ABSTRACT

PELCH, MICHAEL ANTHONY. Exploring how New Teaching Materials Influence the Beliefs and Practices of Instructors and Students' Attitudes about Geoscience. (Under the direction of Dr. David McConnell).

STEM educational reform encourages a transition from instructor-centered passive learning classrooms to student-centered, active learning environments. Instructors adopting these changes incorporate research-validated teaching practices that improve student learning. Professional development that trains faculty to implement instructional reforms plays a key role in supporting this transition. Effective professional development features authentic, rigorous experiences of sufficient duration. We investigated changes in the teaching beliefs of college faculty resulting from their participation in InTeGrate project that guided them in the development of reformed instructional materials for introductory college science courses. A convergent parallel mixed methods design was employed using the Teacher Belief Interview, the Beliefs About Reformed Science Teaching and Learning survey and participants' reflections on their experience to characterize pedagogical beliefs at different stages of their professional development. Qualitative and quantitative data show a congruent change toward reformed pedagogical beliefs for the majority of participants. The majority of participants' TBI scores improved toward more student-centered pedagogical beliefs. Instructors who began with the most traditional pedagogical beliefs showed the greatest gains. Interview data and participants' reflections aligned with the characteristics of effective professional development. Merged results suggest that the most significant changes occurred in areas strongly influenced by situational classroom factors.

Introductory geoscience courses play a crucial role in recruiting new geoscience majors but we know relatively little about how students' attitudes and motivations are

impacted by their experiences in geoscience classes. Students' attitudes toward science and its relevance are complex and are dependent upon the context in which they encounter science. Recent investigations into the attitudes of geoscience students have provided evidence to support this observation. We sought to expand this data set to provide a broader characterization of students' attitudes. We examined students' attitudes about the nature of science and its relevance before and after taking an introductory geology course. To characterize students' attitudes, we employed two quantitative instruments: the revised Scientific Attitude Inventory and the Changes in Attitudes about the Relevance of Science survey. Results show a negative trend in students' attitudes about the nature of science while their attitudes about the relevance of science were more variable. Our data support the findings of previous studies showing only minimal change in students' attitudes about the nature of science and its relevance after taking an introductory science course. The data also highlighted several misconceptions about the nature of science that could have implications toward future investigations of how geoscience courses impact student attitudes about science.

There is consensus among industrialized nations that it is important for its citizens and leaders to be scientifically literate. Therefore, it is important for the educational system to provide students with pertinent scientific knowledge, an understanding of the scientific processes, and the ability to evaluate scientific claims. Students' attitudes toward science and its relevance are important aspects of science literacy. We sought to determine if the repeated and explicit exposure to socioscientific issues through the use of InTeGrate course materials would result in positive changes to students' attitudes about the nature and relevance of science. We collected data on student attitudes using the revised Scientific Attitude Inventory

and the Changes in Attitude about the Relevance of Science survey in a quasi-experimental design over four semesters of an introductory physical geology course. Results show that the emphasis of socioscientific issues can influence both students' attitudes about the nature of science and their perceptions on the relevance of science. Changes were observed in data from both STEM and non-STEM majors. These findings have implications about how we select content for introductory science courses in general, and proves the utility of designing geoscience lessons based around socioscientific issues.

Exploring how New Teaching Materials Influence the Beliefs and Practices of Instructors and Students' Attitudes about Geoscience.

by
Michael Anthony Pelch

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

Marine, Earth, and Atmospheric Science

Raleigh, North Carolina
2016

APPROVED BY:

Dr. David McConnell Committee Chair	Dr. Karen McNeal
Dr. John Nietfeld	Dr. Karl Wegmann

DEDICATION

I would like to dedicate this body of research to my wife who inspired me to pursue a doctoral degree.

BIOGRAPHY

Michael Pelch spent his younger years in several places he would consider home. Formally, he would refer to St. Louis, MO as his hometown. He graduated from Francis Howell High School in 2000 and then went on to attend the University of Missouri and participate in the NRTOC program as a midshipman. In 2004 Michael graduated from the University of Missouri with Bachelor of Science degree in Geology and was also commissioned as a Second Lieutenant in the United States Marine Corps. After college he went on to serve as an infantry officer in the United States Marine Corps and participated in two overseas combat deployments during Operation Iraqi Freedom. Michael left active service as a Captain with and honorable discharge in 2009. Upon leaving active service, he returned to the University of Missouri to begin his graduate career in Geology studying fluid inclusions in Mississippi valley-type ore deposits. After two years of graduate work he completed his Masters in Science in Geology in 2011. It was around this same time that Michael's talented wife Katie finished her doctorate in Biology and they moved to Raleigh, NC so Katie could start her post-doctoral fellowship with the National Institute of Health. After their move to Raleigh in 2012, Michael met David McConnell and he decided to pursue a doctorate in the strange and unknown field of geoscience education at North Carolina State University. The rest you can say is history.

ACKNOWLEDGMENTS

The first people that I think deserve a prominent acknowledgement are all the present and past members of our research group: LeeAnna Chapman, Doug Czajka, Jason Jones, Katherine Ryker, Laura Lukes, April Grissom and Hayley Smith. Without these immensely dedicated, professional and intelligent individuals the last four years would not have made such a positive and important impact my life. A special thanks to LeeAnna Chapman and Doug Czajka who helped co-code interviews and grade hundreds of sustainability question responses. No mention of the Geoscience Learning Process Research Group can go without acknowledging my SOLID advisor, David McConnell. His mentorship and friendship over the past four years are without a doubt the most positive learning experiences I have ever had. I would also like to thank the Science Education Research Center (SERC) for providing the financial and professional support for my dissertation and for allowing me the valuable opportunity of working on the InTeGrate.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: Challenging instructors to change: A mixed methods investigation on the effects of materials development on the pedagogical beliefs of geoscience instructors.

Abstract

STEM educational reform encourages a transition from instructor-centered passive learning classrooms to student-centered, active learning environments. Instructors adopting these changes incorporate research-validated teaching practices that improve student learning. Professional development that trains faculty to implement instructional reforms plays a key role in supporting this transition. The most effective professional development experiences are those that not only help an instructor redesign a course, but that also result in a permanent realignment of the teaching beliefs of participating instructors. Effective professional development features authentic, rigorous experiences of sufficient duration. We investigated changes in the teaching beliefs of college faculty resulting from their participation in InTeGrate project that guided them in the development of reformed instructional materials for introductory college science courses. A convergent parallel mixed methods design was employed using the Teacher Belief Interview, the Beliefs About Reformed Science Teaching and Learning survey and participants' reflections on their experience to characterize pedagogical beliefs at different stages of their professional development.

Qualitative and quantitative data show a congruent change toward reformed pedagogical beliefs for the majority of participants. The majority of participants' TBI scores

improved toward more student-centered pedagogical beliefs. Instructors who began with the most traditional pedagogical beliefs showed the greatest gains. Interview data and participants' reflections aligned with the characteristics of effective professional development. Merged results suggest that the most significant changes occurred in areas strongly influenced by situational classroom factors.

The process of materials development employed in the InTeGrate project is comprised of rigorous, authentic, and collaborative experiences continued over a sufficient duration. Incorporating these characteristics in to a professional development program on a variety of scales can help promote the long-term adoption of reformed teaching strategies. Collaboration among geoscience professionals was one of the predominant drivers for change. Consequently, this research provides insight for the development of future professional development opportunities seeking to better prepare instructors to implement reformed instructional strategies in their classrooms.

Introduction

A principal goal of STEM educational reform is to encourage a shift from instructor-centered classrooms where students are largely passive to student-centered environments where learning is an active process (Singer, Nielsen, & Schweingruber, 2012; Singer & Smith, 2013). This transition has been supported by the development of research-validated teaching strategies that have been shown to improve student learning (Freeman et al., 2014; Handelsman et al., 2004; Kober, 2015; Singer & Smith, 2013) and reduce the achievement gap among student populations (Eddy & Hogan, 2014; Haak, HilleRisLambers, Pitre, &

Freeman, 2011). The effective implementation of research-validated instructional strategies will be referred to as reformed instruction (see MacIsaac & Falconer, 2002). Despite the evidence in favor of the adoption of reformed teaching methods, many instructors in K-12 and higher education institutions have been reluctant to adopt these new strategies (Barak & Shakhman, 2008; Henderson & Dancy, 2007) and participation in professional development may be insufficient to promote the transition of these alternative approaches to instruction (e.g. Ebert-May et al., 2011). University faculty members report several barriers to instructional reform including limited training, insufficient time, and lack of instructional and peer support (Dancy & Henderson, 2010; Fairweather, 2010; Henderson & Dancy, 2007; Sunal, Hodges, Sunal, & Whitaker, 2001; Wieman, Perkins, & Gilbert, 2010).

Classroom change requires that teachers reconsider how they conceptualize the learning environment (Luft & Roehrig, 2007). Instructor professional development has become the major focus of many systematic reform initiatives (Corcoran, 1995; Corcoran, Shields, Zucker, 1998; Garet et al., 2001; Guskey, 2002). Much of the research on this subject has occurred in K-12 or pre-service settings and examined instructors beliefs regarding their roles in the classroom, how students learn best, and the most effective types of student-instructor interactions (Barak & Shakhman, 2008; Fang, 1996; Fishbein & Ajzen, 1975; Hake, 1998; Jones & Carter, 2007; Kagan, 1992; Luft & Roehrig, 2007; Richardson, 1996). Beliefs affect action (Guskey, 1986; Hashweh, 1996; Kang & Wallace, 2005) and any steps seeking to promote lasting classroom change must consider a teacher's pedagogical beliefs (Barak & Shakhman, 2008; Guskey, 1986; Luft & Roehrig, 2007). Successfully shifting the pedagogical beliefs of instructors toward a reformed mindset is essential if professional development is to positively influence teaching in STEM classrooms (Keys & Bryan, 2001).

Professional development is intended to be a primary driver for the adoption of research-validated teaching strategies (Henderson, Beach, & Finkelstein, 2011; Wieman et al., 2010). Instructor professional development is a complex process which requires cognitive and emotional involvement of teachers both individually and collectively (Avalos, 2011). Professional development has been defined as the formal or informal learning opportunities designed to enhance teachers' professional competence, including knowledge, beliefs, motivations, and self-regulatory skills (Richter, Kunter, Klusmann, Lüdtke, & Baumert, 2011; Veenman, 1984). An instructor's preference for either formal or informal professional development is not static (Richter et al., 2011) and may vary throughout their careers (Dall'Alba & Sandberg, 2006; Gregorc, 1973; Henderson, Dancy, & Niewiadomska-Bugaj, 2012; Huberman, 1989; Richter et al., 2011; Sikes, Measor, & Woods, 1985; Unruh & Turner, 1970). The most common types of formal professional development opportunities often feature curricula designed to be large-scale national, state-wide, district-wide or intra-institutional programs (e.g. Choy, Chen, Bugarin, National Center for Educational Statistics., & Institute of Education Sciences (U.S.), 2006; Feiman-Nemser, 2001). Examples of formal professional development are workshops, retreats, and courses where experts disseminate information (Feiman-Nemser, 2001). Informal opportunities do not follow a specific curricula and are often smaller-scale opportunities that happen within a teachers own school setting (Desimone, 2009). These opportunities tend to be less common than formal opportunities (Choy et al., 2006) and often consist of individual learning, collaborative and mentoring activities, and teacher networks (Desimone, 2009; Parise & Spillane, 2010). Informal professional development is often embedded in the classroom or school context allowing instructors to reflect on their experiences and share ideas among colleagues (Putnam & Borko, 2000) and because of this is often a more authentic experience than formal professional development opportunities.

