoirs, or parts of the Earth system where nitrogen is stored, students most frequently discussed nitrogen being found in the atmosphere, which is indeed a major reservoir. The atmosphere was followed by living things, which was followed by soil. In terms of fluxes, students most frequently mentioned nitrogen being consumed and moving through the food chain, followed by decay, and lastly as being absorbed by plants through the roots, connecting back to the soil as a reservoir.

Table 3.2 The codebook for the most occurring nitrogen cycle subcodes

Primary Code	Sub-Code	Description	Example
Reservoirs	Atmosphere	Participant mentions carbon being a component of the atmosphere	“There's a lot more nitrogen in our atmosphere than anything else”
	Living Things	Participant mentions living things (plants or animals) containing nitrogen	“It's also used a lot in our bodies and every animals' bodies”
	Soil	Participant mentions carbon being present in the soil (but not rocks)	“It’s a nutrient in the soil. Water can move it away, actually.”
Fluxes	Eating/ Food Chains	Participate mentions that nitrogen can be passed from plants to animals and from animals to other animals through consumption or predation	“We might have some type of food that has nitrogen in it, not sure. Organisms eat it. Then it goes through their body, passes”
	Decay	Participant mentions decay or nitrogen being released when an organisms dies	“It's when nitrogen is reduced to nitride and nitrate by decomposition. I'm not sure. It's the whole thing from the ground. I know they get it from the ground and then they do something to it and it becomes a gas”
	Absorption from Soil	Participant mentions that plants absorb nitrogen from the soil through their roots	“It's going to the ground. The plants are using it. That's good”
Uses	Fertilizer	Participants mention that nitrogen is used in fertilizer	“A lot of plants utilize it for nutrition. A lot of fertilizers are very nitrogen heavy”
	No Conception	Participants state that they don’t know a use for nitrogen	“I'm sure there are, but I don't know them. I know there's nitrogen fixation and stuff. I don't know if that's good.”
	Biological Function	Participants mention that nitrogen is used by living things	“Yeah, it's like in amino acids. We have to have it to function.”
Environmental Impacts	No Conception	Participants state that they don’t know an environmental impact of excess nitrogen	“I’m sure there are, but I don’t know any.”
	Eutrophication	Participants mention that excess nitrogen causes algae blooms, though they may not have the vocabulary for what it is or the impacts	“It causes a lot of the algae. It's a bloom over a bloom and then it creates a biofoam on top of the water which sucks out all the oxygen and blocks sunlight and as a result kills the fish and other marine life.”
	Global Warming	Participants indicate that nitrogen or nitrogen compounds are related to global warming	“I feel like nitrogen has something to do with...I wouldn't say global warming, but I just feel like that one's always deemed, "The Bad Guy." Don't know why I feel that way because I couldn't give you an example”

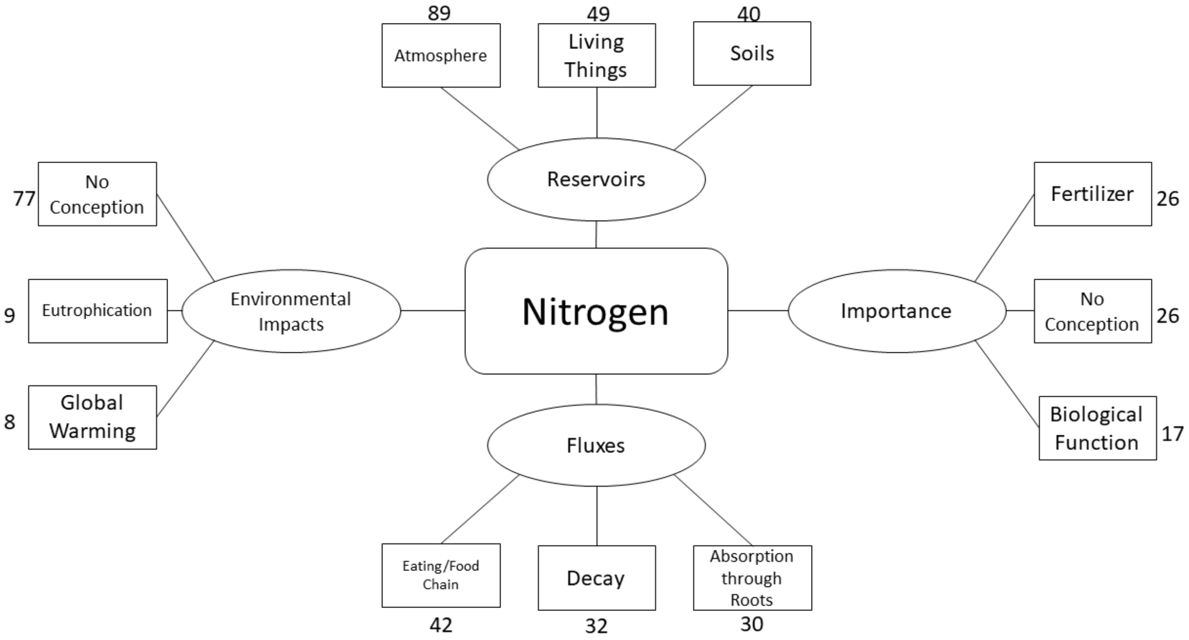


Figure 3.3 Map of most frequently occurring codes as they relate to the nitrogen cycle. The codes are grouped into major themes and the numbers represent the percent of participants who mentioned the code at least once during their participation.

In terms of the importance of nitrogen, in equal amounts students thought of fertilizer or had no conception of what its importance was. Seventeen percent mentioned nitrogen having a major role in biological function, which is contrasted with the 49% who mentioned nitrogen being in living things, indicating a lack of conception as to why nitrogen is so abundant in living things. Of those who did note its biological function, students most typically mentioned its role in DNA as well as its role in proteins. The majority of students were unfamiliar with any sort of environmental issues associated with nitrogen, with a small percentage bringing up eutrophication and global warming. Interestingly, one of the pilot questions asked about the impacts of nitrogen on an aquatic ecosystem, and only 21% of participants correctly indicated that it would increase primary productivity; the majority indicated that nitrogen would directly poison living things or change the pH or temperature of the water, revealing that eutrophication is not well understood by this population.

Microbial activity is a major component of the nitrogen cycle, as atmospheric diatomic nitrogen is inert due to its triple bond. Many students recognized that nitrogen could not be directly taken in from the atmosphere (34% of students made some indication to either plants or animals breathing or taking in nitrogen for use), but only 23% of students (12 students) brought up microbial activity during the interview. When discussing microbial activity, nitrogen fixation came up the most (12 students), followed by nitrification and denitrification (three students), and ammonification coming up the least (two students). The drawings showed similar student understanding, with 21% of drawings depicting microbes engaged mainly in nitrogen fixation. Largely, students were most familiar with nitrogen fixation in name, but did not always understand why it is important:

" I know there's nitrogen fixation and stuff. I don't know if that's good." "What's nitrogen fixation?" "I couldn't tell you, I just know it's a thing." –Sophomore, Applied Mathematics

The drawings of the nitrogen cycle overall were not as detailed as those of the carbon cycle, with 15% of students having no conception to draw. Fifty-five percent of drawings included nitrogen in the atmosphere, which is lower than the percentage in which it was mentioned as a place that nitrogen is found during interviews. Fifty-eight percent of drawings included a biosphere component, largely in the form of plants or animals. Interestingly, while fertilizer was the most common use of nitrogen brought up by students, only two students depicted fertilizer in their drawings, demonstrating a disconnect between fertilizer being an important human use of nitrogen but not knowing what the nitrogen does or how it gets there. Thirty-eight percent of students showed at least one lithosphere component, with 70% of those depicting soil. The hydrosphere was again underrepresented with 11% of drawings containing a hydrosphere element. Half of those again were rain, representing students trying to combine the nitrogen cycle with the water cycle.

The Phosphorus Cycle

Figure 4 shows the most frequently occurring codes and the associated axial codes that link them to the biogeochemical cycling of phosphorous. The numbers represent the percent of study participants who indicated the given code at least once. Table 3 shows the codebook from which these most frequently occurring codes arose. One notable difference between phosphorous and the nitrogen and carbon cycles is the overall lack of conceptions across student categories. In terms of reservoirs, or parts of the Earth system where phosphorous is stored, students correctly and most frequently discussed phosphorous being found in rocks and soil. This was followed by living things, followed by reporting no conception. In terms of fluxes, students typically did not have a mechanism to explain how phosphorous moves. Those who did typically discussed it moving through food chains or being absorbed through the roots of plants, similar to how students described nitrogen fluxes.

Table 3.3 The codebook for the most occurring phosphorus cycle subcodes

Primary Code	Sub-Code	Description	Example
Reservoirs	Rocks and Soil	Participate mentions that phosphorus is found in rocks and/or soil	“I feel like it's just a mineral in the soil”
	Living Things	Participant mentions that phosphorus is found in living things	“It's definitely in all organisms because phosphate is a really big part of biological processes. It's part of lot of amino acids and used in sugars and all kinds of things.”
	No Conception	Participant indicates that they do not know where phosphorus can be found	“I have no idea”
Fluxes	Eating/ Food Chains	Participate mentions that phosphorus can be passed from plants to animals and from animals to other animals through consumption or predation	“ I feel like it's ingested in food but that seems weird”
	No Conception	Participant mentions decay or phosphorus being released when an organisms dies	“If there's a phosphorous cycle, I don't know about that one.”
	Absorption from Soil	Participant mentions that plants absorb phosphorus from the soil through their roots	“The phosphorous cycle, you've got plants taking up phosphorous through their roots.”
Uses	No Conception	Participants state that they don’t know a use for phosphorus	“Once again phosphorus is one of those elements I know about. It's one of them are common ones, but I can't think of what exactly it's used for. I don't know.”
	DNA	Participant indicates that phosphorus is a component of DNA	“You've got your DNA which has phosphorus in it.”
	Energy/ ATP	Participant indicates that phosphorous has a role in cellular energy or a component of ATP	“Phosphorus is part of ATP. This is what we use to produce carbon dioxide and break down glucose.”
Environmental Impacts	No Conception	Participants state that they don’t know a use for phosphorus	“I’m not sure, don’t know.”
	Water Chemistry	Participant indicates that phosphorous may negatively impact water chemistry	“Phosphorus has something to do with water chemistry, because I know I've tested for that before”
	Mining Pollution	Participant indicates that phosphorus may be a product of mining pollution	“It might be part of bad mining production in some place, or like human rights violations but I can't really think of an environmental issue with it”

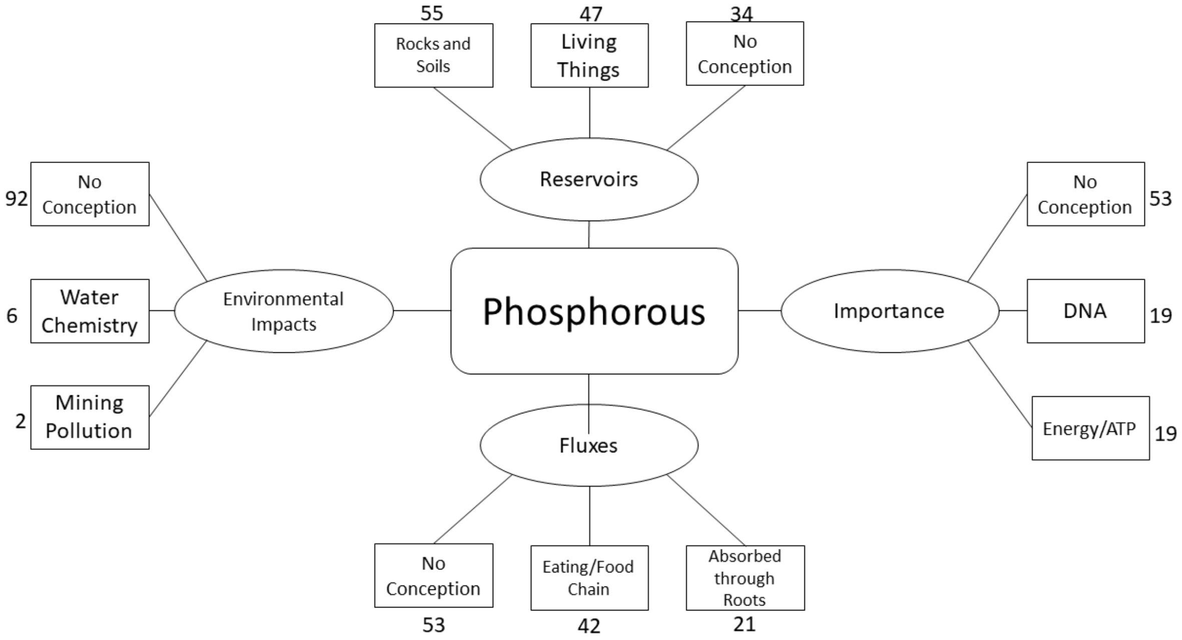


Figure 3.4 Map of most frequently occurring codes as they relate to the phosphorous cycle. The codes are grouped into major themes and the numbers represent the percent of participants who mentioned the code at least once during their participation.

Most students did not know why phosphorous was an important element, but some did report it being found in DNA as well as ATP, the energy currency of cells. Interestingly, many students stumbled into ATP, while discussing respiration and drawing out the carbon cycle. When asked what ATP was, many students correctly revealed that it was Adenosine Triphosphate, upon which they suddenly realized a use for phosphorous. Largely, students were unfamiliar with environmental impacts of phosphorous, while a small number referred to eutrophication and other water pollution issues. One student mentioned mining pollution.

Most students either did not have a conception of phosphorous being in the atmosphere or having an atmospheric component, with only 19% reporting in the interview that phosphorous was found in the atmosphere. 11% of students explicitly stated that phosphorous was not in the atmosphere. Those students who did report phosphorous being in the atmosphere often did it with uncertainty:

"I'm going to say also in the atmosphere. Not as prominent as oxygen and nitrogen, but there too." –Sophomore, Biomedical Sciences

Others expressed a more ephemeral view of phosphorus in the atmosphere:

"If we're using gun powder it burns, then it burns and it gets put into the atmosphere*. It's got to be in the atmosphere at some point or another."* –Sophomore, Business Management

However, when students were asked to label a drawing with carbon, nitrogen, and phosphorous, many students did put phosphorous in the atmosphere, usually because carbon and nitrogen were there. This was during the most structured task, so it is interesting that students were more inclined to put phosphorous in the atmosphere during this task, but not during the interview or drawing portion of the study. Phosphorus rarely occurs in the atmosphere, which is one of the reasons that the phosphorus cycle occurs at rates much slower than the carbon and nitrogen cycles and why it is so limiting in ecosystems (Schlesinger & Bernhardt, 2013).

In terms of drawings of the phosphorus cycle, 13% of participants reported having no conception and not knowing what to draw. Only 15% of drawings included an atmospheric component, which supports the interview data. Twenty-eight percent of drawings depicted an element from the biosphere (again animals were the most common, followed by plants). Despite naming rocks and soil an important source of phosphorous, only 15% of students included it in their drawings (the same as the atmosphere). Unlike nitrogen and carbon, it was typically shown in rocks (88% of occurrences) rather than soil (38% of occurrences). Again, the hydrosphere was only depicted in 11% of drawings, equally as rain, oceans, and rivers. A comparison of the frequency of the four Earth spheres depicted in drawings by cycle is shown in Figure 5.

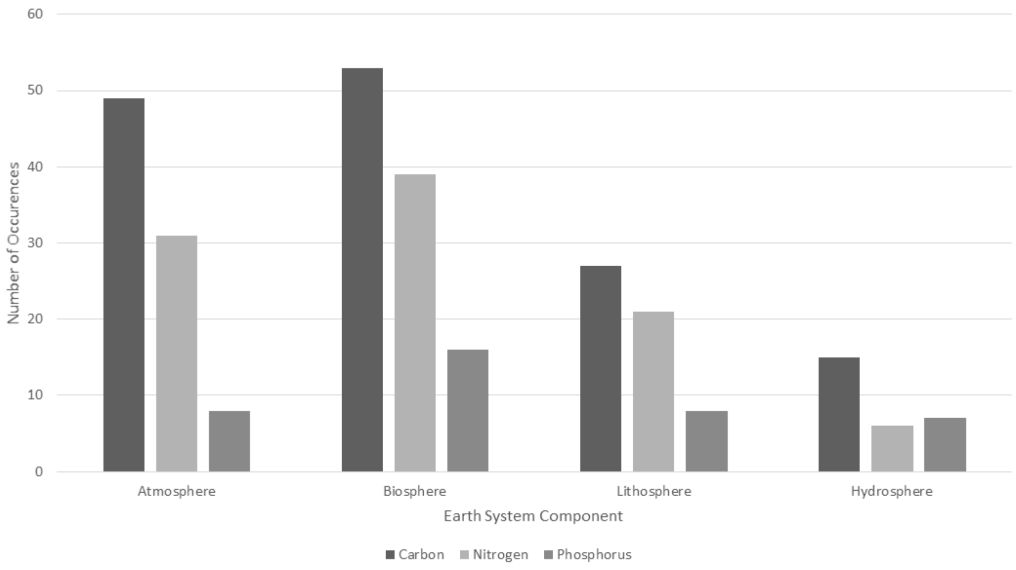


Figure 3.5 Graph depicting the number of occurrences of the four Earth spheres by biogeochemical cycle in student drawings.

Differences by Field

In addition to examining student conceptions of major biogeochemical cycles, we also were curious if a student's major field had any influence on their conceptions. As previously mentioned, the hydrosphere was underrepresented in respect to the other Earth spheres; however, geology and science education students were most likely to include the hydrosphere in their mental models (Figure 6) (73% of codes for carbon cycle came from these two fields). Students in these two fields also provided the bulk of the codes on involving limestone and carbonate rocks (80.5% of codes). Analysis of the questionnaire revealed that these students took many similar courses including physical geology, historical geology, geomorphology, paleobiology, as well as at least two semesters of biology and chemistry.

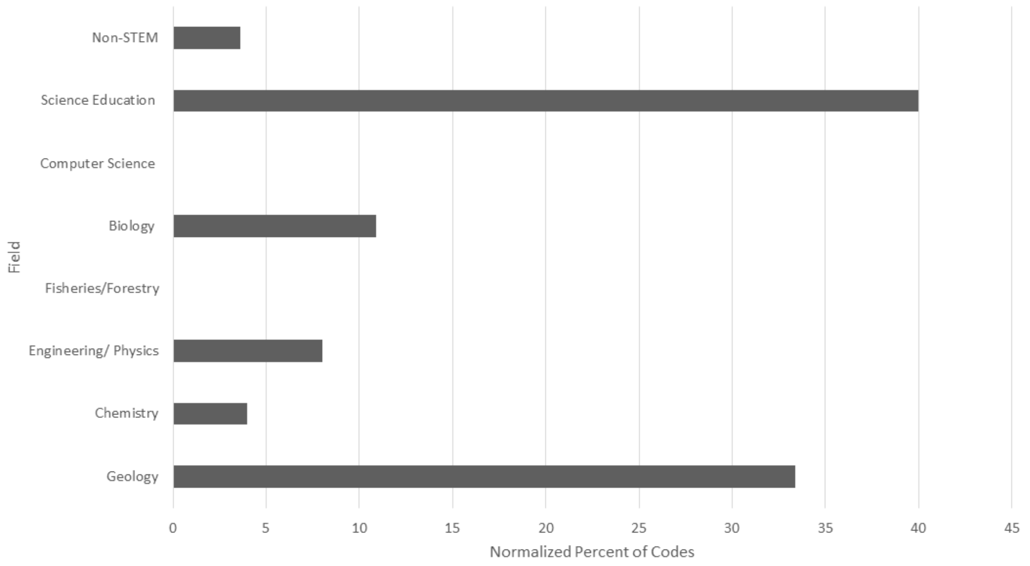


Figure 3.6 Percent appearance of codes involving the hydrosphere by field. Note data are normalized, meaning that proportions are used to not bias any field based on the number of participants.

Biology and science education students accounted for more than half of the codes related to microbial activity in the nitrogen cycle. Students in both groups reported taking basic molecular and cellular biology, organismal biology, and ecology. Chemistry students were least familiar with the phosphorus cycle (almost 25% of "I don't know" codes). Eighty percent of codes relating to phosphorus being in the atmosphere came from non-STEM and science education students. Additionally, the previously noted alternate conception relating to carbon destroying ozone was most pervasive in non-STEM and chemistry students.

Case Studies

In order to give insight on data analysis and triangulation, as well as to explore possible implications for developing learning progressions for basic biogeochemistry, we present three case studies of the data collections process using three students from different science backgrounds, class ranks, and majors. These cases were chosen because they are typical for the participants in non-STEM fields and various stages of a STEM field. In this case, we chose to present a case study typical of a non-STEM student and two STEM students within the same field, in order to demonstrate how these conceptions may vary as students progress in their field.

Case Study 1, Non-STEM

Kelly was a freshman psychology major and she reported taking two STEM classes at the university. In the interview portion, when asked for a definition of “system”, Kelly revealed that she pictured a system as a community that involves different people and groups of people. When asked how a system might apply to planet Earth, Kelly thought about various components of the Earth system but not in a connected sense stating, “[It is] Probably the Earth's magnets and different kinds of animals and people. The weather. I guess that's it.” When asked where carbon can be found, Kelly mentioned that carbon was in the air, likely in the form of carbon dioxide, and in rocks, though she was not sure what kind of rocks may contain carbon. Kelly, however, could not think of any uses for carbon. In terms of the movement of carbon, Kelly remembered maybe seeing a poster in middle school, but couldn’t recall much about it. When asked if she had any thoughts on how carbon might change forms or moves, Kelly mentioned that it had half-lives. When asked what a half-life is Kelly stated, “It's like something deteriorates and it changes forms based on how many protons are in it.” Kelly did not have any additional information to report on carbon or the carbon cycle.

When asked about where you can find nitrogen, Kelly knew about nitrous oxide but was not sure if it was naturally occurring or not, just that she had heard of it. When asked for uses of nitrogen, Kelly stated “I think nitrogen is on one of the noble gases. I think so, and then those are stable, so they are put in things so that it doesn't react with anything like change.” When asked about how nitrogen cycles Kelly admitted that she wasn’t familiar with the nitrogen cycle. Kelly was unfamiliar with phosphorous and the phosphorous cycle and was unable to answer any questions relating to it. The interview portion ended by asking about environmental issues associated with any of the aforementioned elements. Kelly stated that she thought that carbon dioxide might be in the ozone layer and elaborated stating “It’s in the ozone layer, and then the less there is of it, it's a bigger problem. Or maybe more, I don't know. Whatever it is, because it traps in heat, and then the heat can't escape. That's a problem.”

Kelly was next given a piece of paper and was asked to draw out the carbon cycle. Kelly started by drawing a plant and then a cloud indicating the atmosphere. She then drew an arrow connecting the plants to the ground, indicating that carbon moved from the tree to the ground, then a second arrow connecting the ground to the cloud, indicating that carbon moved from the ground to the atmosphere. Kelly explained that the carbon moved with the water and then returned to the tree through condensation. Kelly also indicated that she felt what she drew was incorrect. The final drawing depicted a unidirectional loop. Kelly was unable to depict the nitrogen or phosphorous cycles because she felt that she did not know enough.

To further assess her knowledge, Kelly was given a cartoon image of a simple landscape (Figure 7) and was asked to label any items in the picture that might have carbon with a C. Kelly labeled the tree and rabbit because they were alive. She also put it in the air. Upon looking at the picture she decided that it could be in a lot of the things depicted, putting carbon in the soil and cloud, but not the rain. Next, she was asked to place arrows anywhere she thought that carbon might move between various locations. She put arrows from the tree to the soil indicating that they are physically connected and can get from the tree to the soil through its roots. She then placed an arrow from the soil to the air stating that evaporation may be the process at play. She then connected the air to the tree indicating that condensation was the mechanism. She was then asked to do the same thing with nitrogen and phosphorous. Though she indicated uncertainty, she put nitrogen in the sun, cloud, air, and soil. When asked to put arrows, she was unsure about how nitrogen would move between any of those reservoirs. She was even more unsure about phosphorous but ultimately ended up putting it in the same places she put nitrogen and was unsure of how it would move.

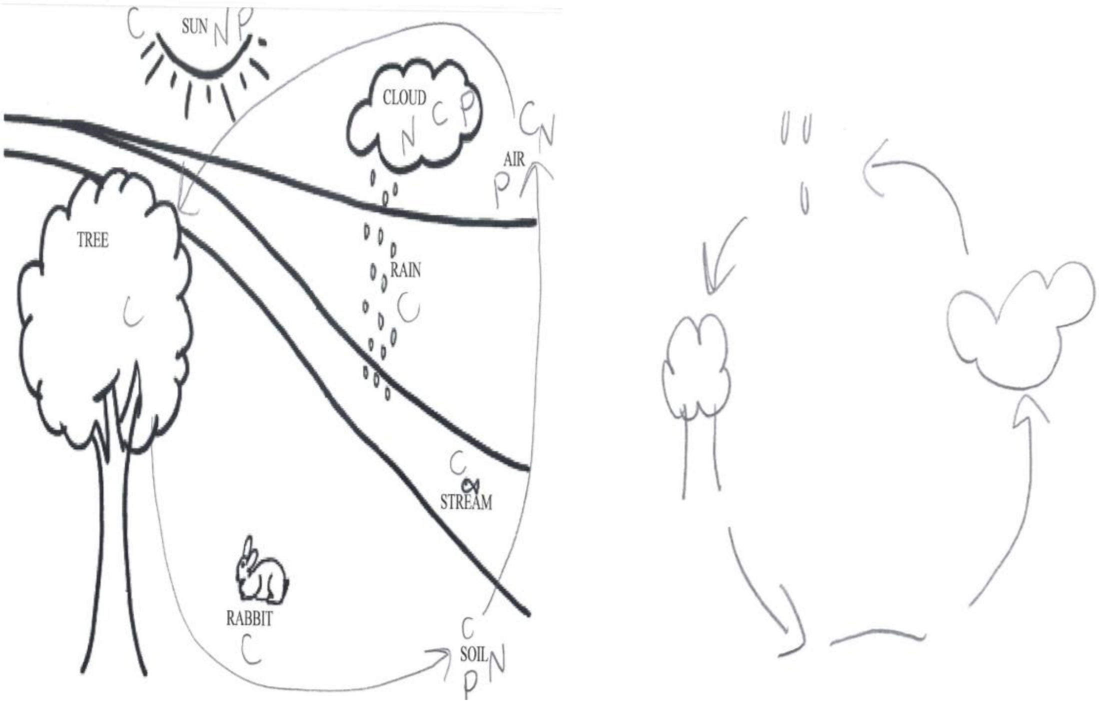


Figure 3.7 *Drawings* produced by Kelly during her interview. The left drawing depicts how she filled in the cartoon and the right is her model of the carbon cycle.

In the post-interview survey, we learned additional information from Kelly. Of the two STEM classes she took, one was a concepts of science class that overviewed the major scientific disciplines, and the other was an introductory chemistry course. In the post-interview survey, she indicated recalling learning about the carbon cycle in high school chemistry but not in college. She did not recall ever learning about the phosphorous and nitrogen cycle. In terms of discussing the Earth system, she remembered learning about things that make up the Earth in her concepts of science course at the university, but in the interview, she focused on the Earth's magnetic poles, which she considered a major part of the Earth system.

Case Study 2- Early STEM

Eric was a sophomore, pre-med major at the time he participated in this study. When asked to define a system, Eric stated that a system is "something that you can put work into or get something out of. Be it energy, or just a product, or something of that nature. It's a very broad term." When asked how the Earth may be considered a system, he stated that he saw the Earth as being made up of subsystems like ecosystems and weather. He also included gravity or other planets, things that accumulate into "making the Earth function the way it does." When asked about where carbon can be found, Eric immediately thought of carbon dioxide coming out of organisms and the fact that organisms are carbon-based. He also thought of carbon dioxide in the atmosphere and thought that carbon was present in most things. He also mentioned carbon dating of rocks. When asked about uses beyond carbon dating, he responded "I know that basic photosynthesis is plants take in the carbon dioxide that we excrete from our own bodies, and then use that to make their food, or whatever. They make oxygen from it. It's just like a cycle." The interview was then able to transition to speaking more specifically about the carbon cycle. Again, when thinking about the carbon cycle, Eric indicated that he always thought of the human body expelling carbon dioxide. When asked why we expel carbon dioxide, Eric stated "I don't know, it's about anatomy. It's just because our bodies can't really do anything with it, I think. It's just not necessary. I forget where the carbon comes from, but we have to excrete it somehow." When pushed further, he recalled from cell biology that cellular respiration came into play.

When asked where nitrogen can be found, Eric initially thought about ozone, but then decided that that was incorrect. He then recalled that one of the layers of the atmosphere is primarily nitrogen or has some sort of nitrogen component. He also stated that he doesn't really think of nitrogen as being that important. When asked if he knew of any uses of nitrogen, he could only think of “general chemistry reactions” but could not think of anything “practical.” When asked about the nitrogen cycle, Eric reported thinking of it as just “moving nitrogen around,” but reported not being aware of it. The interview next moved to discussing where phosphorous is found, with Eric mentioning phosphoric acid, which he recalled using in chemistry labs, and geysers, although he was not sure why he thought of geysers. He was not sure what phosphorous was used for but thought that there were likely many uses. When asked about the phosphorous cycle and how phosphorous might move, he was again not sure but thought that it might have to do with plants. The interview closed by asking about environmental issues potentially associated with any of the elements. Eric immediately thought about global warming, which he noted was a hot topic that he hasn’t done a lot of research on. When asked if global warming was associated with all three elements or just one, he stated that he thought it was just nitrogen.

For the next portion of the study, Eric was asked to draw the three cycles (figure 8). Again, he indicated that he initially thought about carbon dioxide, but chose to start by drawing a plant. He also drew a person. He indicated that carbon dioxide came out of the person and went into the plant, with the plant then releasing oxygen that was breathed in by the person. Eric again brought up carbon dating but was unclear about how other items that contain carbon relate to what he had already drawn. In his drawing, he elected to make a list of materials that he knew could be carbon-dated: cloth, linen, and paper. He noted that all of these materials were plant-based. Eric did not feel comfortable drawing the phosphorous or nitrogen cycle.

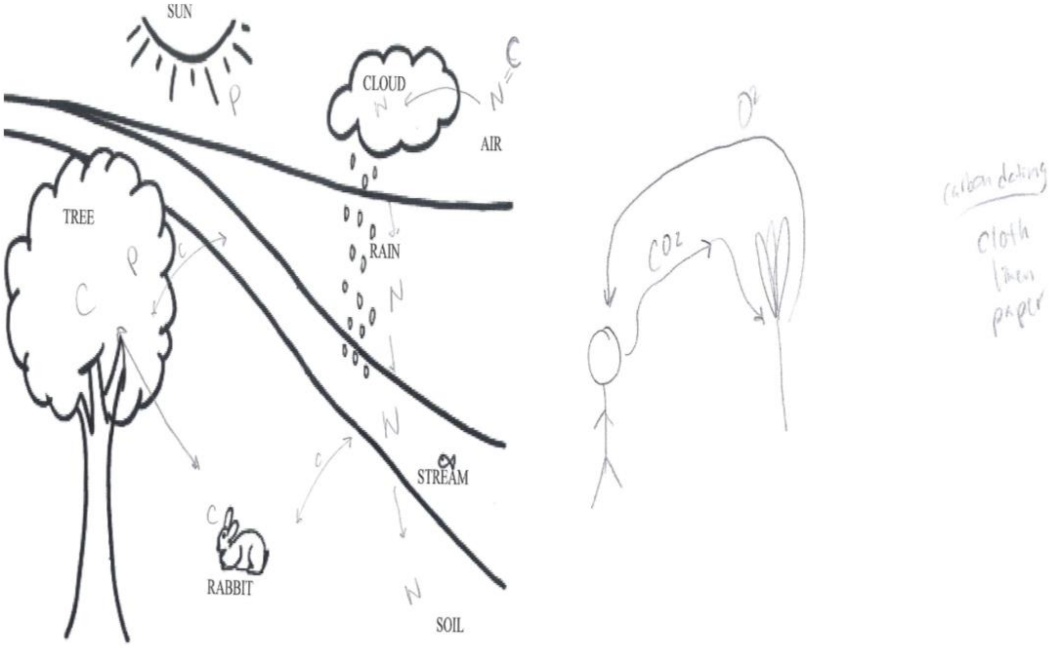


Figure 3.8 Drawings produced by Eric during his interview. The left drawing depicts how he filled in the cartoon and the right is his model of the carbon cycle.

Eric was then given the cartoon image. He placed carbon and nitrogen in the air, nitrogen in the cloud, nitrogen in the soil, carbon in the two living components the tree and the rabbit. He was unsure what to do with phosphorous but ultimately put it in the sun and the tree. When explaining the flow of carbon, he again emphasized it exchanging between the air and tree and then the tree and the rabbit. He also drew a two-way arrow between the rabbit and the atmosphere. With the nitrogen cycle he drew arrows connecting the nitrogen from the air in the cloud, he then showed nitrogen falling as rain and entering the stream and the soil. He indicated that in this case nitrogen probably just followed the water cycle. He did not show any pathways for nitrogen to get to the atmosphere. In terms of phosphorous, he was unsure and did not draw any arrows. During this portion of the study, Eric indicated that he really should know more about phosphorous, indicating that he knew that carbon and nitrogen were important to life, but that he had never thought about phosphorous.

The post-interview survey indicated that Eric had taken general chemistry I and II as well as organic chemistry. He had also taken principles of biology (focused on cellular biology and genetics), organismal biology, and two semesters of general physics as well. He recalled possibly learning about parts of the carbon cycle in high school chemistry as well as his principles of biology class. He did not recall ever learning about the phosphorous or nitrogen cycle. Eric reported that he only ever recalled hearing about the Earth system in his middle school geology class.

Case Study 3- Advanced STEM

Lucy was a senior studying organismal biology at the time she participated in this study. When asked to define a system, she stated “A system would be not just one component, but multiple components having to work together to be able to make a functional...I don't want to use the word system in the definition of system. You have multiple parts, essentially, to a system, other than just one particular thing by itself.” When asked about considering the Earth as a system, Lucy cited plate tectonics, climate, and wind, as well as living components and the oceans all make up the Earth system. She also noted the importance of the sun in this system. When asked where carbon can be found, Lucy stated “Carbon can be found in all of organic matter. It can be found throughout the biome. It can be found also in the environment. It can be found in the atmosphere, like carbon dioxide.” She went on to identify that carbon could be used as fuel, such as with coal. She also identified the importance of carbon’s biological importance as a key component of cells and tissues. When asked to identify processes that move carbon, Lucy initially discussed decomposition and the food web as ways to move carbon. She also discussed how cellular respiration will return carbon to the atmosphere, where it can again be taken in by plants.

Lucy was next asked about nitrogen and where it is found. Lucy responded that "Nitrogen can be found in the atmosphere. It makes up most of the atmosphere. It can also be found in living organisms, DNA, and so on." She also indicated the importance of nitrogen in living things. When asked about processes that moved nitrogen, Lucy identified that many of the same processes that move carbon (i.e. decomposition) also move nitrogen. She also mentioned the importance of nitrifying bacteria in fixing nitrogen and moving it out of the atmosphere. When asked if animals can get nitrogen from the atmosphere, she noted that they cannot. Lucy indicated that she was unsure about how nitrogen could be returned to the atmosphere. When asked about where phosphorus is found, Lucy indicated that phosphorus is also found in living tissue as well as the Earth's crust. She did not think that it was found in the atmosphere. When asked for uses of phosphorus, she said "Phosphorous is part of DNA and RNA. It's combined with oxygen to make phosphate." She identified that phosphorus moves from soil to plants, where it can move up the food chain to animals. Next, the potential environmental impacts of the elements were discussed. Lucy indicated that carbon definitely had associated environmental impacts, but nitrogen and phosphorus mostly did not. In terms of if carbon had an environmental impact, she stated "Carbon definitely does because it acts like a greenhouse. It definitely has an effect on the environment." She went on to note that humans were taking carbon that was locked up in the soil and returning it to the atmosphere at alarming rates.

For the drawing portion (figure 9), Lucy began drawing the carbon cycle by depicting the atmosphere. She showed that carbon being taken up by a plant. She noted that the plant can then decompose or get eaten. Moving up the food chain, she noted that that animal could also decompose or get eaten, and that carbon can be returned to the atmosphere through decomposition or respiration. When asked what was happening when the plants were taking in carbon dioxide from the atmosphere, she stated that “They’re using it to make glucose. They fix carbon dioxide into glucose. They combine those carbon molecules to make sugars through photosynthesis.” It is interesting to note that prior to this point in the interview, Lucy had never mentioned photosynthesis by name.

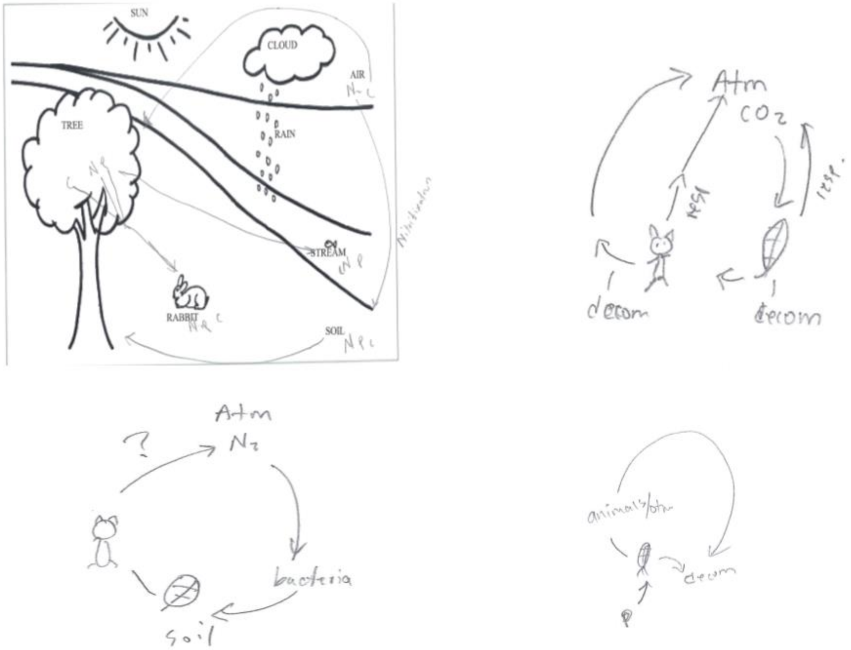


Figure 3.9 Drawings produced by Lucy during her interview. The top-left drawing depicts how she filled in the cartoon, the top-right is her model of the carbon cycle, the bottom-left is her model of the nitrogen cycle, and the bottom-right is her model of the phosphorus cycle.