Professional development opportunities have had an inconsistent impact on classroom practice (Ebert-May et al., 2011; Feiman-Nemser, 2001; Garet et al., 2001; Henderson & Dancy, 2007). Professional development has more effectively served as a medium for disseminating information on reformed teaching strategies (Henderson et al., 2012). Instructors may be more likely to use reformed teaching strategies if they have attended a professional development opportunity (Henderson et al., 2012) but the degree of change towards alternative instructional strategies is relatively modest. Ebert-May et al. (2011) studied the effectiveness of two national professional development programs in biology. They found that 89% of participants reported implementing reformed, active learning strategies in their classrooms. However, videotaped observations of the participants revealed that 75% of them still relied on traditional or instructor-centered methods (Ebert-May et al., 2011). Professional development programs may not provide instructors with the tools necessary to overcome situational barriers to adoption that are often unique to an instructor's teaching environment (Ebert-May et al., 2011; Garet et al., 2001; Henderson et al., 2012; Henderson & Dancy, 2007). Situational factors can vary but commonly include class size and room layout, perceived student resistance to change, expectations of content coverage, and preparation time (Henderson & Dancy, 2007). Implementing institutional changes that promote reformed instruction is culturally and logistically difficult due to the “norms” of practice and organizational structures at most universities (Hora, 2012). Such reform is further complicated in STEM fields since different disciplines have their own set of standards and unique curricular needs (Hora, Oleson, & Ferrare, 2013; Singer et al., 2012).

Effective professional development is characterized by some combination of the following factors: 1) It occurs on an extended timescale beyond a single workshop; 2) It provides guidance and feedback on the design of reformed lessons; 3) It places emphasis on collaboration; and, 4) It is comprised of challenging and authentic experiences (Hawley & Valli, 1999; Loucks-Horsley, Hewson, & Stiles, 1998; Loucks-Horsley & Matsumoto, 1999; Garet et al., 2001; Henderson & Dancy, 2007; Penuel, Fishman, Yamaguchi, & Gallagher, 2007). Changing beliefs in adulthood is rare and belief systems are often developed or realigned over prolonged time (Nespor, 1987). Consequently, successful professional development needs to be of a sufficient duration, allowing for multiple cycles of presentation to provide participants with several opportunities to assimilate new knowledge and practices (Brown, 2004), and to engage multiple cycles of feedback and reflection (Penuel et al., 2007). An in-depth analysis of an instructor’s own teaching strategies may provide the challenge to pre-existing teaching beliefs (Gess-Newsome, Southerland, Johnston, & Woodbury, 2003) that can facilitate a “change in conversation” or gestalt shift that is critical to modifying those beliefs (Pajares, 1992). Collaboration, especially among participants in the same field, has a positive impact on the effectiveness of a professional development program (D’Avanzo, 2013; Garet et al., 2001; Penuel et al., 2007). Collaboration during professional development is most successful when the participants, rather than outside experts, drive and determine their interaction (Hargreaves & Dawe, 1990). For example, one-on-one collaboration between pre-service and in-service teachers has been shown to be an effective model of professional development by actively engaging teachers in their own research projects (Burbank & Kauchak, 2003). Some forms of professional development may be too passive to challenge participants adequately to rigorously explore and reflect on new concepts (Little, 1993; Penuel et al., 2007). Active participation during professional development is also important for the opportunity to be effective (D'Avanzo, 2013; Loucks-Horsley & Matsumoto, 1999; Penuel et al., 2007; Putnam & Borko, 2000). For example, participants should be involved in the examination of effective teaching strategies and should be encouraged to investigate why and how a particular strategy could be effective in their classes. Finally, professional development that models the authentic application of teaching strategies to lessons that are relevant to participants' classes is more likely to support instructional change (Burbank & Kauchak, 2003; D'Avanzo, 2013; Smylie, 1995).

We sought to examine changes in pedagogical beliefs of instructors participating in a national curriculum development program that featured many of the elements of effective professional development discussed above. This study focuses on the degree to which participation in creating resources for the InTeGrate project (http://serc.carleton.edu/integrate/) contributed to changes in the way instructors thought about and described their teaching. The InTeGrate (Interdisciplinary Teaching about the Earth for a Sustainable Future) project is a multi-disciplinary effort to promote geoscience literacy in the context of societal issues, and to increase the numbers of geoscience-related undergraduate majors. While the overall InTeGrate project is multi-faceted, this study focuses on college faculty involved in the development of a series of introductory geoscience modules. Each module was designed around a topic by teams of three geoscience faculty from a variety of institution types. The curriculum development aspect of the InTeGrate project were designed around the theory of participatory design where participants are a central component of the development process (Mao, Vredenburg, Smith, & Carey, 2005; Schuler & Namioka, 1993). While certain aspects of the InTeGrate project mirror formal professional development (e.g., workshops for participants), other aspects more closely resemble those of informal professional development. For example, participants are required to pilot developed materials in their own classrooms. Additionally, participants spent a significant amount of time discussing pedagogy and instructional strategies with peers within their development teams. InTeGrate materials development lasted approximately four semesters (sufficient duration) which included the design of original materials, pilot testing of those materials in classes and subsequent modifications based on participant's reflections, and feedback from other team members and trained observers. Authors worked as a development (collaborative) team but also interacted with a larger project team that included material reviewers, classroom observers, assessment experts, and leadership and evaluation personnel. Activities designed by the teams were rooted in their own experiences and applicable to students in the courses they taught (authentic). Finally, module authors were provided multiple opportunities (e.g., workshops, webinars, technical reviews) to formally and informally place their work in context of discipline-based education research (Singer et al., 2012) and reflect on their experiences under the guidance of an extensive multi-part materials development rubric and expert reviews (rigorous).

Theoretical Framework

Instructors are attracted to professional development because they feel that it will expand their knowledge and skills (Guskey, 2002). However, research on the long-term impacts of professional development suggest that this model is often not sufficient to promote long-term changes in the classroom (e.g., Ebert-May et al., 2011). In a modification to the traditional model of teacher change, Guskey (2002) proposed that instructors would modify their existing beliefs and attitudes if they experienced benefits from a new teaching strategy, such as increased student engagement or improvements in learning outcomes. Participants in the InTeGrate project had the opportunity to assess the effectiveness of new instructional strategies as they pilot the materials they developed in their classes. This practical experience could help secure the evolution of the instructors' pedagogical beliefs if it resulted in improvements in student engagement and learning.

Guskey's linear model for instructional change was expanded by Clarke and Hollingsworth (2002) who proposed a critical distinction between change sequences and growth networks. Clark and Hollingsworth noted that change in one domain can lead to change in another domain (change sequence), such as a teacher learning a new strategy at a workshop (external domain) and then experimenting with it in their classroom (domain of practice). However, they suggested that the concept of growth networks that involve numerous connections of enactment and reflection between multiple domains (Figure 1.1) would be a more effective model for representing permanent changes in instructional practice and teaching beliefs. The goal of professional development programs may be to change teacher practice, but without fostering continued enactment and reflection across multiple domains, it may limit professional growth and may not result in the long-term adoption of effective change. The InTeGrate project sought to influence multiple domains within an interconnected model of professional growth.

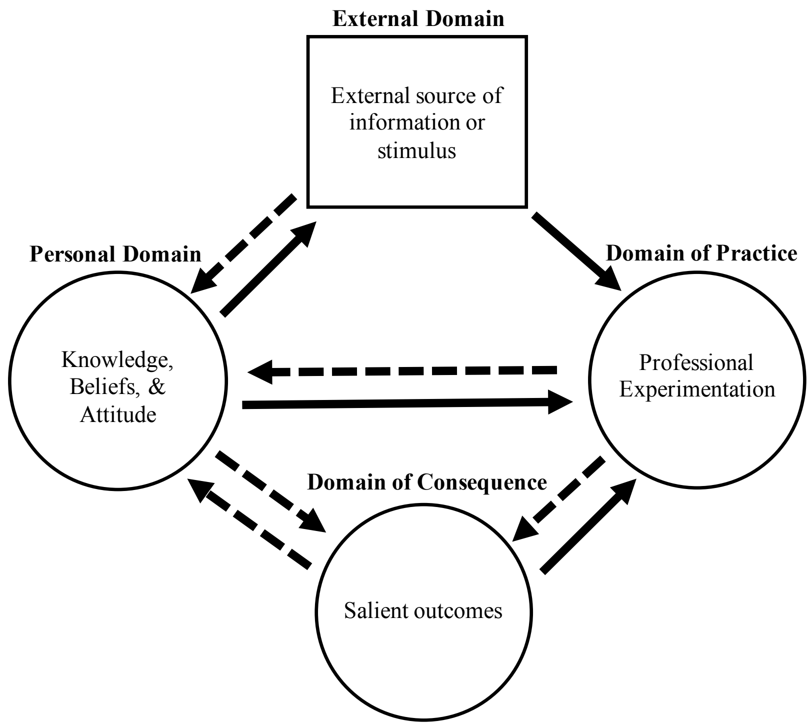


Figure 1.1: Graphical representation of the interconnected model for professional growth described in Clarke and Hollingsworth (2002). Solid arrows represent action or influence and dashed arrows represent reflection.

This study assessed the effectiveness of the materials development process employed in the InTeGrate project as a potential model for professional development capable of influencing instructors' pedagogical beliefs. We investigated the following questions:

1. How are the pedagogical beliefs of instructors altered through the scaffolded development of new instructional materials?
2. How did the materials development process affect different aspects of an instructor's pedagogical beliefs?
3. To what extent did the materials development process affect commonly cited barriers to the long-term adoption of research-validated teaching strategies?

Methods

We employed a convergent parallel mixed methods design for this study. Employing a mixed methods design to characterize beliefs is important because previous research has shown that multiple types of data are necessary to accurately describe beliefs (Pajares, 1992; Richardson, 1996). Mixed methods research combines qualitative and quantitative methods to reach a comprehensive understanding of the phenomenon under investigation, thus providing greater confidence in the study conclusions (R. B. Johnson & Onwuegbuzie, 2004). A convergent parallel mixed methods study implements the quantitative and qualitative instruments during the same phase of the research and integrates the results (Creswell & Plano Clark, 2011). We administered quantitative and qualitative instruments in a two-phase pre-post protocol to analyze how the development of course materials affected participant's pedagogical beliefs. Collecting both qualitative and quantitative data provides an opportunity to determine the degree to which an instructor's description of teaching beliefs (qualitative) converges or diverges with their conceptualizations of those same beliefs as revealed in self-report surveys (quantitative; Creswell & Plano Clark, 2011). A panel of

STEM educators selected the study subjects to participate in the InTeGrate project. All participants were post-secondary instructors from a variety of higher education institutions across the United States. Table 1.1 lists each participants' pseudonym and institutional type.

Table 1.1: Participant List

Pseudonym	**Post-Secondary Teaching Experience (yrs.)**	**Institution Type**
Beth	16	Private
Amy	13	Community College
Jackie	13	Public
Rachel	7	Community College
Ellen	10	Private
Karen	9	Public
Hilary	13	Community College
Mark	13	Community College
Sandy	25	Private
Tyler	5	Public
Mandy	12	Public
Linda	13	Private
John	11	Public
Becky	8	Private
Tammy	12	Private
Lori	2	Private
Nicole	6	Community College
Anne	3	Public
Andrew	8	Community College
Lauren	32	Community College
Henry	10	Community College

Qualitative Instrument

Qualitative data were generated by application of the Teacher Belief Interview (TBI, Luft & Roehrig, 2007), a semi-structured instrument consisting of seven questions (Table 2) designed to assess how instructors characterize various aspects of science teaching and learning. We used the protocol outlined by Luft and Roehrig (2007) to transcribe and code the interview recordings. The coding protocol employs a thematic content analysis by guiding the researcher to assign individual responses to one of five categories: traditional, instructive, transitional, responsive, and reform-based instruction; therefore, any change, in even a single category, is likely practically significant. The traditional and instructive categories encompass participant's statements that often describe science as a series of rules and facts, and represent the most instructor-centered pedagogical beliefs. Responses that fall in the transitional category view science as consistent, connected, and objective, and represent pedagogies that typically include a mix of traditional and reformed strategies. The responsive and reform-based categories include answers dominated by student-centered pedagogical beliefs that characterize science as dynamic and integrated within a social and cultural construct. For comparative purposes each of the five TBI codes can be assigned a numeric value of one to five (1=traditional; 5=reform-based) to generate a total TBI score for each participant. Scores can range from seven to thirty-five representing the end-members for instructor-centered and student-centered pedagogies respectively.

Table 1.2: Seven semi-structured interview questions from the Teacher Belief Interview.

	TBI Question
1	How do you maximize student learning in your classroom?
2	How do you describe your role as a teacher?
3	How do you know when your students understand?
4	How do you know what to teach and what not to teach?
5	How do you decide when to move on to a new topic in your class?
6	How do you know when learning is occurring in your classroom?
7	How do your students learn science best?