Lucy was next asked to draw the nitrogen cycle, and again she started with the atmosphere. She noted that it can be taken out by bacteria and transferred to plants. From here she noted that nitrogen can either move up the food chain or move into the soil through decomposition. She again noted that she was not sure how nitrogen gets back into the atmosphere. For the phosphorus cycle, she began by showing plants taking up phosphorus from the soil through their roots. Again, she depicted decomposition moving phosphorus back to the soil or it being moved to animals through the food chain.

For the final part of the interview, Lucy was asked to depict where the three discussed elements would be found in a cartoon. She placed carbon in the tree, rabbit, soil, and atmosphere; nitrogen in the tree, rabbit, soil, stream, and atmosphere; and phosphorus in the tree, rabbit, soil, and stream. She noted that carbon dioxide can go from the air to the plant, but noted that she didn’t know what that was called (despite mentioning photosynthesis by name previously). She indicated that carbon could be further moved to the atmosphere through respiration by animals or to the soil through decomposition. She noted that nitrogen goes from the air to the soil through microbial activity, once in the soil it can be taken up by plants and moved through the food chain and decomposition. She also noted that nitrogen and phosphorus could both get into the stream from organic matter decaying there. She indicated the phosphorus moves similarly to nitrogen but without an atmospheric component.

The post-interview survey revealed that Lucy had taken a variety of STEM courses of interest to this study including an introductory physical geology course, principles of biology, organismal biology, ecology, microbiology, general chemistry one and two, and engineering physics one and two. She noted learning about the carbon cycle in high school but explicitly stated that she did not retain it well. She recalled going over it again in ecology. She also recalled briefly going over the nitrogen cycle in ecology, though she did not recall learning about the phosphorus cycle at any point. She only recalled discussing the Earth system in her physical geology course.

Discussion

This study investigated two primary questions: (1) How do university students conceptualize the biogeochemical cycling of carbon, nitrogen, and phosphorus? and (2) How does the Earth system manifest itself in these conceptions? In analyzing the data, several key themes emerged that answer these questions.

Student Conceptions of Major Biogeochemical Cycles

Mental models of all three cycles tend to be “bio-centric,” meaning that student conceptions of all three cycles are likely to be framed around the biosphere and its interactions with other spheres. This means that most of the processes that students consider refer to parts of the cycle which tend to be relatively short and on human timescales, as evidenced by the preponderance of codes relating to photosynthesis, respiration, and decay in contrast to more seldom occurring codes such as chemical weathering, sediment burial, or tectonics when discussing the carbon cycle. It is important to note that these conceptions likely will only develop with formal or informal instruction in a topic. For example, without taking a class that specifically deals with carbon cycling, climate change, or chemical weathering, students would have little reason to have any conception about the role of the chemical weathering of silicate rocks and minerals in the carbon cycle.

In comparison to the carbon cycle, the phosphorous and nitrogen cycles were less well-conceived, with student mental models of the phosphorous cycle being particularly limited. Thus, a lack-conception was much more pervasive than alternate conceptions. However, some alternate conceptions did appear. In particular, alternate conceptions about photosynthesis and respiration relating to the transformation of matter did show up in this study, confirming previous work by Lin and Hu (2003), Köse (2009), Hartley et. Al. (2011), and Parker et al. (2012), who found that students did not understand the transformation of matter and its relationship to energy during photosynthesis and respiration. This study also revealed some pervasive alternate conceptions relating to carbon cycling and climate change. The most frequently occurring alternate conceptions concerned carbon destroying ozone or otherwise reducing the "quality" of the atmosphere. This supports work done by Groves and Pugh (1999), which found that elementary education majors hold many misconceptions about the Earth's greenhouse effect, and demonstrates that despite the growing concern of climate change, alternate conceptions relating to the greenhouse effect and global warming are still pervasive across fields. This also supports work done in Turkey by Askan and Çelikler (2015), who found that students erroneously associated the greenhouse effect with the thinning of the ozone layer along with other alternate conceptions. Libarkin and colleagues (2015) found holes in the ozone being related to the greenhouse effect as one of their archetypical student mental models. Harris and Gold (2018) showed the relationship between the ozone layer and the greenhouse effect to be one of the stickiest, or most pervasive, misconceptions that continued to be held even after instruction.

Work from Mohan and colleagues (2009) examined a learning progression for upper elementary and high school students for carbon cycling in socio-ecological systems, which consisted of four levels. Their work found that most students leaving high school were not reasoning at the highest level: using chemical models to understand the transformation of elements. Unsurprisingly, this level of reasoning was mainly seen in students who were several years into a STEM major, and not universally among this population. While there were obvious differences in the complexity and nuance in understanding carbon (i.e. its flexible bonding due to four valence electrons, various carbon compounds including carbonate and methane, etc.) as students progressed through STEM disciplines; however, this knowledge is not integrated into a complete or complex conception of the carbon cycle. Rather, our study suggests a lower anchor for future learning progressions as an understanding of the cycling of these elements is driven by hidden or unknown processes, which was identified as level two (Mohan, et al., 2009).

Work by You and colleagues (2017) found that college students have a more interdisciplinary understanding of the carbon cycle than high school students. This study similarly found that students in more interdisciplinary fields have more interdisciplinary mental models of major biogeochemical cycles. In this case, we define interdisciplinary fields as students who reported taking multiple science classes across disciplines (biology, chemistry, geology, physics). Geology and science education students frequently reported taking at least two semesters in each of the aforementioned disciplines. It is not surprising then that students in these groups were the most likely to include all four Earth system spheres in their mental models.

Biology students frequently reported taking courses in biology, chemistry, and physics, but never geology. Chemistry students reported taking courses primarily in chemistry and physics, with some biochemistry students taking some biology courses. Chemistry students, however, often struggled to conceptualize biogeochemical cycles, and though they were familiar with the elements, they had little context for them. Closer inspection reveals that students in chemistry rarely took biology classes that featured anything bigger than the cell, and frequently did not take microbiology or any environmental chemistry course. Thus, lack of interdisciplinary training seems to be related to rather simplistic models of major biogeochemical cycles and a lack of context for how elements that they are so familiar with in reactions operate in nature.

Student Conceptions of the Earth System in Biogeochemical Cycling

The “bio-centric” view of biogeochemical cycling was also observed in student drawings, with students most frequently depicting the biosphere and atmosphere. Representation of the hydrosphere was low for all three biogeochemical cycling, despite the unique role Earth’s ocean and water plays in elemental cycling on Earth. Past work by Shepardson (2009) and colleagues showed that students’ conceptions of the hydrologic cycle were often disconnected, and it very well could be that students' disconnected conceptions of the hydrologic cycle may relate to their difficulty in including the hydrosphere in their mental models of other biogeochemical cycling.

As a whole, drawings often defaulted to cycles that consisted of simple loops. Very few showed multiple pathways and none showed anything resembling a feedback. Occasionally, drawings would show multiple cycles, but often students struggled to connect the two cycles. Rarely, did students depict multiple fluxes between multiple reservoirs and capture some of the complexity inherent to biogeochemical cycles. Thus, the cycles showed a lack of complexity, an important dimension of Earth System Thinking (Scherer, 2016). The important interdisciplinary nature of Earth system thinking was also not captured in the drawings consistent with work done by McNeal and colleagues (2014), which demonstrated that before instruction, students included few interconnections or flows in their drawings, and even after instruction still showed fewer connections than experts show. Rather, drawings tended to focus on atmosphere and biosphere interactions. When the lithosphere was included, it was usually in the form of soil. This is consistent with work by Sibley and colleagues (2007), which found that the chemical transformations associated with the rock cycle and associated parts of the carbon cycle were difficult for students to understand and articulate. The drawings produced by students demonstrate a lack of important aspects of systems thinking in the development of mental models of major biogeochemical cycles.

Limitations

As with any qualitative work, this study was not designed to discover any sort of universal truth (Mason, 2002); rather, it was designed to systematically analyze participants' words and drawings and present results that may be testable in the future or guide future research questions. We approached this work from the idea that knowledge is co-constructed between the researcher and the researched, a constructivist rather than a positivist view, meaning that data gathered in this study is unique to the researchers and the participants (Charmaz, 2014). However, many of the themes and ideas that have emerged have supported existing literature and have broadened our understanding of what students' mental models of major biogeochemical cycles look like. These findings can now be tested and expanded upon using quantitative and mixed methods approaches (Libarkin & Kurdziel, 2002). The findings also provide a solid lower anchor for future work on undergraduate learning progressions using biogeochemical cycles.

This study also examined student mental models independently of instruction, meaning unlike Arthurs's work (2011), which looked at students in a particular class and could attend to misconceptions and preconceptions, this study analyzes a variety of students from a multitude of different fields in different points of their undergraduate career and their mental models. Mental models have several limitations, including the fact that they can be unstable and difficult to access in their entirety (Stone et al., 2000; Libarkin et al., 2003). Through triangulation, we worked to get the fullest picture possible of each participant's mental model; however, we cannot ensure that each is entirely complete. Additionally, we were intentional in speaking only about alternate conceptions and not misconceptions, as we cannot say how tightly these alternate conceptions are held as we did not explore instructional practices.

Conclusions and Future Directions

Though students do hold some common alternate conceptions about biogeochemical cycling, it is more common for students to have very limited conceptions of major biogeochemical cycles, particularly with respect to nitrogen and phosphorous. Student's mental models tend to be based on temporal and size scales that are readily observed by humans. This means that microbes typically are not conceptualized as part of the nitrogen cycle and that long-term controls on the carbon cycle such as carbonate rock formation or silica weathering are not included in students' mental models. This places the burden on the way instructors teach, as the fact that these processes are not readily observed means that students will likely not conceptualize them into their mental models without instruction, and thus is one of the challenges of working with complex systems (Herbert, 2006).

Exacerbating this is the fact that many students will never take courses that may address some of these concepts. Even STEM students rarely take geology classes, and thus do not have the opportunity to conceptualize important parts of biogeochemical cycles. This is problematic as many of the slower or smaller-scale processes are essential for understanding sustainability issues like climate change and ocean acidification as well as the Earth system itself (Jacobson et al., 2000). While the Next Generation Science Standards (Lead States, 2013) may provide some relief to this in the future, more work needs to be done in the realm of post-secondary education to help students of all disciplinary backgrounds build accurate and complete mental models of biogeochemical cycling. The fact that quintessential "hard science" fields are producing students with vast content knowledge but have no context for that knowledge as it relates to the Earth itself is concerning.

Though this study was conducted only a few hours from the coast, students still overwhelmingly excluded the hydrosphere from their mental models. This ocean-blindness results in a lack of interdisciplinary thinking that directly links to systems thinking when approaching biogeochemical cycling (Scherer et al., 2013). The ocean itself is an important sink for carbon dioxide; however, with modern global change that flux of carbon dioxide threatens ocean chemistry and pushes its natural buffering system to its limits (Schlesinger & Bernhardt, 2013). Thus, understanding the ocean and its relationship to biogeochemical cycling is an important aspect of not only systems thinking, but also sustainability.

This study revealed major shortcomings in how undergraduate students conceptualize biogeochemical cycles and the mental models they construct. This study is not simply about students' content knowledge of biogeochemistry, but rather it is about how they contextualize major transfers of matter and the associated chemical reactions that are central to the function of the Earth system. As such, knowledge of these cycles and their context is essential to tackling issues of global change and sustainability and applying policy to the larger complex systems in which they might occur (Herbert, 2006; Oreskes et al., 1994). Future work is needed to understand how to better help students conceptualize these cycles and develop more accurate and complete mental models, particularly in a wide variety of courses that will reach many students.

Rich qualitative data, such as that generated by this study, is especially useful in instrument development (Libarkin & Geraghty Ward, 2011). Future research is needed to validate concept inventories related to both basic biogeochemistry knowledge as well as Earth system thinking, which can complement work done on carbon cycling, climate change (Libarkin et al., 2018), and interdisciplinary thinking inventories (ISACC; You et al., 2017). Additionally, the Next Generation Science Standards (Lead States, 2013) promise more Earth system teaching at the K-12 level, so it is critical that researchers explore how this is influencing student mental models of these cycles in the context of the Earth system. This work provides a good baseline for future work on undergraduate learning progressions involving biogeochemical cycles. Work by Mohan and colleagues (2009) laid out a learning progression for the carbon cycle in socio-ecological systems (which in this case do not include the marine carbon cycle or the carbon cycle on geologic timescales). Our work supports their finding that most high school students are not graduating with the level of biogeochemical thinking expected by the Next Generation Science Standards (2013). Thus, a major challenge exists in understanding further how to integrate biogeochemical and systems thinking in the K-12 curriculum as well as how to best develop undergraduate learning progressions for non-STEM students in introductory courses and STEM students in their major field. Work may be needed in each discipline (non-STEM introductory courses, geology, biology, chemistry) to understand how student thinking progresses in each discipline and to identify the appropriate upper anchors. The study of student conceptions of biogeochemical cycles is an important area of teaching and research as it opens up valuable windows into how students think about systems thinking and sustainability.

References

Ainsworth, S., Prain, V., & Tytler, R. (2011). Drawing to learn in science. *Science, 333*(6046), 1096-1097.

Arthurs, L., (2011). What college-level students think: Student cognitive models of geoscience concepts, their alternate conceptions, and their learning difficulties. In A.D. Feig & A. Stokes (Eds.) *GSA Special Paper 474: Qualitative Inquiry in Geoscience Education Research* (pp. 135-152), Boulder, CO: GSA.

Ackoff, R.L. (1973) Science in the systems age: Beyond IE, OR and MS. *Operations Research, 21,* 664.

Basri, O., Ben-Zvi Assaraf. O., Cohen, C., and Orion, N. (2015). Understanding the Earth systems: Expressions of dynamic and cyclical thinking among university students. *Journal of Science Education and Technology, 24*(6), 761-775.

Breslyn, W., McGinnis, J. R., McDonald, R. C., & Hestness, E. (2016). Developing a learning progression for sea level rise, a major impact of climate change. *Journal of Research in Science Teaching, 53*(10), 1471-1499.

Cardak, O. (2009). Science students' misconceptions of the water cycle according to their drawings. *Journal of Applied Sciences, 9*(5), 865-873.

Charmaz, K. (2014). Constructing grounded theory. Newbury Park, CA: Sage.

Corbin, J., & Strauss A. (2008). Basics of qualitative research: Techniques and procedures for developing grounded theory. Thousand Oaks, CA: Sage.

Creswell, J. & Plano Clark, V.L. (2007). Designing and conducting mixed methods design. Thousand Oaks, CA: Sage.

Creswell, J. & Poth, C.N. (2018). Qualitative inquiry and research design: Choosing among five approaches. Thousand Oaks, CA: Sage.

Delaughter, J.E., Stein, S., Stein, C.A., & Bain, K.R. (1998). Preconceptions abound among students in an introductory earth science course. *Eos, Transactions, American Geophysical Union, 79*(36), 429–436.

Glaser, B.G., & Strauss, A.L., (1965). The discovery of grounded theory. Chicago: Aldine.

Groves, F. H., & Pugh, A. F. (1999). Elementary pre-service teacher perceptions of the greenhouse effect. *Journal of Science Education and Technology, 8*(1), 75-81.

Gunckel, K. L., Covitt, B. A., Salinas, I., & Anderson, C. W. (2012). A learning progression for water in socio-ecological systems. *Journal of Research in Science Teaching, 49*(7), 843-868.

Hadenfeldt, J. C., Neumann, K., Bernholt, S., Liu, X., & Parchmann, I. (2016). Students' progression in understanding the matter concept. *Journal of Research in Science Teaching, 53*(5), 683-708.

Harris, S. E., & Gold, A. U. (2018). Learning molecular behaviour may improve student explanatory models of the greenhouse effect. *Environmental Education Research, 24*(5), 754-771.

Hartley, L. M., Wilke, B. J., Schramm, J. W., D'Avanzo, C., & Anderson, C. W. (2011). College students' understanding of the carbon cycle: Contrasting principle-based and informal reasoning. *BioScience, 61*(1), 65-75.

Herbert, B. E. (2006). Student understanding of complex earth systems. *Special Papers-Geological Society of America, 413,* 95.

Herrmann-Abell, C. F., & DeBoer, G. E. (2018). Investigating a learning progression for energy ideas from upper elementary through high school. *Journal of Research in Science Teaching, 55*(1), 68-93.

Herrera, J. S., & Riggs, E. M. (2013). Identifying students' conceptions of basic principles in sequence stratigraphy. *Journal of Geoscience Education, 61*(1), 89-102.

Hmelo-Silver, C. E., Jordan, R., Eberbach, C. & Sinha, S. (2017). Systems learning with a conceptual representation: a quasi-experimental study. *Instructional Science, 45*(1), 53–72.

Hogan, K., & Fisherkeller, J. (1996). Representing students' thinking about nutrient cycling in ecosystems: Bidimensional coding of a complex topic. *Journal of Research in Science Teaching, 33*(9), 941-970.

Jacobson, M. J., Markauskaite, L., Portolese, A., Kapur, M., Lai, P. K., & Roberts, G. (2017). Designs for learning about climate change as a complex system. *Learning and Instruction, 52*, 1-14.

Kalkan, H., & Kiroglu, K. (2007). Science and nonscience students' ideas about basic astronomy concepts in preservice training for elementary school teachers. *Astronomy Education Review, 6*(1).

Knoll, A. H., Canfield, D. E., & Konhauser, K. O. (Eds.). (2012). Fundamentals of geobiology. Hoboken, NJ: John Wiley & Sons.

Köse, S. (2008). Diagnosing student misconceptions: Using drawings as a research method. *World Applied Sciences Journal, 3*(2), 283-293.

Lavi, R., & Dori, Y. J. (2019). Systems thinking of pre-and in-service science and engineering teachers. *International Journal of Science Education, 41*(2), 248-279.

Ledley, T.S., Haddad, N., Bardar, Ellins, K., McNeal K., & Libarkin, J. (2012). An Earth system science laboratory module to facilitate teaching about climate change. *The Earth Scientist, 28,*19-24.

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

Libarkin, J.C., Beilfuss, M., & Kurdziel, J.P. (2003). Research methodologies in science education: Mental models and cognition in education. *Journal of Geoscience Education, 51(*1), 121–126.