Pre-development interviews were conducted shortly after participants were selected to work on the project, but before they had any formal interaction with their teams or with project leadership. Post-development interviews occurred when participants completed the development, piloting and final revision of their team's materials. Consequently, the time between participants pre- and post-development interview typically fell between one and two years. Interviews of participants took place in person or over the phone, and consisted of two phases: (1) a preliminary round of informal questioning, and (2) the semi-structured format of the TBI. The semi-structured portion typically lasted between 20 and 45 minutes with an average time of 30 minutes. The preliminary phase began with a short description of the interview process. We asked interviewees to comment on their progress with the development of their instructional materials. Most discussions then addressed the nature of their current classes. We chose this initial questioning strategy to ease participants into the interview and to put them in the frame of mind to think about their teaching. Subsequently, we asked the seven TBI questions in the same order with the only exceptions representing occasions when an interviewee's responses dictated a deviation from the pre-determined sequence. The interviewer took the stance of a passive participant in the second phase of the interview to limit any potential for biasing the interviewees' comments and to preclude leading the participants. Probing questions were asked after the instructor had introduced a topic and most of these supplementary questions guided interviewees to elaborate on the meaning of commonly used pedagogical terms such as "hands-on activities", "engaged", and "facilitating" or to describe what a certain activity would look like in their class. These clarifications were necessary to ensure that we could accurately code each TBI question. Pre- and post-interviews were conducted in the same way with the exception that the post-development interview also included the additional final question, "Do you think that working on the InTeGrate project has affected your teaching in any way?"

Qualitative Reliability

Interviews in the current study were co-coded by two researchers that resulted in a Cohen's Kappa value of inter-rater agreement of 0.84 which qualifies as 'good' co-coder agreement (Cicchetti & Sparrow, 1981). One coder was the primary investigator while the other was completely independent of the research and the study's participants. It was not possible to assess changes in teaching practice for comparison with changes in teaching beliefs due to the broad geographic distribution of participants and the lack of control observations. However, others have reported that the TBI shows a strong significant correlation with classroom observations of teaching practice (Ryker, McConnell, Bruckner, & Manduca, 2013). Ryker et al. (2013) compared TBI scores and observations of geoscience instructors' classroom practices using the Reformed Teaching and Observational Protocol

(RTOP). RTOP (Sawada et al., 2002) describes the degree of reformed teaching in a classroom based on five subscales (Lesson Design and Implementation, Content: Propositional Knowledge, Content: Procedural Knowledge, Student/Student relations, and the Instructor/Student relationship) and has well documented validity (M. Piburn, Sawada, & Turley, 2000; Sawada et al., 2002) and reliability (Amrein-Beardsley & Popp, 2012; Sawada et al., 2002). TBI scores had a strong positive correlation (R^2 = 0.60) with RTOP scores (Ryker et al., 2013). This suggests that teaching beliefs measured by the TBI have complementary classroom practices.

Quantitative Instrument

The Beliefs about Reformed Science Teaching and Learning (BARSTL) survey (Sampson & Grooms, 2013) measures an instructor's construct of reformed pedagogical beliefs. The BARSTL survey consists of thirty-two statements evenly divided among four sub-categories: 1) How people learn about science; 2) Lesson design and implementation; 3) Characteristics of the teacher and learning environment; and, 4) Nature of the science curriculum. Instructors respond to each statement with one of four options on a Likert survey: strongly agree (SA), agree (A), disagree (D), or strongly disagree (SD). Sixteen of the thirty-two items were traditionally-phrased statements and were reverse scored (i.e., the more traditional the instructors teaching beliefs, the lower their score). Traditionally phrased items represent a post-positivist perspective of knowledge emphasizing the transmission of knowledge from instructor to student. The other sixteen statements were reform-phrased items that represented a constructivist view that the individual creates knowledge and it can be unique to each student. The higher the total score on the BARSTL survey the more aligned an instructor's pedagogical beliefs were with reform ideologies (Sampson & Grooms, 2013). The BARSTL survey requires participants to respond to Likert items that generate ordinal data. This form of data is commonly used in the social sciences to create overall scores by assigning the ordered responses a numerical value (Boone et al., 2011) and these scores are used for various parametric statistical tests. However, ordinal data often do not meet the fundamental assumptions necessary to apply parametric statistical tests (Bond & Fox, 2007). Ordinal data are not linear, equal interval, and often are not normally distributed (Bond & Fox, 2007; Boone et al., 2011). Additionally, ordinal data are not additive and simply changing the numeric values assigned to categories can severely affect the overall scores and interpretation of respondents' performances. We did not use raw scores from the BARSTL surveys that could have been obtained from assigning numeric values to the four-item Likert scale. Instead, we report BARSTL scores as scaled values, referred to from this point as *person measures* that were calculated using Rasch analysis. Winsteps Rasch analysis converted raw survey scores into person measures in an effort to better characterize results from the BARSTL survey. Person measures are expressed in logits (log-odds units), which are equal-interval units that can be applied to parametric statistical tests because they avoid many of the issues of the non-linearity of rating scales as well as the non-linearity of raw survey data (Boone, Staver, & Yale, 2014).

George Rasch (1960) developed Rasch analysis that represents a one-way probabilistic approach based on Item Response Theory (IRT). The application of Rasch measurement in the social sciences has been most notably discussed in the 1967 Invitational ETS conference and in a wide-variety of subsequent publications (Andrich, 1978; Boone et al., 2011; Choppin, 1985; Libarkin & Anderson, 2005; Linacre, 1998, 2006, 2010; Siegel & Ranney, 2003; R. M. Smith, 1991; Tong, 2012; Mark Wilson & Adams, 1995; Wright & Stone, 1979; Wright, 1977, 1984). Rasch modeling is both norm-referenced (comparing individuals to the group) and criterion-referenced (measured according to specific standards; Siegel & Ranney, 2003). This is accomplished by considering the difficulty of the item on a continuum, the participant's response to that item (ability), and the probability that a participant will choose a response (Boone et al., 2014). Winsteps software (Linacre & Wright, 2000) was used to compute person measures for the BARSTL survey results for all participants in this study, and to transform the original logit range to a more comparable, but still linear, scale ranging from 0 to 100. Just as with the raw survey data, a higher person measure indicates more reformed pedagogical beliefs while a low person measure indicates beliefs that are more traditional.

Wright Maps (also known as person-tem maps) were constructed using logit values for person measures and items (Wilson & Draney, 2000). Wright Maps are a method of displaying complex rating scale and test data (Boone et al., 2014) and place items and person measures on the same linear scale, much like a ruler. Wright maps can simultaneously display both items and person measures because Winstep Rasch analysis transforms both data types into logits (Boone et al., 2014; Linacre & Wright, 2000).

BARSTL Survey Reliability and Validity

In their validation of the BARSTL survey Sampson & Grooms (2013) showed that the instrument was reliable (α=0.77, p=0.001). Item review panels, sub-scale correlations, and a confirmatory factor analysis suggested that it had reasonable content and construct validity (Sampson & Grooms, 2013). In addition to calculating standard reliability values, Winsteps software allows the researcher to determine both the person and item reliabilities using the scaled logit values as opposed to the non-linear values commonly used in a Cronbach alpha calculation. Any measurement of reliability is limited to considering how items are related to one another (Cortina, 1993; Cronbach, 1951). Rasch analysis simultaneously considers the reliability of items and an individual's responses to those items to generate a reliability values for persons (similar to the Cronbach alpha) and a value for item reliability (similar to a Cronbach alpha but pertains to the inter-item relationship). Winsteps Rasch analysis yielded acceptable reliabilities and separation indices using participants' responses to the BARSTL survey for persons and items (Table 3) indicating that the instrument displays good internal reliability. Separation indices greater than one indicate that the sample size is sufficient to locate items along the trait of interest (Linacre & Wright, 2000; Table 3); reformed pedagogical beliefs. Winsteps Rasch analysis also allows the researcher to quantify pre- and post-survey bias by calculating differential item function (DIF) values for all survey items. DIF is a measure of how consistently an item or survey measures a group of respondents or if a particular group of respondents preferentially performs better or worse on a particular item (Boone et al., 2011). None of the thirty-two items in the BARSTL survey display significant DIF values ($p<0.05$). Reliability values, separation indices, and DIF values calculated in Winsteps suggest that the BARSTL survey has reasonable construct validity as a measurement instrument (Boone et al., 2011; Linacre & Wright, 2000; Linacre, 2010).

Table 1.3: BARSTL Person and Item Reliability Values

	Pre-Development		Post-Development	
	Person	Item	Person	Item
Separation Index	1.90	3.11	5.03	3.46
Reliability	0.78	0.91	0.86	0.92

Results

Teacher Belief Interview Results

A majority (71%) of the participants show gains on their TBI scores from pre- to post-development interviews (Figure 2). TBI scores are approximately normally distributed from the both phases of the interviews based on visual analysis of Q-Q plots. A pre-development kurtosis value of 0.548 (standard error of 0.972) and a skewness value -0.999 (standard error of 0.501); and a post-development kurtosis value of –0.545 (standard error of 0.972) and skewness value of -0.415 (standard error of 0.501) also support the assumption of a normal distribution. No outlier values were detected using the outlier labeling rule and g-factor described in Hoaglin & Iglewicz (1987). Pre-development TBI scores range from 17 to 32 with a mean score of 25.8. Post-development TBI scores range from 23 to 32 and have a mean score of 28.2. There is a statistically significant ($t(20)=3.43$, $p=0.003$, $d=0.70$) change toward more student-centered responses with a moderate to high effect size. Participants with the lowest initial TBI score exhibited the greatest normalized gains in the post-development TBI score (Figure 3). Fifteen of the 21 participants improved their TBI scores representing a shift toward more student-centered pedagogical beliefs; five instructors showed no change in their beliefs, and one instructor’s TBI score decreased in their post-development interview.

Responses to four TBI questions (How do you maximize learning in your class? How do you describe your role as a teacher? How do you know when learning is occurring in your classroom? How do your students learn science best?) do not show significant shifts between the pre- and post-development interviews (Figure 4). Responses to the remaining three questions (How do you know when to move on to a new topic in class? How do you decide what to teach and what not to teach? How do you know when your students understand a concept in class?) showed the greatest divergence among the participants and all demonstrate a significant shift toward more student-centered codes for the post-development interviews (Figure 4).

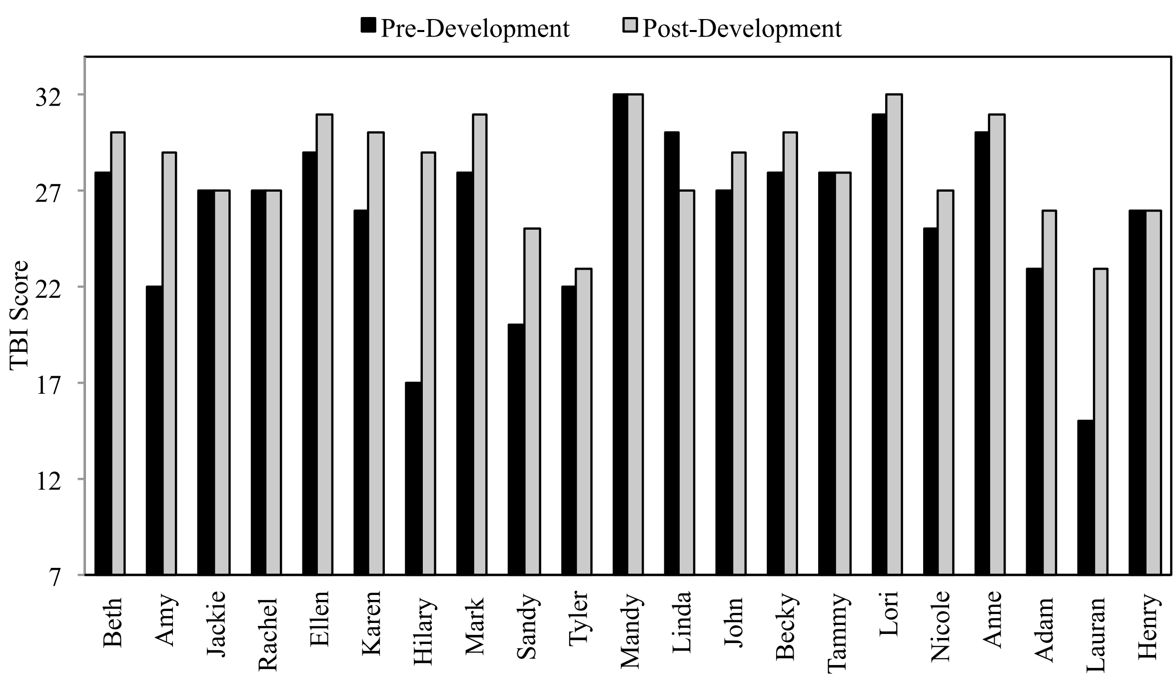


Figure 1.2: Graph displaying participants' pre- and post-development TBI scores.