Libarkin, J.C., & Kurdziel, J.P. (2001). Research methodologies in science education: Assessing students' alternative conceptions. *Journal of Geoscience Education, 49*(4), 378–383.

Libarkin, J.C., & Kurdziel, J.P. (2002) Research methodologies in science education: Qualitative data. *Journal of Geoscience Education, 50,* 195–200.

Libarkin, J.C., & Kurdziel, J., (2006) Ontology and the teaching of earth system science. *Journal of Geoscience Education,54* (3), 408–413.

Libarkin, J. C., Thomas, S. R., & Ording, G. (2015). Factor analysis of drawings: Application to college student models of the greenhouse effect. *International Journal of Science Education, 37*(13), 2214-2236.

Libarkin, J. C., & Ward, E. M. G. (2011). The qualitative underpinnings of quantitative concept inventory questions. *Geological Society of America Special Papers, 474,* 37-48.

Lin, C. Y., & Hu, R. (2003). Students' understanding of energy flow and matter cycling in the context of the food chain, photosynthesis, and respiration. *International Journal of Science Education, 25*(12), 1529-1544.

Lincoln, Y.S. & Guba, E.G. (2013). The constructivist credo. Walnut Creek, CA: Left Coast Press.

Mason, J. (2002) Qualitative Researching. London, Sage Publications.

Mayring, P. (2004). Qualitative content analysis. *A Companion to Qualitative Research, 1*, 159-176.

McNeal, K. S., Libarkin, J. C., Ledley, T. S., Bardar, E., Haddad, N., Ellins, K., & Dutta, S. (2014). The role of research in online curriculum development: The case of EarthLabs Climate Change and Earth System Modules. *Journal of Geoscience Education, 62*(4), 560-577.

Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching: 46*(6), 675-698.

NGSS Lead States, (2013). Next generation science standards: For states, by states. Available at http://www.nextgenescience.org/next-generation-science-standards (accessed 7 August 2018)

Norman, D.A. (1983). Some observations on mental models. In D. Gentner & A. Stevens (Eds.), *Mental Models* (pp. 7-14). Hillsdale, NJ: Lawrence Erlbaum Associates.

Oreskes, N., Shrader-Frechette, K., & Belitz, K. (1994). Verification, validation, and confirmation of numerical models in the earth sciences. *Science,263*, (4), 641–646.

Orion, N. (2002). An Earth systems curriculum development model. In V.J. Mayer (Ed.), *Global Science Literacy* (159-168). Amsterdam: Kluwer Academic Publisher.

Parker, J. M., Anderson, C. W., Heidemann, M., Merrill, J., Merritt, B., Richmond, G., & Urban-Lurain, M. (2012). Exploring undergraduates' understanding of photosynthesis using diagnostic question clusters. *CBE—Life Sciences Education, 11*(1), 47-57.

Prokop, P., & Faněoviěová, J. (2006). Students' ideas about the human body: Do they really draw what they know?. *Journal of Baltic Science Education*, (10).

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

Raia, F. (2008). Causality in complex dynamic systems: A challenge in earth systems science education. Journal of Geoscience Education, 56(1), 81-94.

Schlesinger, W.H. & Bernhardt, E.S. (2013). Biogeochemistry: An analysis of global change. Boston: Elsevier

Sell K.S., Herbert B.E., Stuessy C. and Schielack J. 2006. Supporting student conceptual model development of complex earth systems through the use of multiple representations and inquiry. Journal of Geoscience Education

Shepardson, D. P., Wee, B., Priddy, M., Schellenberger, L., & Harbor, J. (2009). Water transformation and storage in the mountains and at the coast: Midwest students' disconnected conceptions of the hydrologic cycle. International Journal of Science Education, 31(11), 1447-1471.

Sibley, D. F., Anderson, C. W., Heidemann, M., Merrill, J. E., Parker, J. M., & Szymanski, D. W. (2007). Box diagrams to assess students' systems thinking about the rock, water and carbon cycles. Journal of Geoscience Education, 55(2), 138-146.

Stokes, A. (2011). A phenomenographic approach to investigating students' conceptions of geoscience as an academic discipline. *Geological Society of America Special Papers, 474*, 23-35.

Stone, A.A., Turkkan, J.S., Bachrach, C.A., Jobe, J.B., Kurtzman, H.S., & Cain, V.S. (Eds.), (2000). *The Science of self-Report. Implications for research and practice*. Mahwah, NJ: Lawrence Erlbaum Associates.

Thomas, G. V., & Silk, A. M. (1990). An introduction to the psychology of children's drawings. New York: New York University Press.

Tripto, J., Ben-Zvi Assaraf, O., Snapir, Z., & Amit, M. (2018). Recurring patterns in the development of high school biology students' system thinking over time. *Instructional Science, 46*, 639–680.

Trumper, R. (2001). A cross-college age study of science and nonscience students' conceptions of basic astronomy concepts in preservice training for high-school teachers. *Journal of Science Education and Technology, 10*(2), 189-195.

Van Meter, P., & Garner, J. (2005). The promise and practice of learner-generated drawing: Literature review and synthesis. *Educational Psychology Review, 17*(4), 285-325.

Vygotsky, L. (1962). Thought and language.(E. Hanfmann & G. Vakar, Trans.) Cambridge, MA: MIT Press

Wilson, C. D., Anderson, C. W., Heidemann, M., Merrill, J. E., Merritt, B. W., Richmond, G, & Parker, J. M. (2006). Assessing students' ability to trace matter in dynamic systems in cell biology. *CBE—Life Sciences Education, 5*(4), 323-331.

Yoon, S. A., Goh, S. E., & Park, M. (2018). Teaching and learning about complex systems in K–12 science education: A review of empirical studies 1995–2015. *Review of Educational Research, 88*(2), 285-325.

Yoon, S., Goh, S. E., & Yang, Z. (2019). Toward a learning progression of complex systems understanding. *Complicity: An International Journal of Complexity and Education, 16*(1).

You, H. S., Marshall, J. A., & Delgado, C. (2018). Assessing students' disciplinary and interdisciplinary understanding of global carbon cycling. *Journal of Research in Science Teaching, 55*(3), 377-398.

CHAPTER 4: DEVELOPMENT AND VALIDATION OF A CONCEPT INVENTORY FOR EARTH SYSTEMS THINKING SKILLS

Introduction

A serious need exists for large-scale studies to examine the wide range of systems thinking teaching practices in the context of geoscience courses. While many studies have been conducted examining teaching practices and systems thinking on the small scale in classroom-based studies, there has not been work done on a national or broader scale to analyze trends in systems thinking instructional practices. Large-scale work can allow the geoscience community to better understand what resources and professional development are needed to help instructors improve their teaching of Earth system science. Initial work with the National Geoscience Faculty Survey to understand systems modeling teaching practices has been conducted to understand broader Earth systems teaching practices and their relation to active learning practices (Lally et al., 2019; Soltis et al., 2019). This work, however, is based on self-reported data, and research must begin looking at how students respond to these teaching practices and how EST skills develop.

Thus, to move research forward, work on the development of instruments to measure students' systems thinking abilities is critical (Orion & Libarkin, 2014). It is unlikely that one instrument can capture all dimensions of the context of Earth systems thinking, especially after the literature analysis of Scherer and colleagues which found that there are at least four main areas of Earth systems thinking based on the current literature (2017). So rather, a host of different instruments will need to be developed and validation studies completed to better assess how students are developing systems thinking abilities. Arnold and Wade (2017) have developed a complete set of systems thinking skills that cut across multiple disciplines that has the potential to serve as an important starting point. Instruments for systems from fields like ecology (Jordan et al., 2014) and paradigms like ill-structured problems (Grohs et al, 2018) have been developed and may be of use. However, there does not exist a research tool, such as a concept inventory, to easily assess students' EST abilities. An instrument of this nature is essential for moving research on EST forward in the form of classroom studies and interventions.

Earth Systems Thinking

Systems thinking, as used in this study, is an understanding of the circular nature of what happens in the physical world, and the ways that individual systems and their components (sub-systems) relate to and interact with each other. It is an understanding that a system is an indivisible whole, and much more than the sum of its parts (Ackoff, 1973). In the geosciences, qualitative work by Stokes (2011) has demonstrated that the highest level of understanding about the physical world is the ability to think about the Earth as a dynamic system. However, previous work has demonstrated that students have conceptual difficulties with systems thinking and understanding the dynamic nature of systems, particularly when the matter in these systems is not observable or readily apparent (McNeal et al., 2014; Orion, 2002; Sibley et al., 2007). Students also have difficulty when the instructional context of the system is not directly relevant to students' past experiences (Wilson, 2006). Thus, a major area of research in the field of geoscience education is the development of Earth systems thinking.

Earth systems thinking skills are also essential for understanding issues of sustainability and the development of solutions for important environmental issues. A report by the National Research Council (2000) outlined eight major challenges facing humanity, one of them being human alterations to major biogeochemical cycles. Other challenges such as biological diversity and ecosystem functioning, climate variability, hydrologic forecasting, infectious disease and the environment, institutions and resource use, land-use dynamics, and reinvention of the use of materials are directly related to complex environmental systems (NRC, 2000; Herbert, 2006).

These complex systems come with fundamental challenges to student understanding and include the conceptualization of the natural Earth environment as composed of systems, the characterization and explanation of the complex nature of Earth systems, and the application of conceptual and scientific models of Earth systems to support problem-solving and the development of effective environmental policy (Herbert, 2006; Oreskes et al., 1994).

Scherer and colleagues (2017) reviewed the Earth systems thinking literature and synthesized four major Earth system thinking conceptual frameworks: Earth system perspective, Earth system thinking skills, complexity science, and authentic complex Earth and environmental systems. The Earth system perspective emphasizes the interconnections between the major Earth spheres. Systems thinking abilities relate to conceptualizing the Earth system as a whole. The Earth system thinking skills conceptual framework emphasizes the transformation of matter in Earth cycles and the thinking abilities related to identifying and organizing system components. The complexity science conceptual framework emphasizes the scientific study of complex systems and systems thinking abilities related to recognizing complex system characteristics. Finally, the authentic complex Earth and environmental systems framework emphasizes the knowledge of a specific complex near-surface Earth system or phenomenon and systems thinking abilities related to reasoning about the specific system or phenomenon

Concept Inventories

A concept inventory (CI) is a valuable tool for use in both research and teaching for assessing what students know and diagnosing conceptual difficulties or misconceptions. CIs take the form of multiple-choice assessments that differ from other multiple-choice tests by emphasizing reliability and validity. This means that student responses to the CI can be confidently assumed to reflect their knowledge of the topic at hand, and thus the results are repeatable and measure what they claim to measure. This allows CIs to be a valuable research and teaching tool (Libarkin, 2008). In the geosciences several concept inventories already exist, covering areas like general geoscience content (Libarkin & Anderson, 2005), oceanography (Arthurs, Hsia, & Schweinle, 2018), and climate change (Jarrett. Ferry, & Takacs, 2012; Walker & McNeal, 2013; Libarkin et al., 2018).

Reliability

The construct of reliability refers to consistency or stability of results, in other words, whether the assessment or data collection tool consistently captures the same information. Though tools or assessments may be called reliable, reliability actually refers to the results, not the tool itself. While results must be reliable, reliability itself is not sufficient if the results are not also valid (Reynolds et al., 2010). When considering the development of an instrument, there are several methods of assessing reliability with a reliability coefficient. Coefficients may be derived from the administration of the same test or tool on different occasions (i.e. test-retest reliability), coefficients based on the administration of parallel forms of the instrument or test (alternate-form reliability), and coefficients derived from a single administration of a test (internal consistency coefficients) (Reynolds et al., 2010). Often in quantitative educational research, measures of internal consistency are most commonly applied due to the fact that they can be done fairly quickly and easily and require only one administration of an instrument.

Among internal-consistency estimates of reliability, several common statistical methods exist. Split-half reliability involves administering a test or other instrument and dividing the test into two equivalent parts. The results of the first half are then correlated with the second half by calculating the Pearson product-moment correlation. In this case, the Spearman-Brown formula should be applied as a correction to effectively “put the test back together” as the reliability correlation does not take into account the reliability of the test when the two halves are combined. More commonly used is Coefficient alpha or Cronbach’s alpha (1951) and Kuder-

Richardson Reliability (KR-20) (1937). Both of these approaches examine the consistency of responding to all individual items on an instrument or a component of the instrument. Meaning that estimates produced here are analogous to the average of all possible split-half coefficients. Due to this fact, these estimates are sensitive to content heterogeneity, or the degree to which the instrument measures similar constructs (Reynolds et al., 2010). In this case, if the internal structure of an instrument is known to measure multiple constructs, these estimates are applied to items related to a specific construct that the instrument is designed to measure. A composite estimate of reliability then be generated. The reliability of composite scores is typically greater than the measures that contribute to the composite (Reynolds et al., 2010).

KR-20 is one of multiple reliability equations presented by Kuder and Richardson (1937) but is one of the most commonly used estimates. It is applicable when items are dichotomously scored as right or wrong (0 or 1). Cronbach’s coefficient alpha (1951) is a more general form of KR-20 that deals with items that may produce multiple values (0,1,2, etc.). Due to this fact, coefficient alpha has become the preferred statistic for calculating reliability (Keith & Reynolds, 1990). This is especially true on surveys, which tend to have items that are not dichotomous. Typically, researchers look for a Cronbach’s alpha value of .70 or above, though this number may be considered somewhat arbitrary. Despite its frequent use, there has been some criticism that Cronbach’s alpha is unrelated to the internal structure of the test and its usefulness is limited (Sijtsma, 2009).

Validity

Like reliability, validity is a fundamental psychometric property that is a major concern in research. Validity is a somewhat more subtle and nebulous concept, and essentially describes the closeness of what we intend to measure and what we actually measure. In the sense of employing an instrument, validity refers to the appropriateness or accuracy of the interpretation of the score or result (Reynolds et al., 2010). Reliability is a necessary component for validity, but it is not sufficient on its own, in other words, an instrument can produce reliable results, but may not be measuring what it is intended to measure. However, an instrument that does not produce reliable results can never produce valid interpretations. Typically, making a case for validity involves gathering several lines of evidence of various “types” of validity (Reynolds et al., 2010).

There exist several major types of validity when it comes to instruments: content validity, criterion-related validity, and construct validity (Reynolds et al., 2010). Content validity is particularly important in concept inventories and educational tests as it involves how adequately the test samples the content area of an identified construct. It is often evaluated based on professional judgments by experts on the appropriateness of the content. Criterion-related validity is associated with the examination of relationships between the test and external variables that are thought to be direct measures of the construct. Determining the relationship between the test and external variables are completed quantitatively through correlation or regression analyses. Construct validity involves integrating evidence that relates to the meaning or interpretation of test scores. This evidence can be accrued using a variety of research strategies and designs such as exploratory factor analysis and Rasch analysis. Of increasing interest is the concept of cultural validity, which refers to the effectiveness in which an instrument or assessment addresses the sociocultural influences that shape thinking, though this type of validity is not as frequently addressed as the other types (Solano-Flores and Nelson-Barber, 2001).

The classification listed above has been widely accepted by researchers; however, in the ’70s and ’80s measurement professionals began moving towards viewing validity as a unitary construct, in which the different “types” of validity, in reality, are different ways of collecting evidence to support validity.   Thus, validity is now viewed as a unidimensional construct with five categories of evidence that are related to the validity of interpretation: Evidence based on test content, evidence based on relations to other variables, evidence based on internal structure, evidence based on response processes, and evidence based on consequences of testing (AERA et al., 1999; Reynolds et al., 2010). Despite differences in understanding and defining validity, traditional methods of examining validity employ three major types of validation studies: content validation, criterion-related validation, and construct validation (Crocker & Algina, 2008).

The purpose of a content validation study is to assess whether items on an instrument adequately represent a performance domain or construct of specific interest. At the minimum, this should entail the four steps: 1. Defining the performance domain of interest, 2. Electing a panel of qualified experts in the content domain, 3. Providing a structured framework for the process of matching items to the performance domain, 4. Collecting and summarizing the data from the matching process (Crocker & Algina, 2008). Content validation is most often employed with achievement tests, so the performance domain is often defined by some sort of list of objectives or a table of specifications. Also, associated with content validity is the idea of face validity, which typically refers to the extent to which items appear to measure a construct that is meaningful to a layperson. In other words, does the instrument look like it measures what it claims to measure (Crocker & Algina, 2008)?

Criterion-related validation is employed when a test user is looking to make inferences from test scores to examinee behavior on a performance criterion that cannot be directly measured by a test. This typically breaks down into two types of criterion-related validation: predictive and concurrent. Predictive validity refers to the degree to which test scores predict criterion measurements that will be made in the future. For example, the SAT scores have some degree of predictive validity with respect to college grade point average (thus the justification for using SAT scores in making admissions decisions). Concurrent validity refers to the relationships between test scores and criterion measurements made at the time the test was given (Crocker & Algina, 2008). An example of this may be a teacher taking a test designed to measure their knowledge of pedagogy and immediately after that being observed teaching. A positive correlation between their test score and performance during the observation would suggest using the test in place of the more time-intensive observation.

Construct validation evidence is typically assembled through a series of studies. Correlational studies may be conducted to relate scores on a given test or instrument and some other measure of performance. Often multiple regression is used so that contributions of the construct of interest to variance in the criterion can be assessed in relationships to the contribution of other variables. Factor analysis is another approach that may be used to determine whether item responses cluster together in patterns that are reasonable when considering the theoretical structure of the chosen construct. In this case, the factor discerned by factor analysis can be considered a construct. Now, these constructs can be considered in relation to the construct the test is meant to measure to provide evidence for or against validity (Crocker & Algina, 2008). In instrument development, validity is an important consideration in all parts of the process, including the very beginning and through a rigorous process of development.

Classical Test Theory

Classical test theory is a traditional quantitative approach to test a scale's reliability and validity based on its items (Cappelleri et al., 2014). Classical test theory is based on the premise that each observed score (X) is a combination of an underlying true score (T) on the concept of interest and random error (E). Thus *observed score (X)= true score (T) + error (E)*. Thus, true scores (which cannot be observed) quantify values on whatever is intended to be measured, in this case, EST abilities. Classical test theory assumes that item responses are coded so that higher response scores reflect a greater understanding of the concept of interest. Another assumption of classical test theory is that random errors are normally distributed (thus the expected value of random fluctuations is assumed to be 0) and uncorrelated to the true score (Crocker & Algina, 2008).

In classical test theory, the means and standard deviations of each item are important indicators of the quality of the items. In general, the higher the variability of the item scores and the degree of closeness of the mean score to the median, the better the item will perform in the population of interest. Item difficulty can be derived from a *z*-score metric or a proportion of correct responses. Item discrimination, or how well an item can differentiate between high and low performers can be determined by partitioning respondents into overall high and low performing groups and comparing the proportion of correct and incorrect responses and subtracting the proportions (i.e. if 80% of the upper group and 30% of the lower group select the correct response, the item discrimination index would be calculated as (.80-.30=.50). Discrimination can also be determined based on how well an item correlates with the sum of the remaining items on the same scale. Item discrimination and difficulty can be plotted together to illustrate how well items in a scale span across the range of difficulty as well as how well each item represents the concept (Cappelleri et al., 2014).