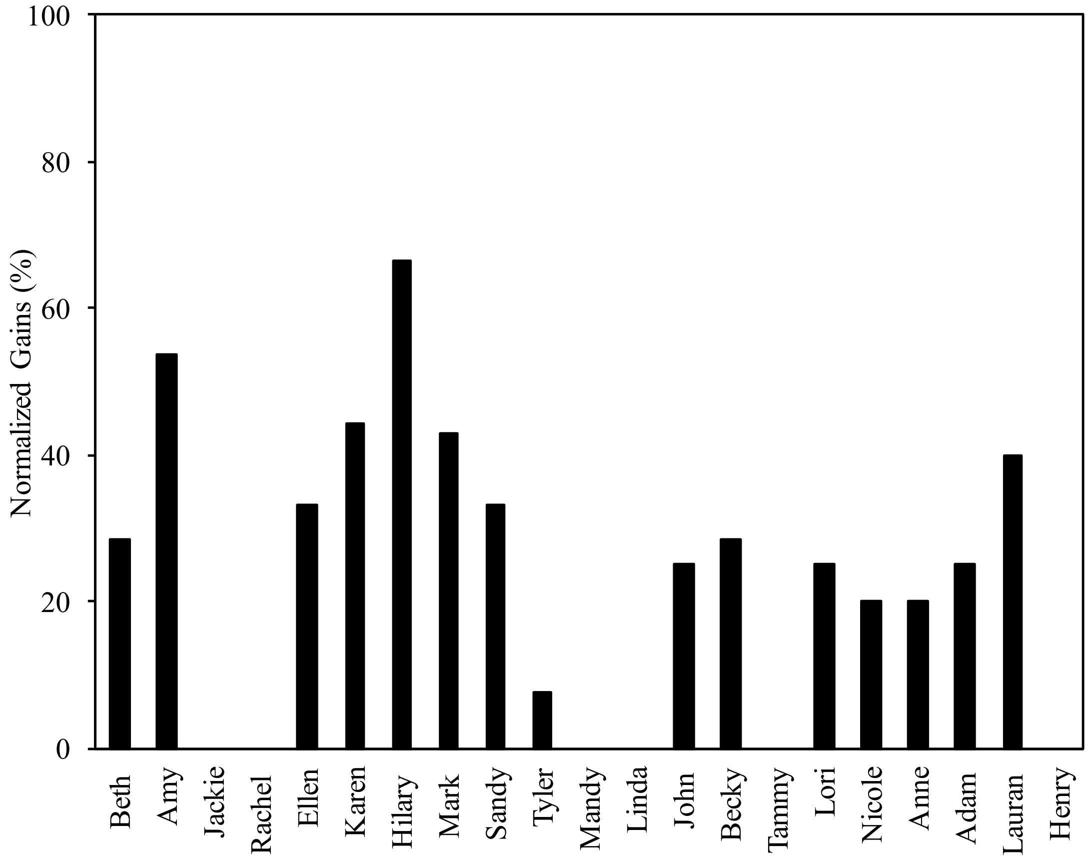


Figure 1.3: Normalized gains for the pre- to post-development interviews.

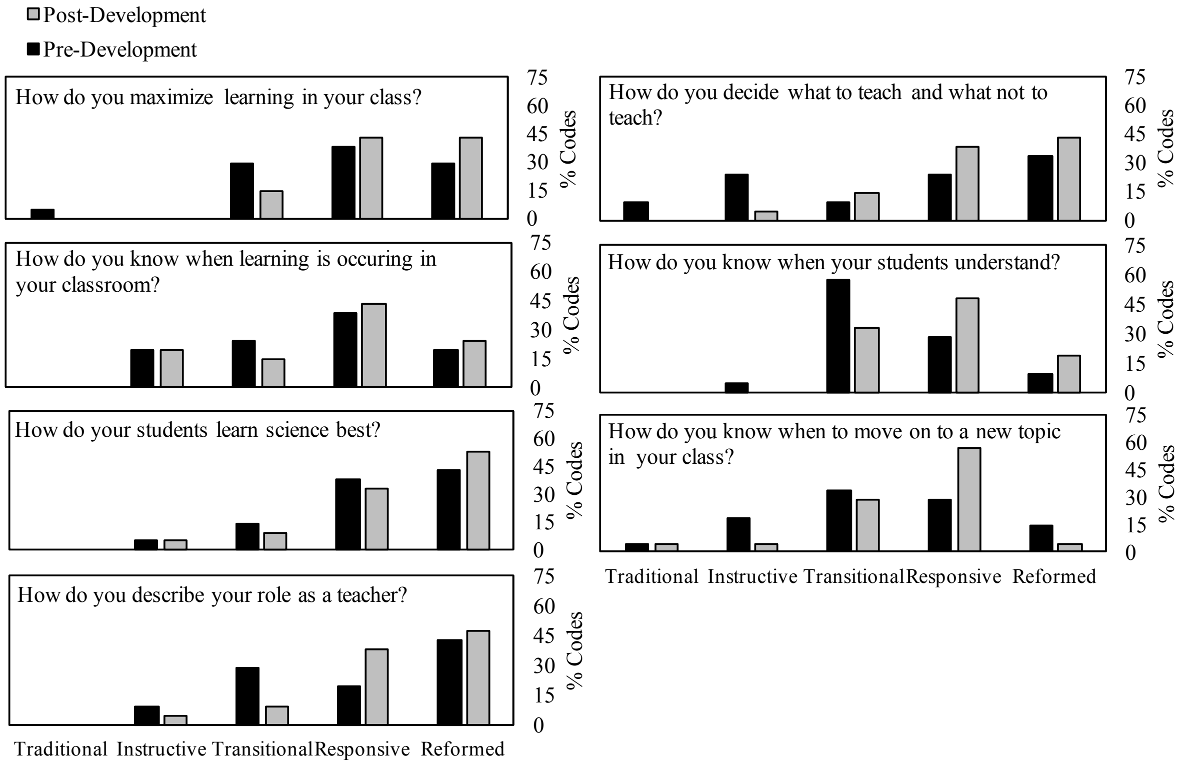


Figure 1.4: Histograms showing the distribution of codes on the pre- and post-development interviews for all seven TBI questions. Questions in the left column show little change between pre- and post-development surveys while questions listed in the right column show more significant change.

Description of variability in TBI responses

When asked how they maximized learning in their classrooms, responses ranged from transitional where Sandy stated: "I try to make them feel very open and I hope to convey to them that I welcome questioning and commenting at any time…" to more student-centered responses where Anne noted: "I think setting up the environment where they can take responsibility for their own learning and really process things on their own…" Sandy focused on creating positive environment (transitional) while Anne focused on creating an environment where students can take charge on their own learning (reformed).

When questioned about how they view their roles as teachers, some participants described a transitional role, for example, Amy reported "I've got to make sure I tell them that I struggled with certain things when I was a student…I share things like that with them so they realize that they are at a normal place…" Nicole's response provides a more reformed, student-centered, conceptualization of the role of a teacher "I customize my classes to what I think students will most relate to. As long as they met the student learning outcomes I'll still customize it so that I feel that they can relate to the material and carry it kind of away and use it in their real life."

The question "How do you know when learning is occurring in your classroom?" seeks to clarify an instructors conceptualization of assessment (Luft & Roehrig, 2007). When discussing how to assess learning Hilary specified that she emphasized "The write-ups for the labs that they turn in" and with "bigger assessments [exams] at the end of three or four chapters…" This is an instructor-centered response where she primarily focuses on formal summative assessments. Conversely, Mandy stated "[In class] when students are starting to ask the level of detail of questions showing that they are really noticing the nuances of the data…" This description is more student-centered, focuses on the student's and their knowledge, and demonstrates that she is looking for students to initiate a significant interaction with either her or possibly asking these questions with their peers.

Participants responses to "How do your students learn science best?" were mostly student-centered. However, the responses ranged from transitional, such as Tyler who stated that his students learn science best “by actually working problems” and “if they sort of get the chance to apply what they have learned” to more student-centered responses from Jackie asserted that her students learn best “when they are doing something that has no right or wrong answer. When things that they do have multiple answers, you [students] can ask multiple questions” and she elaborates further by stating that when “The answer is not the end. An answer can be questioned…” is an experience when where students often best learn science.

Responses to three questions (How do you know when to move on to a new topic in class? How do you decide what to teach and what not to teach? How do you know when your students understand a concept in class?) were the more likely to be coded as “traditional” in the pre-development interviews. When asked about when to move on to a new topic the responses ranged from an instructor-centered response from Hilary during her pre-development interview that focused on schedule and time constraints “I really can’t spend more than the…on whatever topic we are scheduled before there is just so much material that needs to be covered.” Moving along the coding spectrum, we can examine a transitional response from Tyler’s pre-development interview that included the students in the decision about when to move on “I pause when they are producing some results…So there are sort of intermediate steps along the way that I can see if they identify this rock sample correctly…” Unlike Hilary’s response, Tyler based his decision to move on partly on the students. An example of a reformed, student-centered reply can be read in Ellen’s post-development interview response “In that case I might run the lecture over into the next time or have a couple of slides, because that’s the point is to get it and the that obviously helps me next time to redesign the lecture.” Ellen is not only considering whether or not students “get” the information but also verbalized the option of revisiting concepts and redesigning future lessons based on her current students’ experiences with the content.

The responses to the question “How do you decide what to teach and what not to teach?” show a similar variability. For example, a traditional response from the pre-development interviews comes from Hilary who stated “Well of course time is always a constraint…and then if there is extra time then I have more that I can teach”. She also noted, “I have topics that I want to make sure that I cover over the course of the semester.” This indicated a traditional adherence to syllabus and time constraints. A transitional response, such as John’s, from his pre-development interview where he declared “I think of what skills and knowledge in geology contribute to understanding…that’s what focuses my teaching, because majors make up 90% of my class.” John’s statement described a distinct shift away from instructor-centered decisions and situational factors to thinking more about his students when he chooses class topics. A reformed response comes from Anne’s post-development interview when she described something she had done recently in one of her courses. She said “I kind of modified the class the second time around to give them [students] a lot more chances to practice peer teaching so that they felt more comfortable before teaching their own classes.” In this response, Anne not only considered the needs of her students, (much like John), she also used that student-generated feedback to modify future iterations of her class.

Participant responses were again variable ranging from instructor-centered to student-centered when asked about assessing students’ understanding of concepts. An instructor-centered response was evident in Sandy's pre-development description of this process "I guess, you know, I give them [students] quizzes, pop quizzes, but it is not so much to gauge what they understand as to keep them, you know, on track and constantly engaging in the material." Again, in a similar phrasing for instructor-centered responses to learning, she focused on formal assessment delivered by the instructor. Reformed responses to this questions are defined by an instructors ability to determine when students have shown the application of knowledge outside of class or in a novel setting (Luft & Roehrig, 2007). An example of this can been read in Mark's post-development response "I know that they've [students] understood a concept if they can relate something in the real world, based on that concept…" Tammy's post-development response also places an emphasis on the application of knowledge outside of classroom constraints: "I guess when they tell me something that I don't know, or something that I haven't thought about or they have an inspiration or something that makes them want to go further than the original constraints of the assignment."

Responses from participants showing the highest gains.

Instructors with the greatest normalized gains between pre- and post-development interviews display the most significant changes in their responses to interview questions. Amy, Karen, Hilary, Mark, and Lauren all show a 40% or greater normalized gain between their pre- and post-development TBI interviews (Figure 2). We can gain some insight into individual instructors changing beliefs by examining Amy and Hilary's responses to TBI questions. These two instructors exhibited the broadest change between the pre- and post-development interviews. For this study a pedagogically significant change is one in which the participants overall code for a response shifts among the three broad categories of pedagogical beliefs (instructor-centered, transitional, and student-centered) defined in Luft & Roehrig (2007). An example of a significant shift would be if an instructor initially responded to a question in a manner that received a traditional code, while their post-interview response was transitional. Amy showed significant shifts toward more student-centered beliefs in her responses to several questions. Amy described her role as a teacher in the pre-development interview stating "I've got to make sure I tell them that I struggled with certain things when I was a student" and "I try to share things with them so they realize that they are at a normal place." Her words illustrate transitional pedagogical beliefs by focusing on establishing a welcoming and positive classroom environment. Her response shifted from transitional to student-centered in the post-development interview where she described her role as "fostering questions, having them ask questions and be really curious about the world" and she elaborates on this point stating, "I want them to be curious and I want to foster that curiosity." Amy had instructor-centered ideas about how to decide when to move on in class during the pre-development interview. Initially she stated "I focus mostly on time constraints" and "I assign a schedule, it's not incredibly rigid, but it is structured." This instructor-centered focus became more student-centered in post-development because Amy incorporated feedback from students with the possibility of revisiting topics. For example, she stated "when they tell me they are still not understanding then I will build something into the future classes" and "If they [students] need an extra day to review or synthesize everything then I am okay with adding an extra day to do that." This student-centered theme also carried over in her response to assessing learning in the classroom. In the pre-development interview she described the process of assessing learning very transitionally "I try to arrange the course material so that things sort of repeat themselves" and she does not continually assess students, "I don't necessarily assess students along the way…" During the post-development interview, Amy describes more student-centered practices stating "I listen to how they are communicating with their group mates" and she elaborates on this theme stating that she knew they were learning when "they were questioning each other and working through it".

Hilary also showed significant shifts towards more student-centered responses on several questions (Figures 2, 3). During the initial interview, Hilary described how she assessed student understanding, "There are always the assessments, and with assessments I do try to mix it up a little bit…" This suggests a focus on measures given by the instructor. She went on to say, "If they actually come to a…some kind of solution [correct] to this problem…" and here she was not only focusing on assessments but on the correct solutions to those assessments to gauge understanding. Together these two responses align with a transitional pedagogical belief system. However, in the post-development interview Hilary's response shifted to a more student-centered view, stating "when they can come up with a very…an original idea or thought" and "when they [students] are getting the information and maybe making comments on how this is useful or important to them…" The question about determining what to teach in class received the most traditional responses (Figure 4). During her initial pre-development interview Hilary responded in an instructor-centered way where she stated "time is always a constraint…I do have topics that I want to make sure I cover over the course of the semester. Her response on the post-development interview shifted towards student-centered where she stated that she chooses topics “that they can internalize and see the importance of those things, whether you are a geologist, a teacher, or construction worker things that are going to affect their everyday life.” This response showed a student-centered focus on science literacy and she elaborated further when asked why she chooses topics in this way “It’s not because they are students, but because they are humans.” When asked during the pre-development interview how she assesses learning in the class Hilary stated “I kind of have these bigger assessments at the end of three or four chapters.” Initially she has an instructor-centered focus on summative assessments given after she covered topics in class. This view changed during the post-development interview where she responded in a more student-centered manner by stating that she assessed learning by “the discussions really, again they [students] give me a very good idea of what’s going on during a week to week basis.” She was focusing on interactions among students to determine if learning occurred and she was continually doing this as opposed to focusing on summative assessments given after the fact.

BARSTL Survey Results

Figures 5 graphs person measures calculated by Winsteps Rasch analysis and Figure 6 shows each participant’s normalized gains. BARSTL scores, as represented by person measures from the pre- and post-development phases, are normally distributed based on the visual analysis of Q-Q plots. Pre-development kurtosis value of -0.325 (standard error of 0.972) with a skewness value of -0.157 (standard error of 0.501), and a post-development kurtosis value of 0.262 (standard error of 0.972) with a skewness value of 0.562 (standard error of 0.501) also support the assumption of a normal distribution. No outlier values were detected using the outlier labeling rule and g-factor described in Hoaglin & Iglewicz (1987). There was a statistically significant ($t(20)=2.74$, $p=0.013$, $d=0.50$) improvement toward more reformed responses on the survey with a moderate effect size. Fifteen of the twenty-one participants' BARSTL survey scores increased on the post-development survey. Pre-development person measures range from 47.17 to 64.21 with a mean of 56.70. Post-development person measures range from 47.39 to 69.89 with a mean of 59.13. In contrast with the TBI scores, participants with the lowest initial person measures do not show the most significant gains (Figures 5, 6). Many of the participants do show improvement but there is little correlation between the degree of improvement and their pre-development survey score.

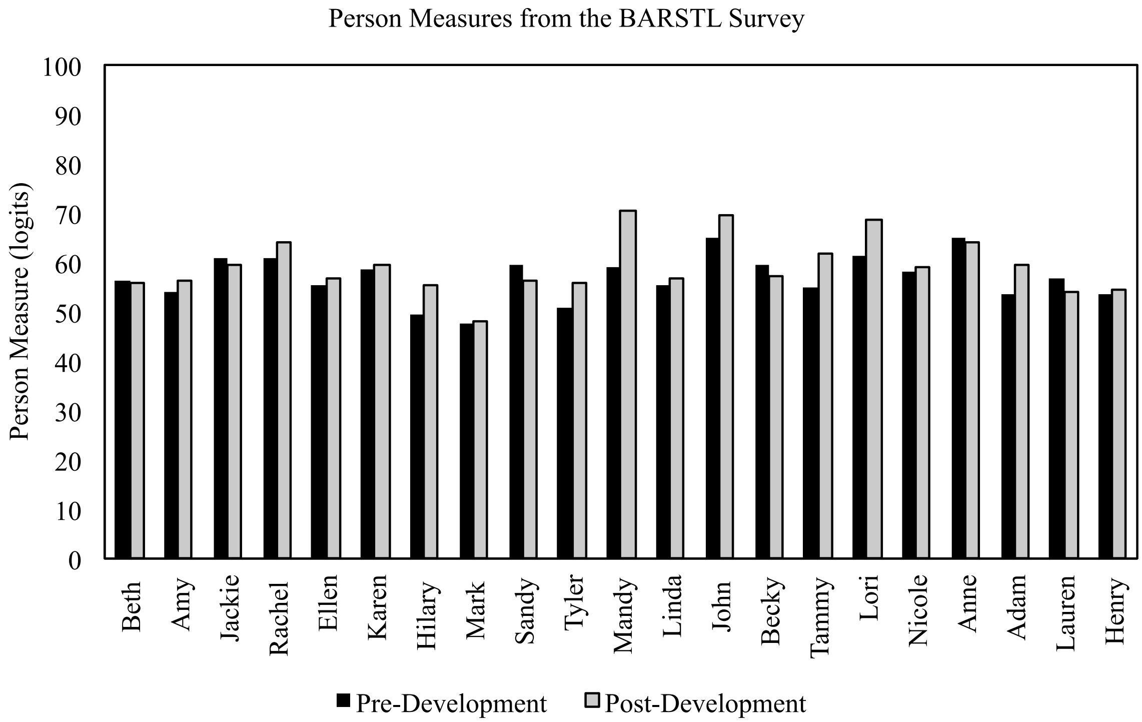


Figure 1.5: Graph displaying participants' pre- and post-development BARSTL survey measures.

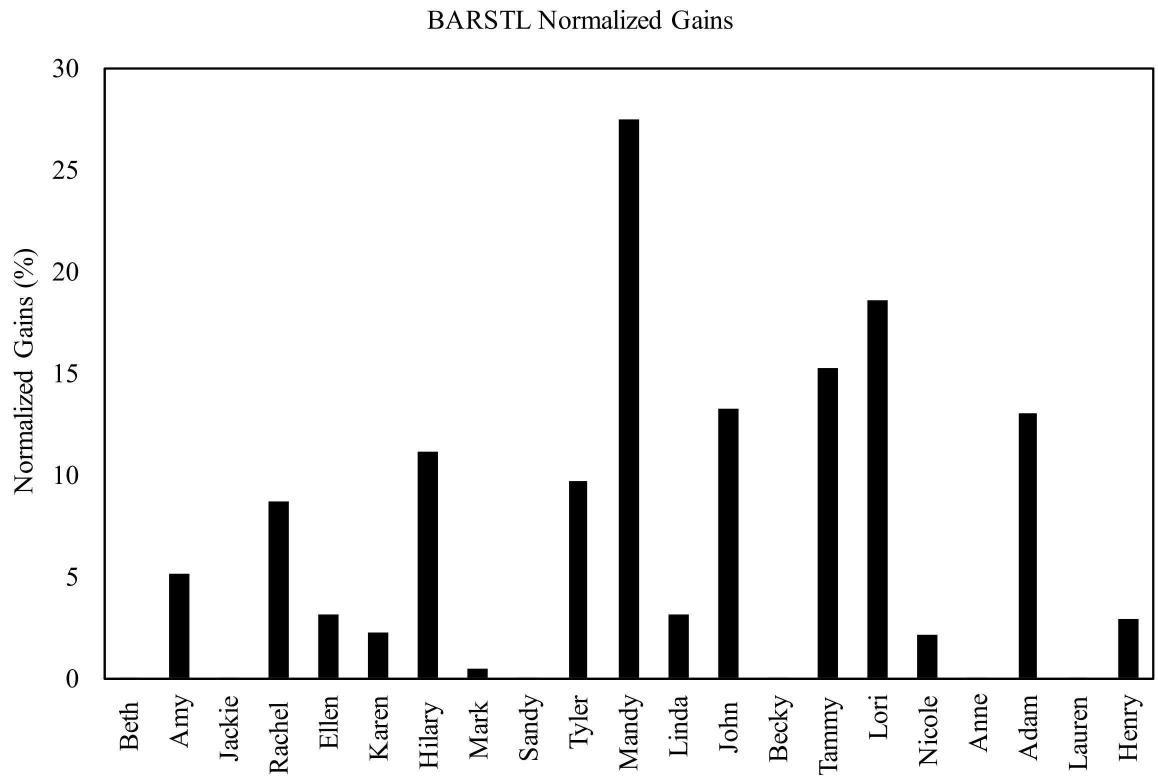


Figure 1.6: Normalized gains for the pre- and post-development BARSTL surveys.

Wright Maps in Figure 1.7 display the person measures of each participant and the item measures for the pre- and post-development surveys. The right side of each Wright map displays item measures and item position. Higher item measures (located toward the top of the map) represent the most difficult items and items at the bottom of the map represent the least difficult items. Only the higher scoring participants agreed or strongly agreed with difficult items while both high and low scoring participants agreed with less difficult items. The BARSTL survey contains equal numbers of reform-phrased and traditionally-phrased items and the latter are scored in reverse. Therefore, participants who stated that they disagreed or strongly disagreed with traditionally-phrased items would obtain a higher total point value and be placed higher on the Wright map (see X values on left side of each Wright map in Figure 1.6). The left side of the Wright Map places all of the participant's person measures (shown in logits) on the same linear scale as the items. More reformed instructors will have a higher score and be located toward the top and instructors with more traditional teaching beliefs will have lower scores and be located toward the bottom. Participants person measures calculated from the post-development surveys show a significant shift toward more reformed pedagogical beliefs (positioned higher on scale along left side of figure in post-development result; Figure 7).