Dimensionality, or the extent at which the items measure a hypothesized concept distinctly, can be evaluated through factor analysis. Exploratory factor analysis is used to generate hypotheses about the structure are the data when there is uncertainty as to the number of factors being measured. Exploratory factor analysis is also useful in determining items to prune to cut because they contribute little to the presumed underlying factor or construct. Exploratory factor analysis should be complemented by confirmatory factor analysis in later stages of instrument development, by imposing the hypothesized structure from the exploratory factor analysis on new data to confirm that structure (Cappelleri et al., 2014).

Item Response Theory

Item response theory (IRT) is a collection of measurement models that work to explain the connection between observed item responses on a scale and an underlying construct. IRT develops models that are mathematical equations that describe the association between participant's responses on a latent variable and the probability of a particular response to an item. Parameters of items are estimated directly using logistic models rather than the proportions of classical test theory. In IRT, items responses may be evaluated on multiple parameters: in the simplest case, just difficulty (Rasch model), or a two-parameter model that adds an item discrimination parameter to the model (Cappelleri et al., 2014). IRT also differs from classical test theory in that it correlates test and item scores based on the mathematical relationship between abilities and item responses (Hambleton, Swaminathan, & Rogers, 1991). This means that a high ability respondent would be predicted to have a higher probability of answering an item correctly than a low ability respondent on an item, whereas CTT assumes that item difficulty is group dependent. Thus, IRT assumes that difficulties of items are not specific to the group of respondents who take a test and that raw scores are dependent on test difficulty, but the estimated ability is independent of test difficulty. IRT has been shown to provide more stable estimates of item difficulty among samples, more stable internal consistencies, and significantly fewer measurement errors than the CTT approach (Magno, 2009).

The Rasch model is often used in IRT studies and is expressed both in terms of the item and the instrument. This model focuses on developing a model of the probability of observed responses. This differs from other models that may develop a model of the responses themselves (Wilson, 2005). The item characteristic curve (ICC), which can be thought of as the probability of choosing the correct response to an item. These curves indicate which items are harder and which items are better discriminators of the attribute of interest by looking at individual items. Item information is a component that provides an assessment of the precision of measurement of an item in terms of distinguishing or discrimination amongst subjects across different levels. Higher information implies more precision, and the amount of item information (precision) decreases as the item difficulty differs from the respondent's attribute level. The person separation index is used as a reliability index in Rasch measurements, as reliability reflects how precisely the scores separate or discriminate among individuals (Cappelleri et al., 2014; Hambleton et al., 1991).

Wright maps (Wright & Masters, 1982) visually represent the data and demonstrate how well item location matches the respondent location. The item location is the item difficulty or item scale value, whereas the respondent location is the respondent ability. In the model itself, the difference between the person and item location determines the probability of the person choosing the correct answer on an item. Thus the Rasch model can be defined as the probability of the item response for item *i* is X_i and is a function of the respondent location (θ) and the item location (δ_i). This can be written in the equation:

$$\text{Probability}(X_i = 1 \mid \theta, \delta_\mathrm{i}) = \frac{e^{(\theta - \delta\mathrm{i})}}{1 + e^{(\theta - \delta\mathrm{i})}}$$

IRT has two major assumptions. Monotonicity is the assumptions of correct model specification and is met if the probability of selecting correct responses increases with the person' location on the individual's total score. Unidimensionality for items in a scale is the other major assumption. This can be satisfied by fitting a factor analytic model to the data to determine the extent to which there is sufficient unidimensionality (no residual correlations greater than or equal to .20) (Cappelleri et al., 2014; Hambleton et al., 1991). WinSteps (Linacre, 2019) was used for all IRT analyses in this study.

Study Goals and Research Questions

The overarching goal of this study is to validate an instrument to measure systems thinking skills in the context of the Earth system in undergraduate students and the public. This study also aims to use statistical techniques such as factor analysis to understand the latent structure of EST abilities. That is, do items related to the cycling or movement of materials through different components of the Earth system relate to those that have to do with systems concepts such as fluxes and feedbacks? We hypothesize that this CI will not be unidimensional and that participant responses to items will correspond to frameworks proposed by Scherer and colleagues (2017). Embedded in the survey are also items concerning feedback loops presented as both "positive" and "negative" feedbacks as well as "reinforcing" and "balancing." This will give practitioners who use the instrument options in the wording used concerning feedback loops to match their instruction and will also allow us to understand if language impacts conceptions of feedback loops or the reliability and validity of the instrument. Thus, in addition to creating a validated, research-grade instrument our research questions are (1) What are the dimensions or constructs that makeup Earth system thinking skills? (2) How does wording affect the performance of tasks relating to feedback loops?

Methods

Instrument Design

EST items were written in a single response format (*n*= 25), in which participants were directed to choose the best answer. Multiple steps were taken to ensure validity and reliability during CI construction (Table 2). The content of the test was determined through the review of literacy documents (NGSS Lead States, 2013) as well as on the previous study on systems thinking and biogeochemical cycles (Soltis et al., In Review). Items were initially written by the lead author in accordance with a concept table based on the literature (Table 1). Work by Scherer and colleagues (2017) was especially instrumental in developing items relating to two major perspectives: interconnections between Earth system components (Earth system perspective), applying systems concepts to transformations of matter (Earth system thinking skills perspective), and feedback loops (complexity sciences perspective). This work also corresponds to work on teaching practices relating to Earth systems thinking identified by Soltis et al. (2019) (Figure 1). Items were written using guidelines from Libarkin (2008) and Libarkin and Anderson (2006). Table 2 overviews modes of assessing various modes of reliability and validity used throughout the process of this study.

Table 4.1 Framework used for developing the CI. Asterixis indicate questions that fall within two frameworks.

Framework	**Description**	**Skill**	**Questions**
Earth System Perspective	High-level interconnections between major Earth spheres	Recognizing interactions between Earth system components, Applying basic Earth system science concepts	1, 4, 5, 7, 8, 11, 15, 18, 19*, 20*
Earth System Thinking Skills	Transformation of matter in Earth cycles and specific system thinking concepts	Applying Systems thinking concepts and vocabulary to Earth system processes	2, 3, 6, 9*, 10, 13, 16, 17, 21
Complexity Sciences	Scientific study of complex systems	Identifying and making predictions about feedback loops, using systems diagrams	9, 12, 14, 19*, 20*, 22, 23, 24, 25
Authentic Complex Earth and Environmental Systems*	Knowledge of a specific Earth system or phenomenon	Context-Specific	*

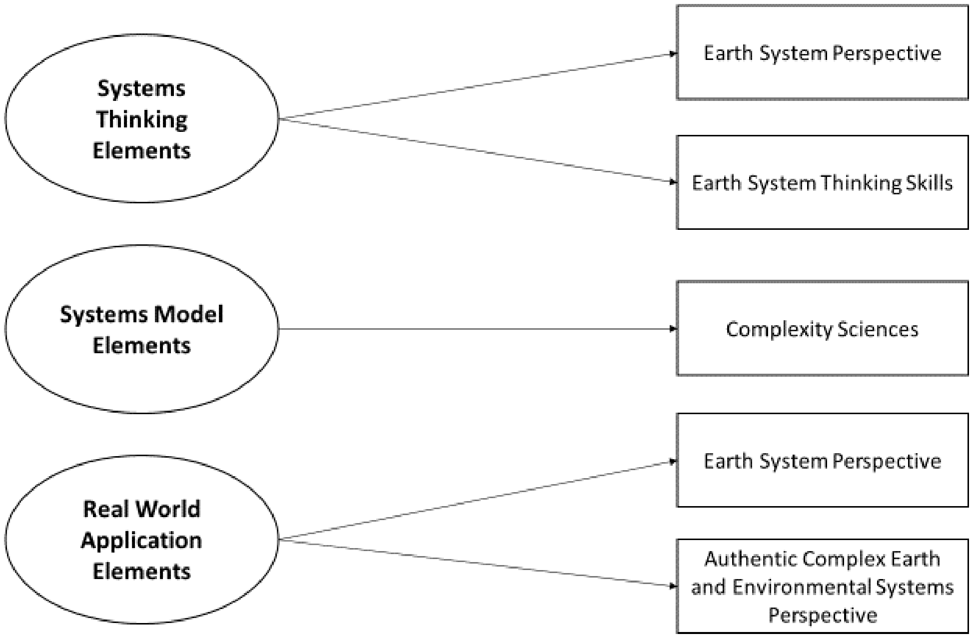


Figure 4.1 Schematic showing the relationship between self-reported teaching practices (circles) from work by Soltis et al. (2019) and Earth systems thinking framework as described in the literature (rectangles) by Scherer et al. (2017).

Upon development, items were then reviewed by other members of the research team before being sent out for expert review. Five experts who have completed research on systems thinking skills in the context of Earth science agreed to review and comment on the items and the items were revised accordingly. The instrument was piloted in conjunction with a think-aloud interview with seven undergraduate students to ensure that the intended population understands the instrument and what it is asking. Participants for this portion of the study were compensated for their time with a $20 amazon gift card. The CI was then revised accordingly.

Table 4.2 Validity and reliability measures utilized (based on work by Libarkin et al., 2018).

Validity/Reliability	Description	Approach used in this study
Content validity-test blueprint	Alignment of item content with discipline, expert, and student views of the domain	(1) Test developed with table of specification (2) Review of literature (3) Qualitative interviews with undergraduate students (4) Expert review of items
Content validity-item content	Examination of items to ensure content spans latent trait	(1)Item content and structure was discussed among researchers and items were revised until consensus was reached (2) Expert review of items
Content validity-item appropriateness	Ensuring items are usable for all target populations	Readability: Flesch reading ease is 53.5, indicating CI is easily understood by 9th graders Think-aloud interviews with undergraduates participating in pilot
Content validity- design principles	Extent to which items are written in accordance with research-based best practices	Iterative revision based on item-writing standards (Libarkin, 2008) until team reached consensus.
Reliability	Precision of trait level estimates Internal consistency estimates	Rash-based estimate of person separation reliability Cronbach’s alpha for each scale
Construct validity-dimensionality	Evaluation of whether items measure one latent trait	Rasch dimensionality analysis using PCA of residuals
Construct validity- item analysis	The extent to which items function as intended and cover the full latent trait range of a population	INFIT and OUTFIT statistics and IRCC curves used to evaluate item performance. Examination of trait map. Removal of items balanced with need to maintain reliability
Cultural validity	The extent to which items perform equally across subgroups within the population	Between-group comparisons via DIF analysis

CI Participants

After piloting and think-aloud interviews with undergraduate students were conducted, CI participants were recruited from an online crowdsourcing system (MTURK) based on MTURK documentation of reliable performance completing other MTURK tasks. MTURK samples are representatively similar to traditional research subject pools in terms of race, gender, age, and education (Paolacci et al., 2010). Workers were prescreened to ensure only those with good performance records completed the study. Workers were compensated for completing the study and compensation for task completion was within MTURK standards for similar tasks. The target of this study was to recruit 1,000 individuals, 250 participants for an initial pilot to screen items and to perform exploratory factor analysis to understand the dimensionality of items. After necessary revisions were made an additional 750 responses were collected to perform additional confirmatory factor analysis and Rasch analysis. MTURK directed participants to the Qualtrics survey where they were asked to provide basic demographic information (age range, gender, education level) and then completed the CI multiple-choice based assessment. Basic demographic information from all stages of the MTURK study can be found in table 3.

Table 4.3 Demographic information of MTURK participants. Values in parentheses indicate percent.

Category	Response	Pilot (n=252)	Revised (n=752)	Total (n=1004)
Gender	Male	162 (64.29)	395 (52.53)	557 (55.48)
	Female	88 (34.92	343 (45.61)	431 (42.93)
	Choose not to Identify	2 (0.79)	14 (1.86)	16 (0.02)
Age	18-25	39 (15.48)	72 (9.57)	111 (11.06)
	26-35	115 (45.63)	307 (40.82)	422 (42.03)
	36-45	49 (19.44)	212 (28.19)	261 (30.00)
	46-55	23 (9.13)	95 (12.63)	118 (11.75)
	55-65	17 (6.75)	48 (6.38)	25 (2.49)
	Over 65	7 (2.78)	17 (2.26)	24 (2.39)
	Choose not to respond	2 (0.79)	1 (0.13)	3 (0.30)
Education	Some Highschool	1 (.40)	4 (0.53)	5 (0.50)
	Highschool Diploma	32 (12.70)	75 (9.97)	107 (10.66)
	Some College	62 (24.60)	198 (26.33)	260 (25.90)
	Undergraduate Degree	109 (43.25)	344 (45.74)	453 (45.12)
	Advanced or Professional Degree	43 (17.06)	126 (16.76)	130 (12.95)
	Choose not to respond	5 (1.98)	5 (.66)	10 (1.00)

Analysis

The CI was analyzed using both classical test theory and item response theory (using a Rasch model). The analysis was completed using the initial pilot (*n*=250) for a basic quality check and to see if more items were needed based on item difficulty. Once any changes were made to the instrument, it was tested again (*n*=750). No items were eliminated between trials, so the pilot data were combined with the data from the second administration. Feedback related items were written using the terminology “positive” and “negative feedbacks,” but alternate items were given at the end of the test using the terminology “balancing” and “reinforcing” feedbacks to provide options to future users of the instrument and to see if the phrasing influenced the results of the CI.

Classical Test Theory

For the CTT analysis, excel and SPSS were used to analyze the data. The standard statistical methods used in this study included item difficulty, item discrimination, and the point biserial coefficient (Table 4). Item difficulty is the fraction of correct responses based on the total number of responses, this results with values ranging from 0-1 with lower items being more difficult. It is recommended that tests consist of a range of difficulties (Barder et al., 2006), with recommended values typically ranging from 0.2-0.9 (Bardar et al., 2006; Ding et al., 2006). Item discrimination assesses how well items distinguish between high and low performers (Ding et al., 2006). In this study, discrimination was determined by dividing participants into thirds based on their total score and looking at differences in performance (number of correct answers divided by the number of participants in the group) between high performers and low performers. Item discrimination should not be negative, and the typical acceptable minimum value is 0.3 (Doran, 1980). The point biserial coefficient looks at the correlation between the performance of individual items with the test as a whole (Ding et al., 2006). Typically, items with a point biserial coefficient of .2 or greater are interpreted as acceptable (Kline, 1986).

Table 4.4 Statistical measures used in classical test theory and values used in this study. Modified from Jarrett, Ferry, & Takacs (2010).

Name of Measure	What it is	Recommended Values
Item Difficulty	Fraction of correct responses	0.2-0.9
Item Discrimination	Assesses how well items differentiate between high and low performers	≥0.3
Point Biserial Coefficient	Consistency of individual items with the test as a whole	≥0.2
Cronbach's Alpha	Measure of internal consistency; estimate of reliability	≥0.7

Cronbach's alpha, which is an estimate of internal consistency, was utilized to calculate reliability. Typically, most concept inventory researchers set 0.7 as the acceptable value for Cronbach's alpha (Nunally, 1978; Litwin, 1995). However, since concept inventories tend to not be homogenous tests, tests of internal consistency can seriously underestimate reliability (Miller, 1995). Due to this fact, some researchers have given 0.6 as the minimum acceptable value for the equivalent Kuder-Richardson 20 (Grolund, 1993; Anderson et al., 2002). To test the dimensionality of the concept inventory and understand how many latent factors were being measured, an exploratory factor analysis was completed using SPSS.

Item Response Theory

Rasch analysis for IRT was performed using Facets (Linacre, 2019). Dimensionality was evaluated with principal components analysis of residuals (for each individual scale), which is a standard approach in Rasch analysis. Rasch fit statistics, INFIT and OUTFIT, were used to examine item functions. Infit is the inlier-sensitive fit, meaning that it is more sensitive to responses to items with difficulty targeted to the location of that person, and outfit is the outlier-sensitive fit, meaning it is sensitive to responses to items with difficulty far from person location.

Infit and outfit are measured using mean-square fit statistics that show the size of randomness, with an expected value of 1.0. Mean-squares near 1.0 indicate little distortion of the measurement systems, values greater than 1.0 indicate unpredictability (unmodeled noise, data underfit the model), and values less than 1.0 indicate that observations are too predictable (redundancy, data overfit the model). Standardized fit statistics for infit and outfit are *t*-tests of the hypothesis “Do the data fit the model?” and are reported as z-scores. These values have an expected value of 0, with items less than 0 being too predictable and more than 0 lacking predictability. Typically, values between -1.9 and 1.9 indicate reasonable predictability (Linacre, 1999).

Table 4.5 Values and interpretation for infit and outfit statistics (Linacre, 1999)

	Value	Implication for Measurement
Mean-Square	>2.0	Distorts or degrades the measurement system.
	1.5-2.0	Unproductive for measurement, but not degrading
	0.5-1.5	Productive for measurement
	<0.5	Less productive for measurement, but not degrading. May produce misleadingly high reliability and separation coefficients
Standardized	≥3	Data likely does not fit the model. But, with large sample size, substantive misfit may be small
	2.0-2.9	Data noticeably unpredictable
	-1.9-1.9	Data have reasonable predictability
	≤-2	Data are too predictable.

Differential item functioning for gender was used to determine possible bias in items. Wright maps were generated and, particularly for the pilot, were analyzed to determine where a mismatch between item location and person location existed to determine if additional items geared toward a particular ability needed to be added. Item characteristic curves were also generated. An item characteristic curve plots the probability of a correct response as a function of the ability of an individual (Linacre, 1999)

Results

Pilot Data

Pilot Data was collected from 252 individuals. Attention checks within the CI were used to identify participants who may not have been giving the instrument their full attention. Participants who failed an attention check were not considered in the pilot (n=10). Meaning 242 participants' data were considered in the analysis of the pilot data. For the pilot, data on the rephrasing of feedbacks (balancing and reinforcing) were not considered. The purpose of the pilot data was to initially screen the items using metrics from both classical test theory and item response theory to get an overview of item difficulty and discrimination (CTT) and to examine the relationship between items and ability of participants (IRT). At this stage of the development, the goal was not to throw out any items; rather it was to identify problematic items and to understand what additional items may have been needed. Measures of Item difficulty (figure 3) indicated that on average the test had an acceptable level of difficulty of .33 when compared to the minimum acceptable value of .3. While this value is acceptable, it does indicate that the CI is difficult. The average item discrimination (figure 4) was near but below the minimum acceptable value of .3, indicating that the test was not discriminating between high and low performers as well as it could. The average point biserial coefficient (figure 4) of .096 was well below the minimum acceptable value of .20, meaning that at this stage the individual items were not predictive of overall performance.

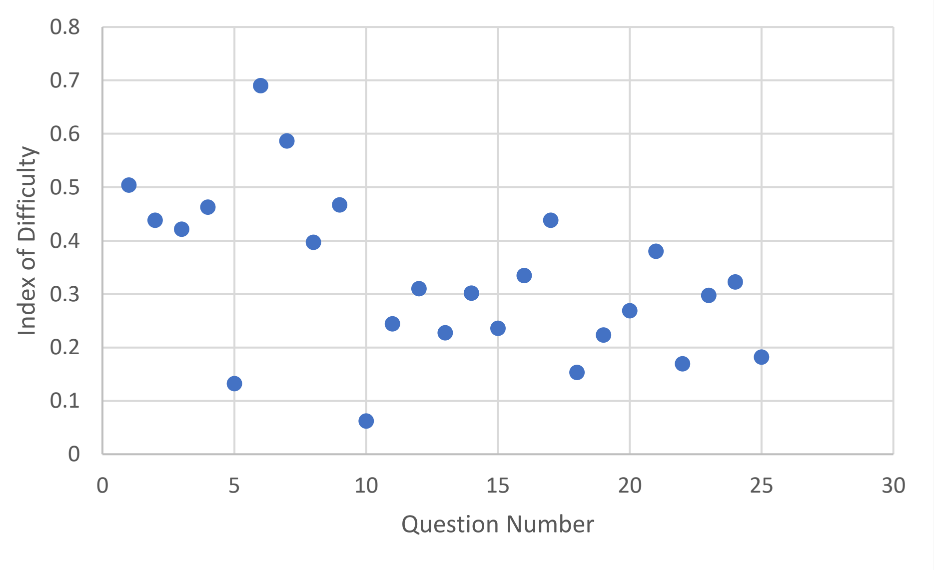


Figure 4.2 Item difficulty by question from the pilot data. Average Index of Difficulty was .33.

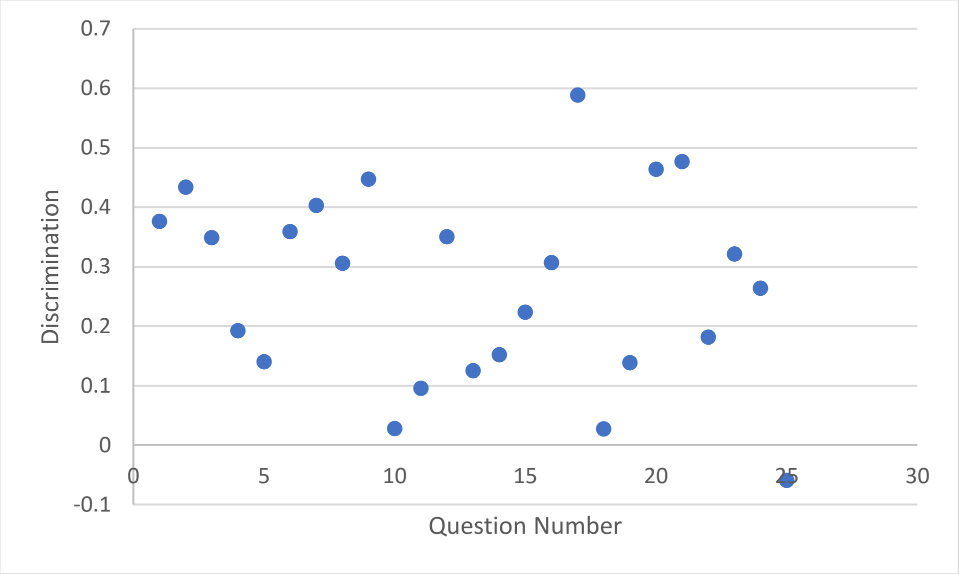


Figure 4.3 Item discrimination by question from the pilot data. Average Index of Difficulty was .27

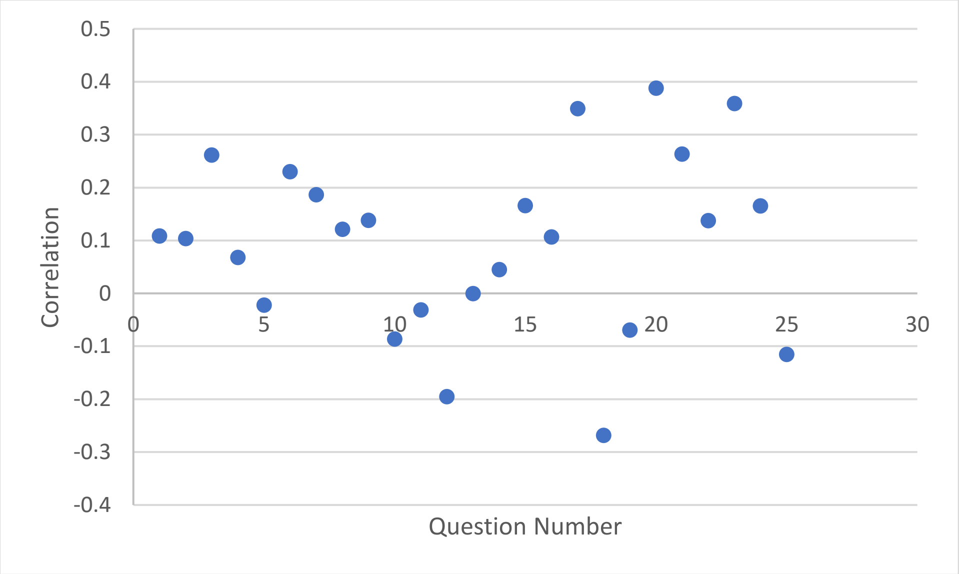


Figure 4.4 Point Biserial Coefficient by question from the pilot data. Average Index of Difficulty was .096.

According to principles of IRT, participants and items were plotted on a Wright Map and graphically depicted in figure 5, this revealed that the mean difficulty of items was higher than the mean ability of participants. Or in other words, the items were too difficult. This data combined with the CTT analysis demonstrated a need for several easier items to be added to the CI. Dimensionality was also assessed using both CTT and IRT (exploratory factor analysis and principal components analysis of residuals), both revealed that with this group of participants the data were unidimensional.

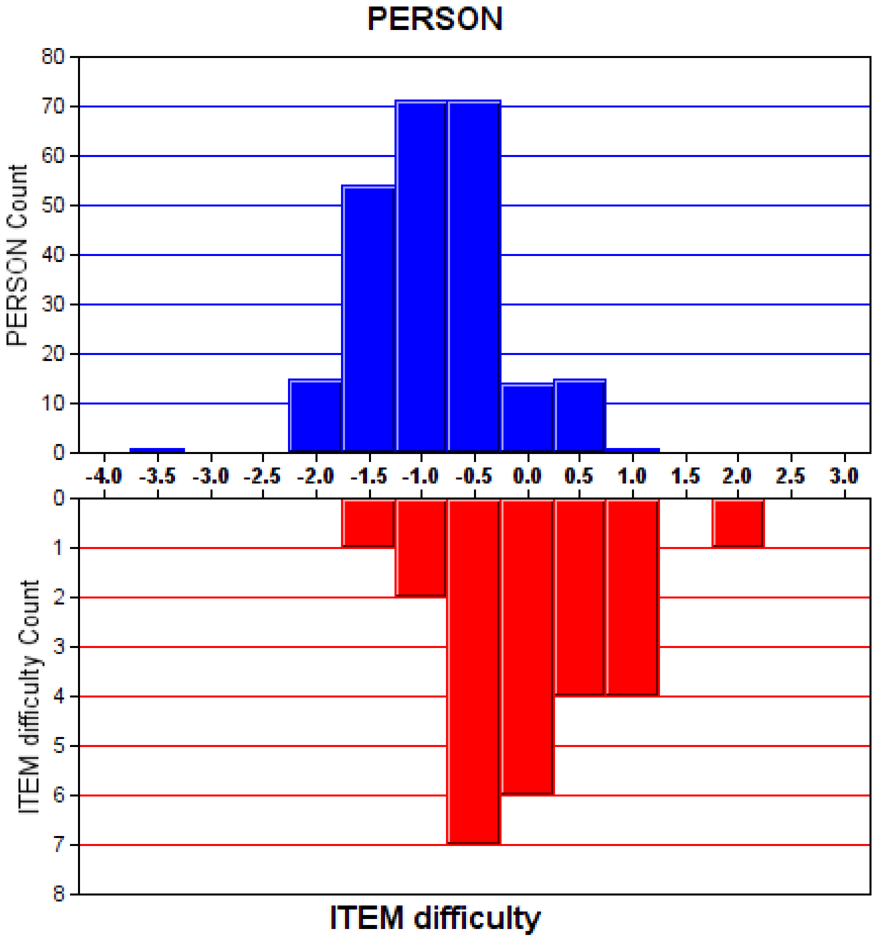


Figure 4.5. Graphical representation of a Wright Map showing relationships between participants' abilities and item difficulty after the pilot.

Full Implementation

Since no items were eliminated and only new items were added, data from the second iteration of the instrument and the pilot data were combined for analysis. Based on exploratory factor analysis and principal components analysis of residuals, the data was still shown to be unidimensional.

Classical Test Theory

As with the pilot data, item difficulty, discrimination, and point biserial coefficients were calculated for each item. The average item difficulty of all items (figure 6) was .36, above the accepted minimum value of 0.30. Average item discrimination (figure 7) was .27, approaching the acceptable minimum value of .30. The average point biserial coefficient (figure 8) was .27, above the acceptable minimum value .20. Thus, the addition of new items seemed to help the metrics of the entire test. However, it was still necessary to examine each item individually to compare these statistics. Upon examining each item (appendix B), it was clear that six items (5,10, 11,14, 19, 25) failed to meet the minimum threshold for all three statistics. Upon removing these items from analysis, average item difficulty rose to .40, average item discrimination rose to exceed the acceptable minimum value at .33, and the point biserial correlation rose to .31. In terms of reliability measures, the full implementation of the CI had a Cronbach's alpha value of .583 before items were removed. Once items were removed Cronbach's alpha climbed to .644.

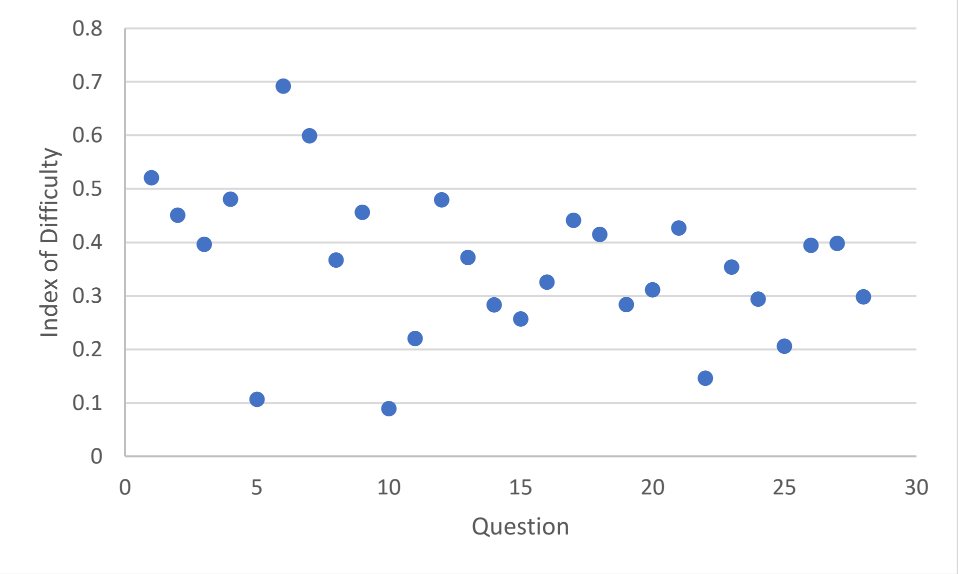


Figure 4.6 Item difficulty by question from the full administration. Average Index of difficulty was .36, with items removed it was .40

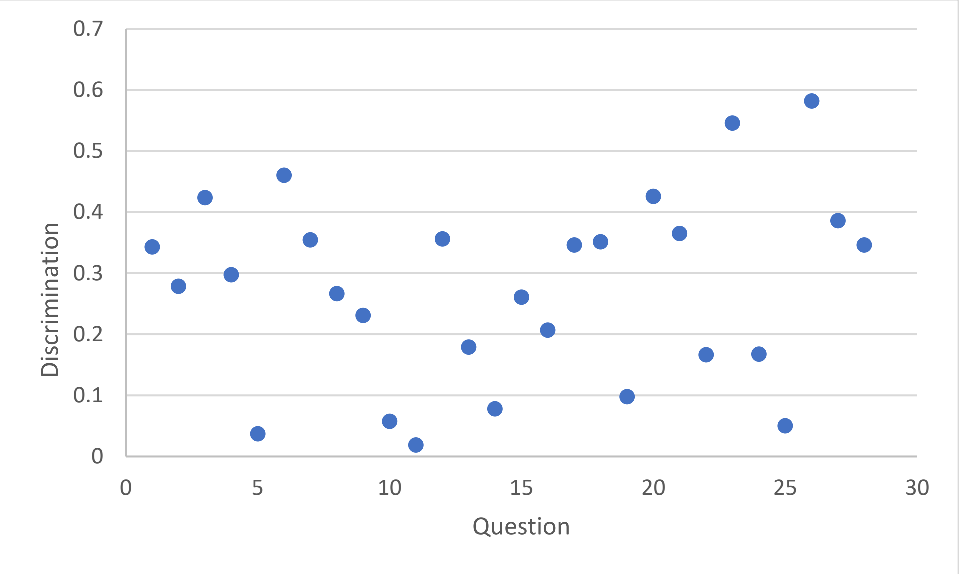


Figure 4.7 Item discrimination by question from the full administration. Average item discrimination was .27, with items removed it was .33

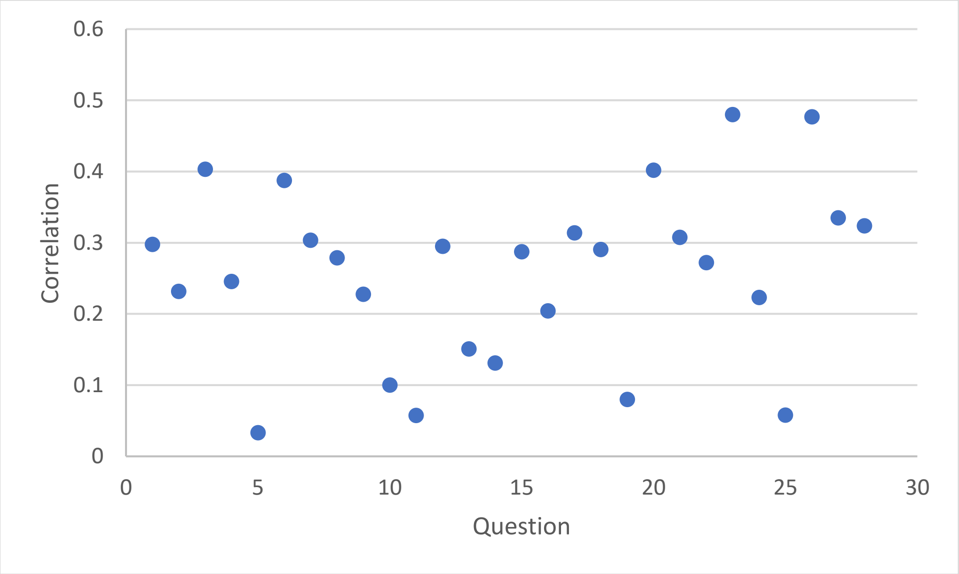


Figure 4.8 Point biserial coefficient by question from the full administration. Average point biserial coefficient was .27, with items removed it was .31

Item Response Theory

The Wright map produced revealed the additional items did help bring the mean of items closer to the mean of ability, though it still was not a perfect match. This can be seen in the graphical representation in figure 9. The items added also did fit where they were expected with other lower ability items. Infit and outfit statistics (Table 6) were analyzed to look for items that did not behave as expected. Values that were excluded after the CTT analysis, aside from 14, also showed higher than expected infit and outfit statistics, though none were unproductive for measurement. Item characteristic curves were also analyzed to check that data fit the Rasch model. Figure 10 shows a comparison between a well-performing item and one that was identified to be excluded. All items identified in classical test theory did not plot against the model item characteristic curve as well as expected. Table 7 compiles metrics from CTT and IRT.

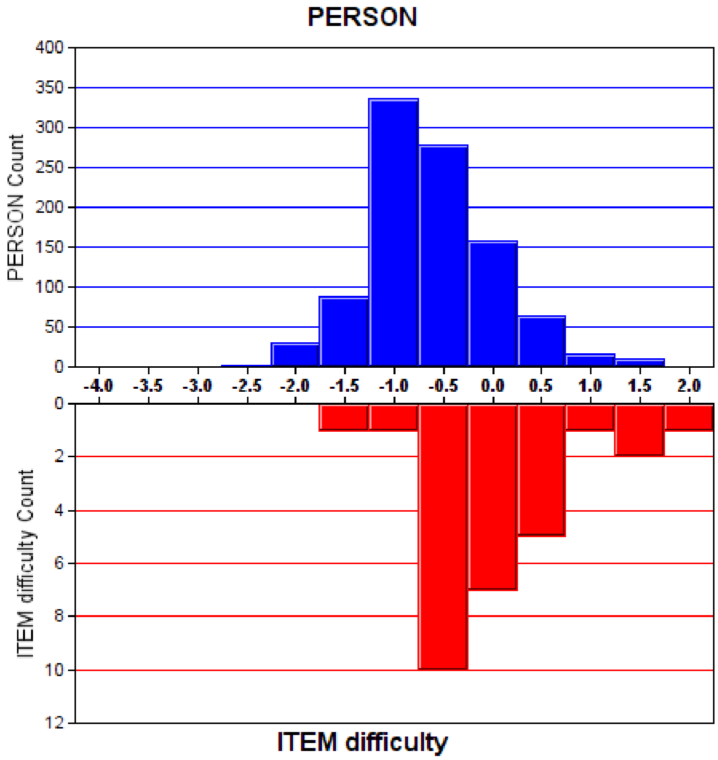


Figure 4.9 Graphical representation of a Wright Map showing relationships between participants' abilities and item difficulty after full implementation of the CI.

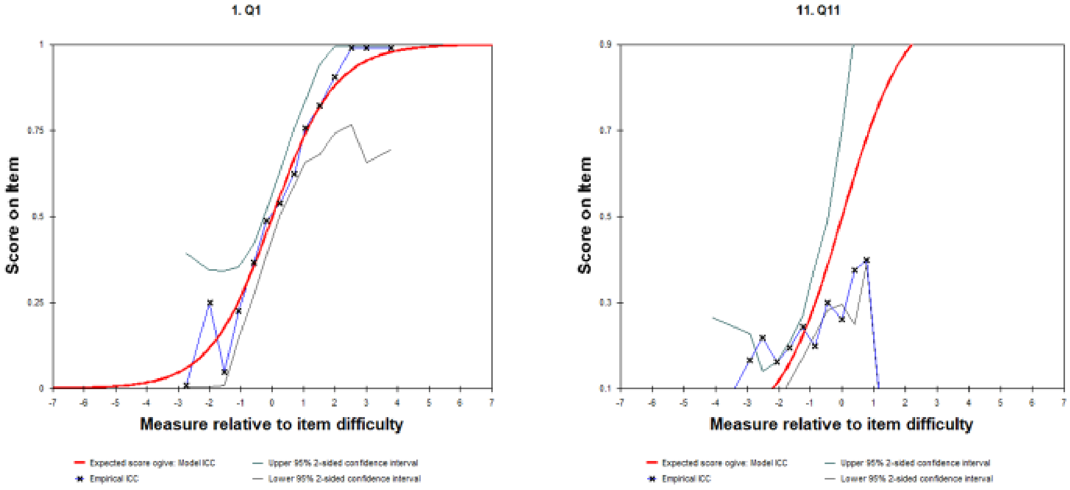


Figure 4.10 Empirical item characteristic curves plotted against the model characteristic curve of a well-performing item (1) and a poorly performing item (11).