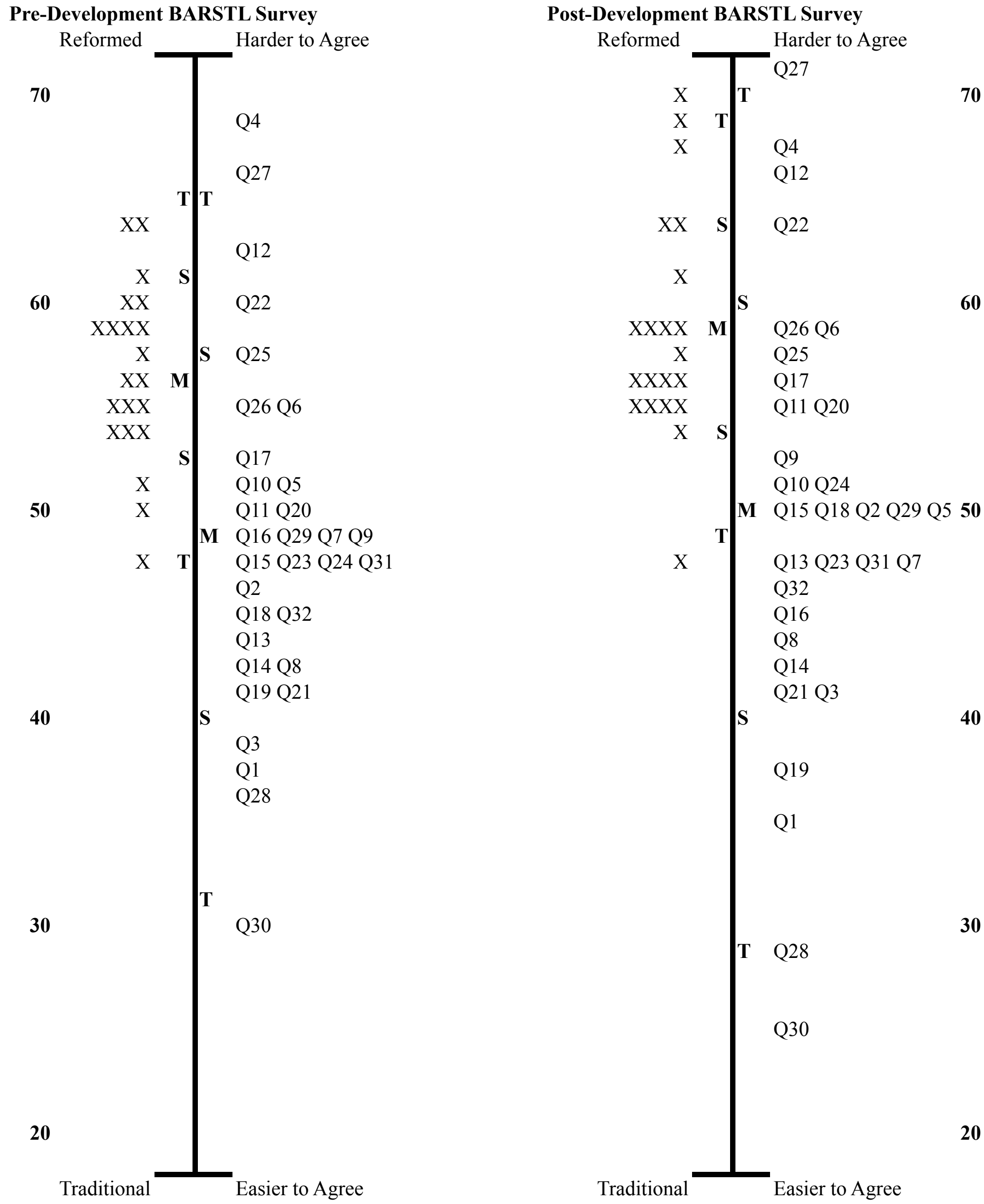


Figure 1.7: Wright maps created using Winsteps software for the pre- and post-development BARSTL surveys. Item measures are displayed on the right side of each map and participants' scores (person measures) are displayed on the left. Along the line dividing items and participants, the letter "M" represents the mean item or person measure, "S" represents one standard deviation away from the mean, and "T" represents two standard deviations away from the mean.

The most difficult item on the pre-development survey was the traditionally-phrased Q4: “Students are more likely to understand a scientific concept if the teacher explains the concept in a way that is clear and easy to understand.” Q4 is located in the “How people learn about science” subcategory and represents a post-positivist view of knowledge. The least difficult item was the reformed-phrased Q30: “The science curriculum should help students develop the reasoning skills and habits of mind necessary to do science.” Q30 remained the item with the most agreement on the post-development survey, while Q4 is replaced by another traditionally-phrased item, Q27: “Students should know that scientific knowledge is discovered using the scientific method.” as the item with the fewest participant disagreements. Q27 is located in the “Nature of the Science Curriculum” subcategory and, similar to Q4, represents a strongly post-positivist view of knowledge in which the only way to create scientific knowledge is through a prescribed common method.

Analysis of Wright maps allows the identification of items that represent the statistical significant difference between groups by comparing the overlap of mean person measures from the pre- and post-development surveys (Boone et al., 2014). Responses to four statements comprise most of the difference between the pre- and post-development BARSTL surveys (Figure 6). These four statements are: Q6 (traditionally-phrased), “Learning is an orderly process; students learn by gradually accumulating more information about a topic over time”; Q17 (reform-phrased), “Students should do most of the talking in science classrooms”; Q25 (reform-phrased), “A good science curriculum should focus on only a few scientific concepts a year, but in great detail; and Q26 (traditionally-phrased), “The science curriculum should focus on the basic facts and skill of science that students will need to know later”. Items Q25 and Q26 explicitly deal with the design of the science curriculum, which is influenced by situational factors, and items Q6 and Q17 can be influenced by situational factors but are also shaped by other aspects of an instructor’s pedagogical beliefs.

Discussion

Twenty of twenty-one participants showed increases between their pre- and post-development scores on either the TBI and/or BARSTL survey (Table 4). Nearly half (10) of the participants showed congruent increases between their scores on each instrument and another four exhibited an increased score on one instrument but no change on the other (Table 4). Participants who entered the InTeGrate project with the most traditionally aligned beliefs recorded a shift toward more reformed pedagogical beliefs. Positive gains in scores were also recorded for participants who already had student-centered beliefs when they began the project. This suggests that the scaffolded materials development process in the InTeGrate project is an effective method of faculty change for instructors who exhibit a range of pre-existing pedagogical beliefs. A key component of the InTeGrate project is the collaborative development of new instructional materials. The fact that the project supported collaboration of peers with similar teaching backgrounds and experiences sets InTeGrate apart from many traditional professional development experiences. Rogers (1995) posits that such teams are efficient at passing information among group members. This is important for the InTeGrate project because much of the communication occurs at the team level (e.g. developing materials, discussing pilot experiences, and sharing information). Dissemination amongst peers is also important because team members started the project with different pedagogical beliefs. Therefore, participants with more reformed beliefs, or who had more experience using student-centered activities in their classes, were able to efficiently share their “lessons learned” from those experiences.

Table 1.4: Merged Results from the TBI and BARSTL survey.

Participant	**BARSTL**	**TBI**
Beth	-	+
Amy	+	+
Jackie	-	
Rachel	+	
Ellen	+	+
Karen	+	+
Hilary	+	+
Mark	+	+
Sandy	-	+
Tyler	+	+
Mandy	+	
Linda	+	-
John	+	+
Becky	-	+
Tammy	+	
Lori	+	+
Nicole	+	+
Anne	-	+
Andrew	+	+
Lauren	-	+
Henry	+	

Six of the participants showed inconsistencies between their pre- and post-development phases of the BARSTL survey and TBI. Five participants had a decrease in BARSTL survey score but showed an increase in their TBI score. This discrepancy between the qualitative and quantitative data could be the result of limitations in the design of the BARSTL survey or an unavoidable limitation of Likert instruments designed to measure complex psychometric constructs as any form of personal belief is often difficult to measure (e.g. Fang, 1996; Fishbein,. and Ajzen, 1975; Jones & Carter, 2007; Rokeach, 1968). However, one of the reasons for employing Rasch analysis was to minimize the negative effects of quantitatively analyzing ordinal surveys. The validity and reliability of the BARSTL survey has not been rigorously confirmed outside of the initial design and validation study by Sampson & Grooms (2013). Consequently, differences in the results may reflect limitations in the BARSTL survey used in the context of this study. We interpret the more consistent changes in TBI scores to be due to the more effective manner in which pedagogical beliefs are characterized using qualitative methodology. While interviews can be considered a form of self-report, interview data tend to be a more authentic representation of real-world teaching practices (Luft & Roehrig, 2007; Ryker, 2014). Providing instructors with the opportunity to elaborate and describe their views on teaching has the potential to more accurately capture their beliefs (Ambrose & Clement, 2004; Munby, 1982). An example illustrating the importance of allowing instructors to elaborate on their practices was evident when asked, “How do you describe your role as a teacher?” Many instructors described their role as a facilitator. Facilitating learning is a characteristic of reformed instruction (Luft & Roehrig, 2007), however, when instructors were probed to elaborate on what they meant by the term "facilitator" stark contrasts with that definition were often revealed. For example, during the pre-development interview Lauren described her role as "I give them the material or I set up the demo, or maybe I will have to do some lecturing as well". This description is strongly traditional and not aligned with reformed pedagogical beliefs despite using a descriptor often associated with reformed teaching strategies. Instructor's teaching practices and beliefs are better reflected by a continuum where the categories of traditional and reformed represents end-members rather than strict dichotomous groups (Smith, Vinson, Smith, Lewin, & Stetzer, 2014). The nuances of a continuum may be more accurately represented by qualitative coding methods like those use for the the TBI than in a Likert-type instrument such as the BARSTL survey.

Qualitative and quantitative results both suggest the greatest change in participants' pedagogical beliefs originate from questions and items partially or directly influenced by situational classroom factors. Situational factors are broadly defined as anything unique to an instructors' academic environment that can inhibit the adoption of new teaching strategies. Common examples include expectations of content coverage, lack of time, departmental teaching norms, student resistance, class size and room layout, and allotted class schedule (Henderson & Dancy, 2007). Comparing mean person measures and the positions of items on Wright maps (Figure 7) revealed four items related to situational factors that comprised the majority of the statistical significance calculated between the pre- and post-development BARSTL surveys. Additionally, during many of the pre-development interviews instructors often referred to dealing with situational factors, or cited such factors as having an impact on their teaching decisions when discussing TBI questions about moving on in class and choosing what to teach (Figure 4). References to situational factors were less common in the post-development interviews during which instructors were more likely to discuss student-centered factors as influences toward choosing topics and deciding when to move on to new material.

The InTeGrate project challenges instructors to situate the design of course materials first within their own classrooms, and then secondly within the classrooms of their collaborators and others who may use the materials. This authentic and practical application may help explain why the pedagogical change shown in the qualitative and quantitative data focused on items and questions influenced by situational factors. Participants' responses to regularly scheduled written reflections completed throughout their development experience provide support for this interpretation. For example, Rachel stated that "Our course structures all being different has been challenging as we develop materials" and Linda (who worked with a different group of instructors), noted "Our different experiences in the classroom, and also teaching outside the classroom, helped us think about multiple experiences". The value of thinking about the implementation of materials in different academic environments was also summarized in Beth's reflection where she acknowledged that her collaborators "helped me see different types of assignments that I could write and ways I could think about my own classroom and students". She went on to discuss the value of thinking about applying teaching materials to different environments "I like that I was forced to think about how an exercise would work in a classroom different from my own with students that were also different". Mandy reflected that her group had to struggle and adapt to "disparate approaches to assessment in our classes due to class size, student populations, etc." Developing materials for a diverse audience can also be beneficial after instructors have had the opportunity to pilot their activities in their classrooms. For example, Amy expressed the view that "I liked classroom testing, that this was done in three different institutions, and the results shared, and how this sparked revisions…" The diverse application of teaching strategies required by participants in the InTeGrate project may better equip instructors with the pedagogical content knowledge and experience necessary to overcome situational barriers. Additionally, information and experiences gained from the piloting of their new materials was readily transferred within each team because group members all had similar teaching backgrounds.

The InTeGrate project fulfilled many of the requirements of effective (sufficient duration; collaboration; rigorous and authentic; provides guidance) and the the advantages of collaboration, authenticity, and rigor were confirmed through participants' statements during interviews and in their written reflections. Participants frequently discussed the value of collaboration stating "Hearing different ways of doing things; expanding the way I think about things", "The best part was sharing knowledge and classroom experiences with faculty from different institutions with diverse students", and "Working with my colleagues to develop the module I learned different approaches that I didn't know or had not tried before". Almost every participant referred to the advantages of working collaboratively and only one participant commented on the team-focused aspect of this project as a negative. Many short-term professional development opportunities are limited in scope and do not have the opportunity to effectively create an environment where instructors are supported to work intensively and to think and about and reflect upon their teaching beliefs and practices. For example, Linda stated "InTeGrate has been really helpful because most other workshops that

I have been to are less involved and you don’t get to spend much time with people. You just grab at stuff and leave. It is a lot less intensive”. Beth presented another important perspective on the value of a rigorous and authentic professional development experience “Instead of saying, nope I am not interested in that activity or I don’t know how to do that activity, the project certainly forced me out of my comfort zone and that was a good thing.” The structured and supportive design of InTeGrate also facilitated multiple cycles of reflection and revision. This is a crucial aspect to the InTeGrate project because these processes drive instructional change within interconnected model (Clarke & Hollingsworth, 2002).