Table 4.6 Difficulty, Infit, and outfit statistics from IRT. Highlighted items were items that were identified as problematic by CTT. Criteria for each metric can be found in table 5

Item	Difficulty	Infit Mean-Square	Infit Standardized Fit Statistic	Outfit Mean-Square	Outfit ZSTD Standardized Fit Statistic
1	-.74	.98	-1.04	.99	-.65
2	-.44	1.01	.77	1.02	.72
3	-.19	.94	-2.67	.94	-2.32
4	-.56	1.02	1.28	1.02	1.02
5	1.65	1.08	.96	1.29	2.65
6	-1.53	.93	-2.45	.88	-3.11
7	-1.09	.96	-1.95	.95	-1.86
8	-.05	1.01	.29	1.02	.72
9	-.46	1.03	1.84	1.04	1.61
10	1.86	1.04	.47	1.20	1.64
11	.73	1.11	2.48	1.22	3.59
12	-.56	.95	-2.91	.94	-2.82
13	-.08	1.01	.49	1.01	.23
14	.37	1.08	2.32	1.14	3.18
15	.51	.99	-.36	.99	-.27
16	.15	1.04	1.23	1.05	1.35
17	-.39	.96	-1.96	.96	-1.74
18	-.27	.98	-.97	.97	-1.03
19	.36	1.12	3.33	1.14	3.25
20	.22	.91	-2.97	.88	-3.19
21	-.33	.99	-.67	.98	-.76
22	1.27	.96	-.66	.99	-.12
23	.01	.89	-4.40	.87	-4.38
24	.31	1.04	1.11	1.10	2.28
25	.82	1.12	2.45	1.22	3.45
26	-.70	.87	-6.66	.85	-6.10
27	-.72	.97	-1.64	.96	-1.34
28	-.13	.97	-1.12	.97	-1.07

Principal components analysis of residuals was used to assess dimensionality in IRT with the full dataset. The first eigenvalue was 2.34, which is below the maximum of 3.0 expected for unidimensional scales (Linacre, 2019). Further exploration revealed no further evidence for multidimensionality. Largest standardized residual correlations were used to check for local independence, meaning that items were checked to make sure that no items were dependent on any other item. These correlations showed that most correlations were negative (which is expected for independence) and that items that did have positive correlations (Q14 and Q24, Q23, and Q26) were low. Differential item functioning was used to explore differences in performance between male and female participants to establish evidence of cultural validity. Traditionally, significant contrasts between absolute values between 0.43 and 0.64 indicate slight to moderate differential item functioning, while absolute values greater than .64 indicate moderate to large differential item functioning (Zwick, 1999). Only two items showed slight to moderate differential item functioning, questions 10 and 13, with the higher of the two (10), being later eliminated from the test. A plot of male and female performance plotted against expected performance from the model can be seen in figure 11.

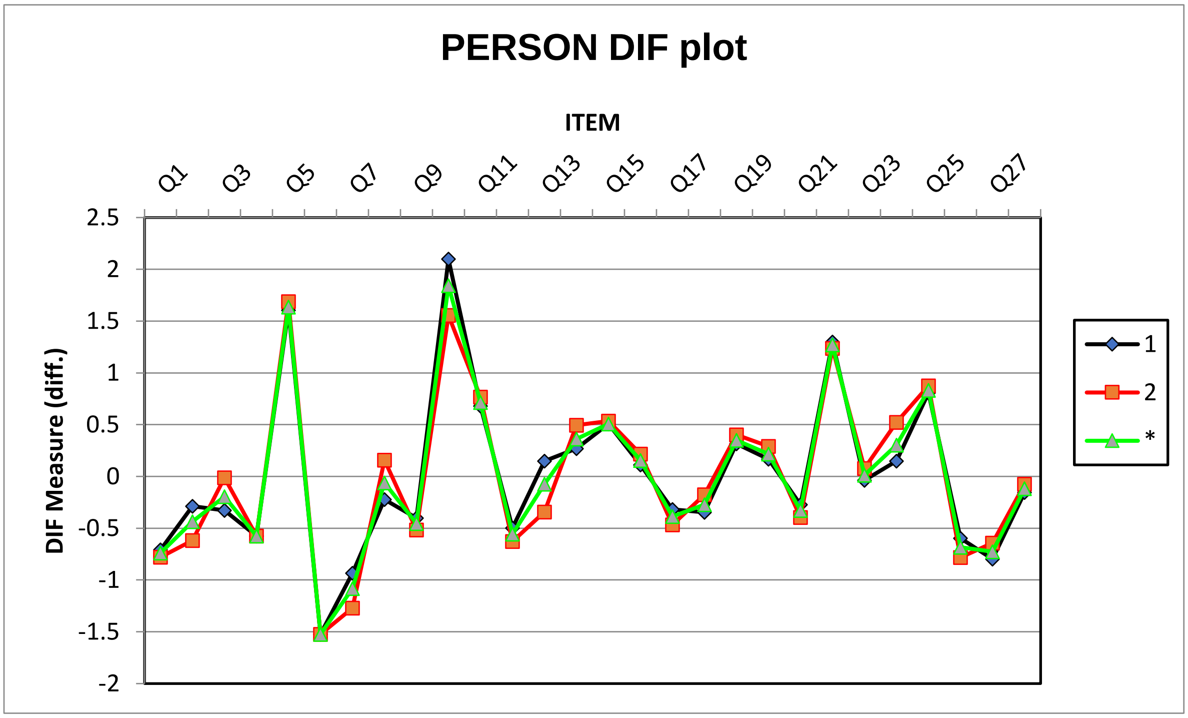


Figure 4.11 Differential Item Functioning Plot. 1 represents males, 2 represents females, and the green curve represents the performance predicted by the Rasch model.

Impact of Wording of Feedbacks

Each participant was given questions on feedback loops phrased in two ways. One way was using the terminology "positive" and "negative" feedback loops. The second was "reinforcing" and "balancing" feedback loops. Using items with the reinforcing balancing wording did improve the overall reliability of the CI, pushing the alpha value to .71, above the recommended value of .7. The rephrasing also improved the IRT Person separation reliability up to .66 (Table 7). The Wright Map produced can be seen graphically represented in figure 12, showed the mean of items was also closer to the mean of participant ability. As the data was not parametric for each item, a Wilcoxon signed ranks test was conducted to see if there was a significant difference between performance within individuals when answering the positive/negative versus the reinforcing/balancing wording. While performing better on the reinforcing/balancing wording on question 9, the difference was not significant (*p*=.082). Participants performed significantly better on 14, 22, and 25 when phrased with the reinforcing/ balancing wording (*p*<.001).

Table 4.7 Comparison of reliability statistics for positive/negative wording versus reinforcing/balancing wording. These values reflect question numbers 5,10, 11, 19, and 25 from the original instrument being removed.

	Positive/Negative	Reinforcing/Balancing
Cronbach's Alpha	.64	.71
Item Separation Reliability	.99	.99
Person Separation Reliability	.58	.66

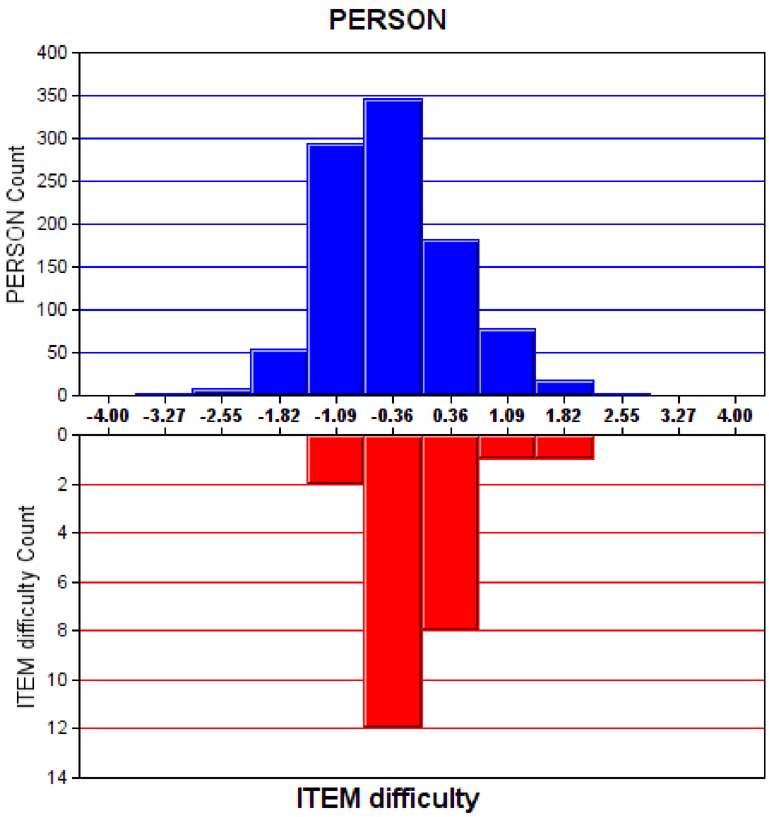


Figure 4.12 Graphical representation of a Wright Map showing the relationship between participants and items based off of ability and difficulty.

Discussion

Existing studies of systems thinking in the geosciences have relied on content-based tests or qualitative observations or assignments. Thus, a need exists for a validated instrument designed to measure systems thinking abilities in the context of the Earth sciences. This paper presents a semi-customizable instrument with evidence of validity and reliability. The instrument is semi-customizable as the individual administering can choose between wording used when discussing feedbacks. The steps taken during test construction and analysis points toward the overall validity and reliability of the developed test for evaluating basic systems thinking skills in the context of the Earth sciences.

This instrument addresses Earth systems thinking frameworks identified by Scherer and colleagues (2017) from the literature relating to Earth systems perspectives, Earth systems thinking, and complexity sciences. Though it is impossible to encompass all aspects of systems thinking in the context of the Earth system in one instrument, this instrument allows for an important step forward in collecting quantitative data on systems thinking in undergraduate students and the efficacy of systems thinking teaching practices. We had hypothesized that the item would be multidimensional, with questions from different frameworks being more likely to be related to each other. However, both CTT and IRT metrics showed the items to unidimensional. This is likely due to the fact that most participants were novices, likely with very little familiarity with systems thinking. Future work should examine the dimensionality of the instrument when taken by both experts and students with some training in systems thinking to see if the unidimensionality holds.

In this study, CTT and IRT were paired to offer a robust analysis and evidence of validity and reliability (table 8). Though many components of both CTT and IRT provide evidence for validity and reliability, we can use the metrics shown in the table to show and track the improvement of the instrument for multiple iterations. CTT allows for standard analysis of item difficulty and discrimination in easily accessible terms. It also allows for measures on how each item relates to overall performance on the instrument and for standard estimates of reliability. While IRT measures may not be as accessible as those from CTT, Rasch analysis, in particular, allows the ability to evaluate along a construct continuum while offering metrics that can be used to evaluate individual items and the overall test. It also offers metrics for measuring differential item functioning and mapping item difficulty against participant ability. Additionally, IRT estimates of item difficulty and internal consistency tend to be more stable across samples than CTT and have significantly fewer measurement errors (Magno, 2009). Though either framework for analysis can be used, using them together can help provide multiple lines of evidence for reliability and validity.

Table 4.8 Summary of CTT and IRT statistics for the pilot administration, combined data, and combined data with problematic items removed.

Metric	Pilot	Full Data	Full Data (Items removed)
Item Difficulty	.33	.36	.40
Item Discrimination	.27	.27	.33
Point Biserial Correlation	.096	.27	.31
Cronbach’s Alpha	.49	.58	.64
Item Separation Reliability	.96	.99	.99
Person Separation Reliability	.37	.52	.58

Through rigorous development and analysis, we were able to improve the quality of the CI as well as provide additional evidence of reliability and validity. As we analyzed pilot data using metrics from both CTT and IRT, we were able to identify a need for easier items to be added to the instrument. Upon doing this, we were able to improve all CTT metrics as well as reliability statistics from both CTT and IRT. It is important to note that the choice of language of feedbacks does impact the reliability of the instrument. In the case of this study, which targeted participants that were representative of the general public, the instrument is most reliable and exceeds the recommended alpha value of .7 (Nunally, 1978; Litwin, 1995) when the balancing/reinforcing language is used. That being said, it has been argued that because concept inventories tend to not be homogenous tests, tests of internal consistency can seriously underestimate reliability and lower values may be acceptable (Miller, 1995). This work on feedbacks suggests the importance of using the balancing/reinforcing language as it is more intuitive and more easily understood by the general public.

Of the items that were eliminated (Table 9), a combination of CTT and IRT metrics were used. Likely, these items showed poorer performance than others since they may have been too content-specific. For example, question five asks specifically about ocean sediments, which may have made the question more difficult than intended due to the fact that participants may not have been familiar with ocean sediments. Questions 10 also asked about the hydrosphere, and research on conceptions of the Earth system has demonstrated that people tend to be less familiar with the hydrosphere than other components of the Earth system (Soltis, McNeal, & Schnittka, in review). Questions 11, 14, and 25 all included compounds or terms that may have been unfamiliar to the general public. Question 25, in particular, was a diagram showing marine biogeochemical feedbacks that likely was too advanced for the intended audience. Though question 14 did not perform well based on classical test theory metrics, it was left in the final CI because it has an alternative version, where the feedback wording is changed to reinforcing/balancing, which performed better.

Table 4.9 Eliminated items from the concept inventory. These items were eliminated due to poor performance based on metrics from CTT and IRT

Question	Text
5	Which of the following contributions to ocean sediments indicates that the ocean is an open system?
10	The production of the Earth's first organic soils by forests resulted in a permanent change of the nutrient input to the Earth's oceans. What best describes this event in the context of the Earth system?
11	Eutrophication results from an increase of nutrients in a body of water. The end effect of severe eutrophication is oxygen depletion and the formation of a dead zone. Which of the following is most likely to lead to oxygen depletion?
14	A warming climate can lead to the melting of methane hydrates in the deep ocean. Being a gas, methane rises, entering the atmosphere where it traps heat. What best describes this process?
19	In which way would the hydrosphere most likely respond to a cooling climate?
25	The following diagram shows the complex feedbacks between organic carbon burial, nutrients, climate, atmospheric composition, and ocean circulation.

Limitations and Future Directions

Though workers on Amazon's Mechanical Turk has been shown to be reflective of the general population, it is not an exact match (Buhrmester, Talaifar, & Gosling, 2018). Due to the anonymous nature of data collection, we do not know much about our participants. As this is reflective of the general public, this instrument has yet to be validated on individuals with geoscience expertise. Important future work involves testing the instrument with individuals of various levels of Earth science training and systems thinking. As we know that systems thinking skills must be explicitly taught (Stillings, 2012), it is especially critical that the instrument be tested on individuals with training in systems thinking. Additionally, data on ethnicity from participants was not collected, so it is important to explore differential item functioning by ethnicity in the future to further establish cultural validity (Solano-Flores, 2001).

Future work also needs to focus on piloting the instrument in classroom settings as a pre-post measure in order to validate its use in accurately measuring systems thinking abilities.

Though to our knowledge no other quantitative instruments to measure Earth systems thinking skills exist, the instrument should also be tested in conjunction with other Earth science concept inventories and qualitative systems thinking tools to establish concurrent validity.

Conclusions

This CI addresses a gap in the literature on Earth systems thinking, as it provides a research-validated tool that is semi-customizable to measure certain dimensions of systems thinking ability. Though the entire construct of Earth systems thinking cannot be captured in one instrument, this instrument can serve as a basic quantitative measure for assessing both students' systems thinking abilities in the undergraduate classroom as well as the effectiveness of practices related to systems thinking ability. Thus, this instrument serves as an important step in moving research on systems thinking in the Earth sciences forward.

This work also presents a framework for using both CTT and IRT in the analysis of CI results. These practices in development and analysis can be of use to both practitioners and researchers looking to develop research-grade instruments. In doing this, practitioners and researchers can develop more ideal measures to conduct research and produce results that are more valid and reliable. Though the design and analysis processes are rigorous, it is manageable and aspects of CTT and IRT can be used together or individually to provide evidence of validity and reliability.

Acknowledgments: We would like to thank Chris Schnittka, Stephanie Shepherd, Joni Lakin, Anne Gold, Leilani Arthurs, Juliette Rooney-Varga, Lisa Gilbert, Hannah Scherer for their insights in the development and revision of this instrument.

References

American Educational Research Association, American Psychological Association, National Council on Measurement in Education, Joint Committee on Standards for Educational, & Psychological Testing (US). (1999). *Standards for educational and psychological testing*. Amer Educational Research Assn.

Arnold, R.D., & Wade, J.P. (2015). A definition of systems thinking: A systems approach. *Procedia Computer Science, 44,* 669-678.

Arthurs, L., Hsia, J. F., & Schweinle, W. (2015). The oceanography concept inventory: A semicustomizable assessment for measuring student understanding of oceanography. *Journal of geoscience education*, 63(4), 310-322.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science*, 13(2), 149-154.

Cappelleri, J. C., Lundy, J. J., & Hays, R. D. (2014). Overview of classical test theory and item response theory for the quantitative assessment of items in developing patient-reported outcomes measures. Clinical therapeutics, 36(5), 648-662.

Crocker, L., & Algina, J. (2008*). Introduction to Classical & Modern Test Theory*. 2008. Cengage Learning, Mason, Ohio.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *psychometrika*, *16*(3), 297-334.

Grohs, J. R., Kirk, G. R., Soledad, M. M., & Knight, D. B. (2018). Assessing systems thinking: A tool to measure complex reasoning through ill-structured problems. Thinking Skills and Creativity, 28, 110-130.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory* (Vol. 2). Sage.

Jarrett, L., Ferry, B., & Takacs, G. (2012). Development and validation of a concept inventory for introductory-level climate change science. *International journal of innovation in science and mathematics education*, 20(2), 25-41

Jimenez-Buedo, M., & Miller, L. M. (2010). Why a trade-off? The relationship between the external and internal validity of experiments. *Theoria. Revista de Teoría, Historia y Fundamentos de la Ciencia*, *25*(3), 301-321.

Jordan, R. C., Sorensen, A. E., & Hmelo-Silver, C. (2014). A conceptual representation to support ecological systems learning. Natural Sciences Education, 43(1), 141-146.

Lally, D., Forbes, C. McNeal, K.S., & Soltis, N. A. (in press). National Geoscience Faculty Survey 2016: Prevalence of systems thinking and scientific modeling learning opportunities. *Journal of Geoscience Education*

Libarkin, J. C., & Anderson, S. W. (2005). Assessment of learning in entry-level geoscience courses: Results from the Geoscience Concept Inventory. *Journal of Geoscience Education*, 53(4), 394-401.