Data from this study allows us to propose a model for how the InTeGrate project promoted instructional change framed by the interconnected model of professional growth (Figure 8; Clarke & Hollingsworth, 2002). Participants’ external domain was being strongly influenced by key components of the InTeGrate project (e.g., workshops, peers; Figure 8). Participants’ desire and interest to work on InTeGrate project was initially driven by their pre-development pedagogical beliefs (Arrow #1, Figure 8). InTeGrate participants were recruited to join a project seeking to create student-centered teaching materials. This characterization served to attract instructors with an interest in reformed instruction and this explains why many of the instructors who applied already had transitional to student-centered pedagogical beliefs. Their initial participation in the project’s workshops (both online and in person), and their interactions with team members, assessment experts, and project leaders represented the earliest opportunities for reflection and revision of their pedagogical beliefs (#2, Figure 8). The InTeGrate project’s rubric requirements and structured feedback from leadership and assessment experts influenced the development of their lessons (#3, Figure 8). Once instructors met the project-wide and pedagogical requirements outlined in the rubric they piloted the new materials in their own classes. The experience of piloting the new materials reinforced and strengthened new-found pedagogical beliefs (#4, Figure 8) in the teaching process but also facilitated reflection leading to revision and refinement of activities and practices for future iterations of the lessons (#5, Figure 8). Evidence of greater student engagement and student learning (#6, Figure 8), as seen in daily formative assessment activities and summative exams, may also have contributed to the evolution of beliefs (#9, Figure 8). Student reactions that showed greater engagement and/or success for specific types of activities undoubtedly also influenced decisions regarding revisions to the pedagogy (#7, Figure 8). Similarly, the success of InTeGrate lessons that were specifically created to promote geoscience literacy in the context of societal issues may cause an instructor to reconsider their teaching and learning goals for other lessons (#8, Figure 8). Different instructors may have been more influenced by some of these factors than others but the interconnected model provides us a means of illustrating the different drivers that lead to changes in teaching beliefs.

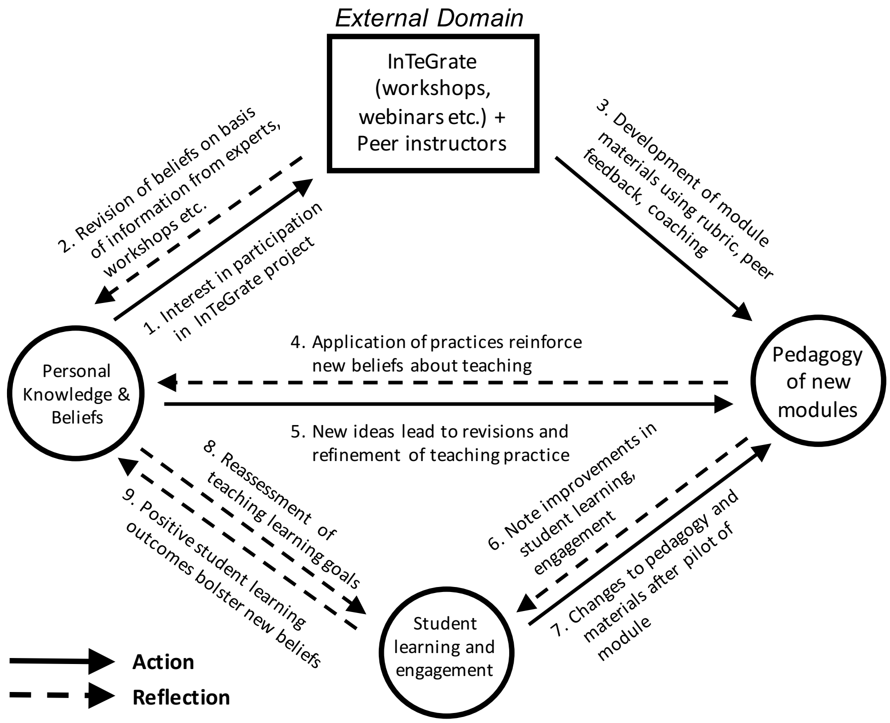


Figure 1.8: Proposed model for instructional change based of off the interconnected model of professional growth developed by Clarke and Hollingsworth (2002).

Study Limitations

Results in this study are limited due to the non-random method participants were chosen and the small sample size. The participants in the study submitted proposals to develop materials for the InTeGrate project, and factors such as teaching experience and pedagogical content knowledge contributed to their selection for the project. All of the participants are post-secondary instructors with several years of teaching experience. Therefore, the results of this study may not be applicable to new instructors with little to no prior experience, or to instructors outside of higher education. While the study generated a sufficient quantity of qualitative information, the numbers of participants used in Winsteps

Rasch analysis remains low. It is recommended that a range of 16 to 36 participants be used to achieve results applicable to ±1 logits within a 95% confidence interval (Wright & Stone, 1979). This minimum sample size is only applicable for exploratory purposes, such as was used in this study, and should not be applied to item design, evaluation, or high-stakes assessment (Wright & Stone, 1979). The distributions of person measures from this study were compared to a larger sample (n=103) of BARSTL surveys completed by similar higher education geoscience faculty. This larger sample of faculty was comprised of other both additional participants from the InTeGrate project working on either whole courses or upper-level classes, and from faculty not related to the InTeGrate project. The mean of this larger sample was similar to the means calculated in this study for the pre- and post-development surveys, and this increases the likelihood that the results of this study are representative of post-secondary geoscience instructors.

The BARSTL survey was originally designed and validated to measure the teaching beliefs of K-12 teachers. This study chose to employ the BARSTL survey because it was explicitly designed and validated around the construct of reformed pedagogical beliefs and thus should align with the TBI. All thirty-two items on the BARSTL survey were reviewed to ensure they were applicable in some way to post-secondary instructors. After a sufficient amount of interview data was collected the items in the BARSTL survey were re-evaluated to ensure that concepts instructors discussed in their interviews aligned with items from the survey. Preliminary explorations into the validity of the BARSTL survey revealed that several items show poor discrimination potentially attributed to ambiguous wording. Additionally, visual analysis of the Wright maps for the pre- and post-development surveys

(Figure 7) shows that many of the items group toward the mean and that several items plot on the same item measure. Items that plot on the same measure are theoretically redundant and could also contribute to a less desirable distribution of scores along the trait of interest (Boone et al., 2011). The narrow distribution of items revealed in the Rasch analysis could also be the likely cause for the correspondingly narrow range of person measures despite the broad range of pedagogical beliefs revealed in the participants' interviews.

Conclusion

The process of materials development employed in the InTeGrate project incorporates many of the characteristics of effective professional development and influences multiple domains within the interconnected model of professional growth. Results suggest that incorporating several of the characteristics of effective professional development into a single opportunity can mitigate commonly cited barriers to the long-term adoption of reformed teaching strategies. The overarching design of the InTeGrate project created a scaffold structure where participants were encouraged to think differently about their teaching, teach differently, and to be continually reflective. Consequently, the InTeGrate project changed the way participating instructors thought about their teaching. Another powerful facet of the InTeGrate project was collaboration situated within the geoscience discipline and was exemplified in the following quote from one of the participants:

> "It [InTeGrate] has helped because it is not just someone in an unrelated field talking about what they do, but it is colleagues in the geosciences teaching the same courses,

> using real tangible examples in different ways to create assignments and activities that you can use in your own classroom".

This model of professional development may be a more effective strategy for promoting the long-term adoption of reformed teaching strategies that promote student learning. We suggest that the project was successful in promoting instructional reform because it addressed the key domains (Personal, External, Practice, Consequence) of an interconnected model of professional development. Creating well-supported, long-term professional development activities such as the InTeGrate project is beyond the capabilities or the needs of many organizations interested in faculty professional development. However, several of the characteristics of the project such as the authentic application of instructional strategies, collaboration, the application of a well-designed rubric, and structured feedback can be designed and implemented at a variety of scales. This research thus provides insight for the development of future professional development opportunities seeking to better prepare instructors to implement reformed instructional strategies in their classrooms.

CHAPTER 2: Do our introductory geology courses help students think beyond a class about rocks? Realities of students' attitudes toward the nature of science and its relevance after taking a physical geology course.

Abstract

Introductory geoscience courses play a crucial role in recruiting new geoscience majors but we know relatively little about how students' attitudes and motivations are impacted by their experiences in geoscience classes. Students' attitudes toward science and its relevance are complex and are dependent upon the context in which they encounter science. For example, students often have positive to neutral attitudes toward the nature of science but frequently have negative perceptions about the science they confront in college classes which they often regard as not relevant to their lives. Recent investigations into the attitudes of geoscience students have provided evidence to support this observation. We sought to expand this data set to provide a broader characterization of students' attitudes. We examined students' attitudes about the nature of science and its relevance before and after taking an introductory geology course. To characterize students' attitudes, we employed two quantitative instruments: the revised Scientific Attitude Inventory and the Changes in Attitudes about the Relevance of Science survey. Results show a negative trend in students' attitudes about the nature of science while their attitudes about the relevance of science were more variable. Our data support the findings of previous studies showing only minimal change in students' attitudes about the nature of science and its relevance after taking an introductory science course. The data also highlighted several misconceptions about the

nature of science that could have implications toward future investigations of how geoscience courses impact student attitudes about science.

Introduction

Over 18 million undergraduate students are enrolled at post-secondary institutions (Snyder & Dillow, 2012). Martinez and Baker (2006) estimated that approximately 400,000 students were enrolled in an introductory geoscience course in 2004-2005. Four years later, a little more than 1% of that number of students graduated with an undergraduate geoscience degree (National Science Board, 2014). For the remaining 396,000 students, an introductory course may be the only formal experience with the geosciences and may therefore play a pivotal role both in recruiting a diverse technical workforce and in developing a scientifically literate citizenry capable of assessing environmental challenges. There are a variety of socio-scientific challenges facing humanity in the 21st century that deal with changes to our environmental system (e.g., sea-level rise) and the sustainable use of our natural resources (e.g., water availability). Addressing these challenges will require a robust personal and collective understanding of earth systems. Unfortunately, there is a paucity of such knowledge in the general population (Jon D Miller, 2004; S. R. Singer, Nielsen, & Schweingruber, 2012a).

In their overview of discipline-based education research, Singer et al. (2012) suggested that research should measure outcomes that extended beyond student performance to include aspects of instruction that promote student engagement in science and engineering. The degree of relevance students perceive from geoscience course content plays an important role in promoting student engagement (Osborne, Simon, & Collins, 2003), and can influence the diversity of students attracted to the geosciences (Huntoon, Peach, & Hopkins, 2005). We sought to examine the impact of an introductory geoscience course on undergraduates' attitudes about science and their perceptions of the relevance of geoscience.

Conceptual Framework

There is a division between students' attitudes toward science in general and their attitude toward what they perceive as school science (Osborne et al., 2003). Surveys show that students' attitudes towards science itself are often neutral or slightly positive (Nadelson & Viskupic, 2010; Osborne et al., 2003). A large-scale market research survey of 1552 students conducted in the United Kingdom for the Institute of Electrical Engineers (The Research Business, 1994) found that high-school aged students saw science as useful (68%) and interesting (58%), and that there was no significant distinction between genders. Additionally, a large proportion of students saw the relevance of science as a reason for studying it (53%) and the majority (87%) of students rated science and technology as 'important' or 'very important' in everyday life.