Libarkin, J. (2008, October). Concept inventories in higher education science. In BOSE Conf.

Libarkin, J. C., Gold, A. U., Harris, S. E., McNeal, K. S., & Bowles, R. P. (2018). A new, valid measure of climate change understanding: Associations with risk perception. Climatic change, 150(3-4), 403-416.

Linacre, J. M. (1999). Understanding Rasch measurement: estimation methods for Rasch measures. *Journal of outcome measurement*, 3, 382-405.

Linacre, J. M. (2019) Facets computer program for many-facet Rasch measurement, version 3.83.0. Beaverton, Oregon: Winsteps.com

Keith, T. Z., & Reynolds, C. R. (1990). Measurement and design issues in child assessment research. *Handbook of psychological and educational assessment of children*, *1*, 29-61.

Kuder, G. F., & Richardson, M. W. (1937). The theory of the estimation of test reliability. *Psychometrika*, *2*(3), 151-160.

Magno, C. (2009). *Demonstrating the difference between classical test theory and item response theory using derived test data*. The International Journal of Educational and Psychological Assessment, 1(1), 1-11.

Orion, N. & Libarkin, J. (2014). Earth system science education. *Handbook of research on science education*, 2, p. 481-496.

Patton, M. Q. (2002). *Qualitative evaluation and research methods (3rd ed.).* Thousand Oaks, CA: Sage Publications, Inc.

Reynolds, C. R., Livingston, R. B., Willson, V. L., & Willson, V. (2010). *Measurement and assessment in education*. Upper Saddle River: Pearson Education International.

Richards, L., & Morse, J. M. (2012). *Readme first for a user's guide to qualitative methods*. Thousand Oaks, CA: Sage.

Scherer, H. H., Holder, L., & Herbert, B. (2017). Student learning of complex Earth Systems: Conceptual frameworks of Earth Systems and instructional design. *Journal of Geoscience Education*, 65(4), 473-489.

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107.

Solano-Flores, G., & Nelson-Barber, S. (2001). On the cultural validity of science assessments. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, *38*(5), 553-573.

Soltis, N.A., McNeal, K.S., Forbes, C., Lally, D. (2019). The relationship between active learning, course innovation and teaching Earth systems thinking: A structural equation modeling approach. *Geosphere*, 15(5), 1703-1721.

Soltis, N.A., McNeal, K.S., Schnittka, C.G., (In review). Undergraduate student conceptions of the Earth system and biogeochemical cycling. *Journal of Geoscience Education*.

Stillings, N. (2012). Complex systems in the geosciences and in geoscience learning. In Kastens, K., and Manduca C., eds. *Earth and mind II: A synthesis of research on thinking and learning in the geosciences*. Boulder, CO: Geological Society of America, p. 91-96.

Walker, S. L., & McNeal, K. S. (2013). Development and validation of an instrument for assessing climate change knowledge and perceptions: The climate stewardship survey (CSS). International Electronic Journal of Environmental Education, 3(1), 57-73.

Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. New Jersey: Lawrence Earlbaum Associates

Wright, B. D., & Masters, G. N. (1981). *Rating scale analysis*. Chicago: MESA Press

CHAPTER 5: CONCLUSION

Understanding Earth systems thinking in the context of the geosciences is a critical step in both improving undergraduate perceptions of Earth phenomena as well as in enhancing understanding of major issues of sustainability and global change. Environmental challenges such as climate change, ocean acidification, biodiversity loss, eutrophication, and pollution all involve a variety of components making up larger systems that operate on a variety of spatial and temporal scales. The studies that compose this dissertation look at EST in three different ways: what teaching practices related to systems thinking look like based off of large-scale survey data, how students conceptualize the Earth system through the lens of biogeochemical cycles, and the development of an instrument to assess EST skills. Together, these studies illuminate the current state of EST thinking, elucidate the challenges associated with conceptualizing complex systems, and provide a path forward through providing an assessment piece for studies the assess teaching and learning about complex Earth systems.

The first study utilized structural equation modeling to explore the relationship between latent variables related to active learning practices, course change, and EST-teaching practices based on 2016 responses from the Geoscience Faculty Survey. An exploratory factor analysis of items related to EST-teaching revealed three major teaching practices that correspond to frameworks to EST from the literature (Scherer et al., 2016), with practices related to systems modeling being more distinct than practices related to adding systems thinking elements or real-world elements into courses. An exploratory factor analysis of items related to course changes also found two major modalities of change, both relating to systems thinking. The full structural equation model that was developed found that active learning did not directly predict EST-teaching practices as hypothesized, and rather than the relationship between active learning and EST-teaching was mediated by course changes. This suggests that explicit professional development related to systems thinking is necessary, particularly as it relates to systems modeling, as active learning enough is not sufficient to enhance Earth systems thinking teaching practices. Individuals who engage in active learning teaching strategies are more likely to make course changes that may enhance systems thinking teaching opportunities.

The second study explored student conceptions of the Earth system through the lens of biogeochemical cycles. Biogeochemical cycles were chosen as a lens to the fact that they connect all spheres are the Earth system and follow complex pathways and feedbacks. Additionally, they transcend any one discipline and cannot be untangled from issues of global change. Both interviews and student drawings were used to assess participants mental models of the carbon, nitrogen, and phosphorous cycles. This work found that all three cycles were poorly conceived by students across disciplines, particularly the nitrogen and phosphorous cycles. Typically, students focused on short term cycling through the biosphere and atmosphere, with longer-term cycling through the geosphere and hydrosphere being largely ignored or poorly conceived. Thus, students often based their mental models on perceptions, those things that were directly observable on human time scales, rather than conceptions of phenomena that must be imagined or inferred. Student mental models also often showed only linear or cyclical pathways with little to no complexity. Students also struggled to connect biogeochemical cycles to environmental issues or matters of sustainability, and several persistent misconceptions such as carbon dioxide destroying ozone and leading to global warming did come up repeatedly. This work suggests that across disciplines more intentional work needs to be done in embedding biogeochemistry into instruction and in helping students develop more complex and accurate mental models of these cycles.

The final study focused on the development and testing of an instrument designed to assess systems thinking skills in the context of the Earth sciences. The study documents the design process but also details how both classical test theory and item response theory can be used together to establish evidence of reliability and validity in an iterative testing process. The instrument development process demonstrated that for the general public, the construct of Earth systems thinking is unidimensional. The design process also showed that using the terms 'balancing' and 'reinforcing' feedbacks versus 'negative' and 'positive' feedbacks results in a more reliable assessment. However, the instrument does allow users to choose the terminology that works best for the target of interest. The instrument developed in this study serves as an important step forward in working related to EST as it provides the first multiple-choice tool to assess EST in the context of geoscience courses. This allows for many new opportunities in studies that aim to assess EST-teaching and learning.

While enhancing our understanding of EST teaching and learning, these studies also offer insight into exciting future research opportunities. The development of the Earth Systems Thinking Concept Inventory provides the tool needed to assess the efficacy of systems thinking teaching practices as identified by the first study. It also provides a way to assess ways in which teaching about biogeochemical cycles may relate to the development of systems thinking skills. In the broader sense, the development of the CI allows other researchers to have another tool to quantitatively assess systems thinking abilities both in how they may naturally develop in courses and as the result of interventions targeted to assess systems thinking abilities. With the metrics offered by the CI, it is now possible to explore the relationship between EST abilities, spatial thinking abilities, temporal thinking abilities, as well as to content knowledge in fields such as climate change science, geology, and oceanography.

Regardless of discipline, undergraduate students must grapple with a variety of issues relating to global change and sustainability. Thus, being able to conceptualize the Earth as a complex system and to understand the nature of systems is an essential skill for all students. The studies presented in this dissertation present three unique lenses on EST teaching and learning. It also offers new possibilities for assessing systems thinking skills in the context of the Earth sciences. This work is critical, as through developing systems thinking abilities through undergraduate courses, we as educators take part in meaningful work in equipping students to better understand the complexity of the planet they live on and the interactions between all of its components. Beyond that, this work offers a path forward and new research tools to deepen our understanding of what Earth systems thinking is and how to most effectively teach it and help students make sense of complex Earth systems.

APPENDICES

Appendix A

Semi-Structured Interview Script

What comes to mind when you hear ‘Earth System?’
What parts make up the Earth System?
Can you tell me about those parts?
What is a system?
Where can I find carbon?
Why is carbon important?
What are some ways in which carbon moves?
Where can I find nitrogen?
Why is nitrogen important?
What are some ways in which nitrogen moves?
Where can I find phosphorus?
Why is phosphorus important?
What are some ways in which phosphorus moves?
Of these elements, are you aware of any environmental issues associated with any of them?
Can you think of any other elements that cycle?

Questionnaire

ID Number (from investigator):
Class Rank:
Major or proposed major:
Minor:
From each subgroup, select any courses you have taken or are currently taking

Geology

Dynamic Earth Earth and Life through Time Environmental Geology
Geomorphology Paleobiology Geochemistry Hydrology

Chemistry

Survey of Chemistry Fundamentals of Chemistry General Chemistry I General Chemistry II
Organic Chemistry Biochemistry Environmental Chemistry

Biology

Intro to Biology Principles of Biology Organismal Biology Ecology
Biology of Marine Systems General Microbiology

General Science

Concepts of Science Methods of teaching Science I Methods of Teaching Science II

When do you recall learning about the movement of carbon? Please include any college courses or classes taken in high school or middle school. If you don’t recall learning about this, please indicate so.

When do you recall learning about the movement of nitrogen? Please include any college courses or classes taken in high school or middle school. If you don’t recall learning about this, please indicate so.

When do you recall learning about the movement of phosphorus? Please include any college courses or classes taken in high school or middle school. If you don’t recall learning about this, please indicate so.

In what classes have you heard about or learned about the Earth System? Please include any college courses or classes taken in high school or middle school. If you don’t recall learning about this, please indicate so.

1.) Increased temperature can cause what change in the carbon cycle?
  A.) Increased weathering drawing down carbon
  B.) Increased volcanic activity increasing carbon dioxide levels
  C.) Increased ice melting moving carbon to the deep ocean
  D.) Increased photosynthesis results in an increase in carbon dioxide
2.) Phosphorus differs from Nitrogen in…
  A.) Nitrogen is a critical nutrient to plant life and phosphorus is toxic
  B.) Nitrogen in an important component of the atmosphere and phosphorus is not
  C.) Microbial activity is essential for moving Phosphorus, but not nitrogen
  D.) Nitrogen is typically locked in rocks and minerals, whereas phosphorus can be found in a variety of settings

3.) Excess nitrogen introduced to a body of water is likely to…

- A.) Poison living things
- B.) Change the temperature of the body of water
- C.) Initially increase primary productivity
- D.) Significantly modify the pH of the water

4.) Photosynthesis moves carbon from the ____________ to the ____________.

- A.) Biosphere to Atmosphere
- B.) Atmosphere to Biosphere
- C.) Hydrosphere to Biosphere
- D.) Geosphere to Atmosphere

5.) Atmospheric Nitrogen can be taken up and used by living organisms through…

- A.) Respiration
- B.) Photosynthesis
- C.) Microbial Activity
- D.) Diffusion

Appendix B

Earth System Thinking Concept Inventory

Start of Block: Default Question Block

Q30 Thank you for participating in our study! Please read the following document to learn about procedures and terms of consent before participating. After you download and read the document please indicate if you consent to participate. You may print the consent form for your records.
Consent Document

This survey will consist of 28 multiple choice questions, followed by four follow up questions and several demographic questions. Please take your time and try your best.

○ I consent

○ I do not consent

X→

Q1 Photosynthesis moves carbon between which components of the Earth system?

○ Atmosphere to Biosphere

○ Biosphere to Geosphere

○ Biosphere to Atmosphere

○ Geosphere to Biosphere

Q2 Which of the following is an example of an important flux in the rock cycle?

○ Erosion

○ Sediment

○ Sublimation

○ Magma

Q3 If the ocean absorbs more carbon than it releases which term best describes the ocean's role?

○ A flux

○ A sink

○ A reservoir

○ A source

Q4 A volcanic eruption moves materials between which different parts of the Earth system?

○ Hydrosphere to Atmosphere

○ Atmosphere to Geosphere

○ Geosphere to Atmosphere

○ Geosphere to Biosphere

Q5 Which of the following contributions to ocean sediments indicates that the ocean is an open system?

○ Cosmogenic dust

○ The remains of organisms

○ Products of terrestrial weathering and erosion

○ Chemical precipiation

Q6 Which is the flux moving water molecules from the open ocean to the atmosphere?

○ Clouds and fog

○ Evaporation

○ Water vapor

○ Precipitation

Q7 What is responsible for the Earth’s modern oxygen content?

○ The evolution of photosynthesis

○ Outgassing of the early Earth

○ The Earth's ocean

○ The sun and its distance from the Earth

Q8 Which is a way the biosphere may impact the geosphere?

○ Changing temperatures of magma

○ Formation of soils

○ Radioactive decay of heavy elements

○ Freeze thaw cycles

Q9 When atmospheric carbon dioxide concentrations are high, more carbon dioxide is absorbed by the Earth's ocean, lowering atmospheric carbon dioxide. What best describes this event in the context of the Earth system?

○ A positive feedback loop

○ A negative feedback loop

○ A perturbation

○ A forcing

Q10 The production of the Earth's first organic soils by forests resulted in a permanent change of the nutrient input to the Earth's oceans. What best describes this event in the context of the Earth system?

○ A positive feedback loop

○ A negative feedback loop

○ A perturbation

○ A forcing

Q31 Please select Choice 2

○ Choice 1

○ Choice 2

○ Choice 3

○ Choice 4

Q11 Eutrophication results from an increase of nutrients in a body of water. The end effect of severe eutrophication is oxygen depletion and the formation of a dead zone. Which of the following is most likely to lead to oxygen depletion?

○ The growth of algae blocking sunlight

○ The poisoning of living things by chemicals like mercury

○ The decay of dead organisms

○ The increased sediment making the water cloudy

Q12 Based on the diagram, which of the following would enhance the rate of soil erosion?

○ Planting more grass

○ Increasing the grazer population

○ Introducing a fatal cow disease

○ Adding fertilizer to enhance grass growth

Q13 Which of the following would be most likely to be a perturbation on river flow?

○ A seasonal drought

○ Tectonic uplift

○ Buildup of glaciers

○ Long term erosion of a mountain range

Q14 A warming climate can lead to the melting of methane hydrates in the deep ocean. Being a gas, methane rises, entering the atmosphere where it traps heat. What best describes this process?

○ Negative feedback loop

○ Positive feedback loop

○ Perturbation

○ Forcing

Q15 Monsoonal circulation occurs due to major atmospheric pressure differentials created by large landmasses and vast expanses of ocean. Which part of the Earth system contributes least to monsoon formation?

○ Biosphere

○ Geosphere

○ Atmosphere

○ Hydrosphere

Q16 Which role does photosynthesis play in the carbon cycle?

○ A reservoir

○ A flux

○ A source

○ A forcing

Q17 When considering the carbon cycle, which process is part of a long-term cycle?

○ Formation of fossil fuels

○ Photosynthesis

○ Decay

○ Ocean-air diffusion

Q18 The construction of a large dam would have the least impact on which component of the Earth system?

○ Hydrosphere

○ Geosphere

○ Atmosphere

○ Biosphere

Q19 In which way would the hydrosphere most likely respond to a cooling climate?

○ Decrease in dissolved oxygen

○ Decrease in alkalinity

○ Increase in rates of ocean circulation

○ Increase in sea level

Q20 Which geosphere process would result in a long-term decrease in biosphere diversity?

○ Formation of a supercontinent

○ Opening of an ocean

○ Separation of landmasses

○ Mountain building activity

Q32 5-3=?

○ 1

○ 2

○ 3

○ 4

Q21 On Earth, which best describes the state of most surface environments?

○ Open system

○ Closed system

○ Isolated system

○ Equilibrium system

Q22 Based on the following diagram, which of the following is being depicted?

○ Negative feedback loop

○ Positive feedback loop

○ A forcing

○ A perturbation

Q23 Considering the diagram in the question above, which of the following could increase atmospheric carbon dioxide concentration?

- ○ Reduction in respiration
- ○ Enhanced weathering of silicate minerals
- ○ Increased absorption of carbon dioxide by oceans
- ○ Widespread forest fires

Q24 As climate warms, more ice melts. This results in less solar radiation being reflected by the ice, further warming the climate. What best describes this relationship?

- ○ Negative feedback loop
- ○ Positive feedback loop
- ○ Stasis
- ○ Equilibrium

Q25 The following diagram shows the complex feedbacks between organic carbon burial, nutrients, climate, atmospheric composition, and ocean circulation.
Based on this diagram, which would cause an increase in marine nitrogen fixation?

- ○ A decrease in ocean circulation
- ○ An increase in oxygen
- ○ A decrease in nutrient aqueous phosphrous
- ○ An increase in temperature

Q40 Evaporation moves water between what components of the Earth system?

- ○ Hydrosphere to Biosphere
- ○ Atmosphere to Hydrosphere
- ○ Atmosphere to Biosphere
- ○ Hydrosphere to Atmosphere

Q41 If a volcano releases more sulfur to the atmosphere during an eruption than it takes it, a volcano is best thought of as what?

- ○ A sink for sulfur
- ○ A source of sulfur
- ○ A forcing on the sulfur cycle
- ○ A feedback on sulfur cycling

Q42 Which component of the water cycle takes place on the longest timescale?

- ○ Evaporation of surface waters
- ○ Condensation of water to form clouds
- ○ Movement of groundwater through an aquifer
- ○ Melting of seasonal snow

End of Block: Default Question Block

Start of Block: Block 1

Q26 When atmospheric carbon dioxide concentrations are high, more carbon dioxide is absorbed by the Earth’s ocean, lowering atmospheric carbon dioxide. What best describes this event in the context of the Earth system?

○ Reinforcing feedback loop

○ Balancing feedback loop

○ A perturbation

○ A forcing

Q27 A warming climate can lead to the melting of methane hydrates in the deep ocean. Being a gas, methane rises, entering the atmosphere where it traps heat. What best describes this process?

○ Balancing feedback loop

○ Reinforcing feedback loop

○ Perturbation

○ Forcing

Q28
 Based on the following diagram, which of the following is being depicted?

○ Balancing feedback loop

○ Reinforcing feedback loop

○ Forcing

○ Perturbation

Q29 As climate warms, more ice melts. This results in less solar radiation being reflected by the ice, further warming the climate. What best describes this relationship?

○ Balancing feedback loop

○ Reinforcing feedback loop

○ Stasis

○ Equilibrium

End of Block: Block 1

Start of Block: Demographics

Q34 What is your gender?

○ Male

○ Female

○ Choose not to identify

Q35 What best describes your age range?

- ○ 18-25
- ○ 26-35
- ○ 36-45
- ○ 46-55
- ○ 55-65
- ○ Over 65
- ○ Choose not to identify

Q36 What best describes your level of education?

- ○ Some high school
- ○ High school diploma
- ○ Some college
- ○ Undergraduate degree
- ○ Advanced or Professional degree
- ○ Choose not to respond