In contrast, Bezzi (1999) used qualitative methods to investigate the attitudes of first-year university students taking geoscience courses and showed that they had little concern for the relevance of geoscience in society, and this view remained intact after completing the course. Osborne & Collins (2000) noted that high school students struggled to identify with the relevance of the physical sciences. For example, their study found that chemistry students had a high level of disdain regarding the periodic table. Further investigation revealed that not only did they display a low-level of comprehension on the topic but they also failed to perceive its relevance to everyday life; either at present or in the future (Osborne et al., 2003). Ebenezer & Zoller (1993) investigated the attitudes of 1564 undergraduate students and found that, while they perceived science to be valuable (72%), and they thought that science in schools was important (73%), nearly 40% indicated that they found science classes boring, and irrelevant. This contrast between some students' views of science and science classes was also noted by Sundberg, Dini, & Li (1994) who compared the attitudes of majors and non-majors among a population of 2257 college biology students. They concluded that introductory classes were effectively "turning off" potential majors by unintentionally emphasizing science as disconnected from society (Sundberg et al., 1994). This disconnect may be due to a gap between the socially relevant idea of science as a field that many students are routinely exposed to, and a decontextualized version of science presented in academic settings (Ebenezer & Zoller, 1993; Sundberg et al., 1994; Zeidler, Osborne, Erduran, Simon, & Monk, 2003). These conclusions were colorfully summarized in Osborne et al. (2003, p.15): "In essence, the vision that school science offers is a backward-looking view of the well-established scientific landscape, whereas, in contrast, what appeals to and excites students is the 'white heat' of the technological future offered by science." Osborne (2003) suggested that science should be "less retrospective and more prospective". This view was echoed by Bralower, Feiss, and Manduca (2008) who suggested that there was a need to revise the way we teach about Earth, to shift from a focus on the past towards a focus on the future with an emphasis on aspects of earth systems science that affect society.

Students' attitudes toward science are complex and distinct from their opinions about a specific class or content topic (Osborne et al., 2003). Student's attitudes are affected by many factors, some related to what happens in the classroom and others linked to external forces. For example, student attitudes may be influenced by gender, as women's attitudes toward science are less positive than men's (Breakwell & Beardsell, 1992; Erickson & Erickson, 1984; Harding, 1983; Harvey & Edwards, 1980; Hendley, Stables, Parkinson, & Tanner, 1996; Johnson, 1987; Jovanovic & King, 1998; Kahle & Lakes, 1983; Robertson, 1987; Smail & Kelly, 1984). Environmental factors such as socio-economic standing can also influence student's attitudes (Breakwell & Beardsell, 1992); for example, students from wealthier families had more positive attitudes. Parental support and parental attitude toward science (Simpson & Oliver, 1985) have also been connected to students' attitudes toward science. For example, students whose parents support an interest in science tend to have better attitudes. Additionally, factors such as the classroom environment (Haladyna, Olsen, & Shaughnessy, 1982; Myers & Fouts, 1992; Talton & Simpson, 1985) and the teaching strategies used by instructors (McMillan & May, 1979; M. D. Piburn & Baker, 1993) may also have an impact on students' attitudes towards science.

Some have noted that attitudes toward science influence students' academic performance (Papanastasiou & Zembylas, 2004; Tuan, Chin, & Shieh, 2005) as well as how they utilize scientific information to make decisions related to socio-scientific problems (Sadler, 2004; Zeidler et al., 2003). Further, attitudes can influence decisions to take future courses in a discipline or in the selection of an academic major. Instructors have an opportunity to address some of the personal, environmental and academic factors that influence student attitudes toward science in the design and management of their courses.

Recent investigations into the attitudes (separate from the affective domain) of undergraduates taking introductory geoscience courses are limited. A study by Nadelson & Viskupic (2010) used the revised Science Attitude Inventory (SAI II; Moore & Foy, 1997) to measure the intellectual and emotional perceptions of the nature of science among undergraduates taking lower and upper division geoscience courses. They found that both groups of students' intellectual perceptions of the nature of science (e.g. understanding of the scientific method) showed no significant change over the semester while their emotional perceptions (e.g. the value of scientific information) showed a statistically significant decrease (Nadelson & Viskupic, 2010). These results reiterate the earlier findings of Bezzi (1999) which suggested that introductory geoscience courses were doing little to reduce the negative attitudes toward science among undergraduates. This decrease in the emotional perceptions of science could have been the result of students perceiving the content in geoscience courses as not relevant to their lives.

Studies have shown that students' attitudes about the nature of science are unaffected after taking a geoscience course (Bezzi, 1999; Nadelson & Viskupic, 2010), but what about the attitudes towards the relevance of science? Research suggests that enhancing relevance of a course can help to eliminate the disparity between how students perceive science in society and science in an academic setting (Ebenezer & Zoller, 1993; Osborne et al., 2003; Sundberg et al., 1994). We sought to determine how students' attitudes towards the nature of science and their attitudes about the relevance of science changed over the course of a semester in a large lecture format introductory physical geology course presenting a relatively standard curriculum. We investigated the following research questions:

1. How are students' attitudes about the nature of science and the relevance of science impacted by completing a physical geology class?
2. How do attitudes about the nature of science and the relevance of science vary among students on the basis of degree type, gender, and class year?

Study Population

This study took place over two consecutive semesters (Spring and Fall) and involved 168 students enrolled in two moderately large physical geology classes at North Carolina State University (NCSU). Each participant was over the age of 18 and provided their written consent to be included in this study. Our sample primarily consisted of freshman and sophomores most of whom were male and most students were pursuing a degree that was not in a STEM (science, technology, engineering, and mathematics) discipline (Table 2.1). The distributions of students among the four academic levels (freshman, sophomore, junior, senior) were similar for both semesters.

Table 2.1: Study population.

	Spring Semester (n=79)	Fall Semester (n=87)
Men	58	53
Women	21	34
Freshman	31	37
Sophomore	19	26
Junior	19	16
Senior	10	8
STEM	20	9
Non-STEM	59	78

Course Characteristics

Content coverage in the physical geology course at NCSU is similar to that taught in many introductory geology courses. The course is sub-divided into eight modules (Table 2.2) that incorporate a selection of topics presented in almost every Physical Geology textbook. Physical geology classes at NCSU have an active learning format where students frequently participate in think-pair-share activities, respond to conceptests (conceptual multiple choice questions) using clickers, and collaborate in small groups to make predictions or discuss interpretations of data. The course includes four exams, each testing students on material from two content modules. In addition, students completed online 'learning journals' prior to each class meeting and answered questions on an online quiz for each module. The learning journals consist of a series of multiple choice and open-ended questions answered on the basis of reading assignments or short videos that are linked to upcoming lecture topics. Student grades were calculated on the basis of exams (60%), in-class clicker questions (conceptests, 10%), online quizzes (10%) and learning journals (20%). Classes from both semesters lasted 75 minutes and met twice per week. The room layout was auditorium style with non-movable chairs arranged in rows facing the front of the class.

Table 2.2: Course content structure

Module Number	Topic
1	Introduction and the Scientific Method
2	Plate Tectonics
3	Rocks and Minerals
4	Earthquakes and Volcanoes
5	Geologic Time
6	Earth's Past Climate
7	Water and Society
8	Energy Resources and Future Climate Trends.

Instructional Reliability

The principal investigator of this study taught the class during both semesters. Using the same instructor minimized variability in instructional practice and provided an enhanced level of internal consistency and confidence. Controlling for instructional practice is especially important when measuring student's attitudes because research has shown that both the instructor and instructional practices used have a significant influence on student's attitudes about a course (Osborne et al., 2003). A formally trained, independent observer employed the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002) to score three non-consecutive classes per semester in order to provide a quantitative gauge on the consistency of instruction. The RTOP instrument is designed to be used to identify key dimensions of active learning, including the design and implementation of the lesson, the content (propositional) knowledge of the instructor, the nature of the tasks the students were asked to complete (procedural knowledge), communication among students, and the relationship between the instructor and students (Piburn and Sawada., 2002). Each of the RTOP's 25 items is scored from 0 (never occurred) to 4 (often occurred). The total scores, as well the sub-category totals, for all six observations are shown in Table 2.3. Taken collectively these scores suggest that instruction over the course of the semester and between semesters was consistent and was representative of active learning or student-centered instruction (Budd et al., 2013).

Table 2.3: RTOP scores for spring and fall semesters.

	Spring			**Fall**		
Observation	**1**	**2**	**3**	**1**	**2**	**3**
Lesson Design & Implementation	8	9	11	9	10	11
Propositional Knowledge	19	19	15	18	18	18
Procedural Knowledge	7	8	4	6	8	8
Communicative Interaction	13	9	11	10	11	13
Student/Teacher Relationships	7	8	9	8	10	7
Total	**53**	**53**	**50**	**51**	**57**	**57**
Semester Average	**52.0**			**55.0**		

Methods

We employed a quantitative pre- and post-comparison to characterize how students' attitudes towards the nature of science and the relevance of science change over the course of a semester. The Scientific Attitude Inventory (SAI) II was used as our metric for students' attitude toward science in combination with the Changes in Attitude About the Relevance of Science (CARS) survey which we used to measure students' perception of relevance of science and how it changed over the course of the semester. The Scientific Attitude Inventory (SAI) was first developed by Moore and Suntman (1970) and was subsequently the subject of a thorough re-evaluation by Munby (1983) which highlighted several issues with the surveys validity. Munby's review led to the revision of the SAI by Moore & Foy (1997) resulting in an improved instrument, the SAI II which measures a respondent's intellectual and emotional perceptions of the nature of science. This is a forty item survey with each item featuring a five-point Likert scale where respondents choose an option between strongly agree and strongly disagree (Appendix 3). The validity of the SAI II has also been questioned by Lichtenstein et al. (2008) for application to adolescents. However, Nadelson and Viskupic (2010) found the instrument valid and reliable for use with college students which they attributed to the greater maturity level of their study population. We administered the SAI II on the first day of class to all participants and then again during either the 13^{th} or 14^{th} weeks of the semester.

The CARS survey was developed by Siegel & Ranney (2003) and features a series of statements and a five-point Likert scale. For example, students can choose to strongly disagree (0), disagree (1), have a neutral opinion (2), agree (3), or strongly agree (4) with the statement “Science helps me to make sensible decisions”. The CARS survey was designed to gain insight in to how students’ attitudes about the relevance of science changes over time and to minimize the impacts of repeated testing bias (Siegel & Ranney, 2003). This was accomplished by dividing 59 items among three different versions of the CARS survey, each version consisting twenty-five items (seventeen version-unique items and eight repeated items; Appendix 4, 5 & 6). Each version of the survey contains a few items that are reverse-coded (e.g., “Science has nothing to do with my life outside of school.”), that is agreement with the statement indicates that the student views science as having low relevance. The first of the three versions were administered on the first day of class, the second version during the middle of the semester, and third version was administered at the end of the semester along with the post-class iteration of the SAI II. Some of the items from the original CARS survey have been slightly re-worded to make them more geology-specific. For example, item 17 on the first version of the CARS survey was initially worded as follows: “Science class will help prepare me for a career.” The word ‘science’ was replaced with ‘geology’ and now reads as “Geology class will help me prepare for a career.” Only minor wording changes were made to items in order to maintain the instruments original validity and reliability.

Data Analysis

We chose not to combine students from spring and fall semesters into a single population as there were numerous statistical differences within the data between semesters and we wanted to provide a more comprehensive examination of students’ attitudes than would have been possible by combining the data. When these data are stratified by academic rank the resulting numbers of juniors and seniors are significantly smaller than freshman and sophomores. Repeated-measures ANOVA tests are used to analyze the CARS data and consequently the small sample sizes of juniors and seniors make them more susceptible to violate the assumption of sphericity and result in type I or II errors. To minimize this risk, we combine juniors and seniors into a single group labeled upper class (UC) when analyzing stratified data from the CARS survey. The SAI II and CARS surveys require participants to respond to Likert items that generate ordinal data. This form of data is commonly used in the social sciences to create overall scores by assigning the ordered responses a numerical value (Boone et al., 2011) and these scores are used for various parametric statistical tests. However, ordinal data often do not meet the fundamental assumptions necessary to apply parametric statistical tests (Bond & Fox, 2007). Ordinal data are not linear, equal interval, and often are not normally distributed (Bond & Fox, 2007; Boone et al., 2011). Consequently, we did not use raw scores from the surveys that could have been obtained from assigning numeric values to the Likert scale items. Instead, we report SAI II and CARS survey scores as scaled values, referred to from this point as *person measures* that were calculated using Rasch modeling and analysis.

Rasch analysis was developed by George Rasch (1960) and represents a one-way probabilistic approach based on Item Response Theory (IRT). The application of Rasch measurement in the social sciences has been most notably discussed in the 1967 Invitational ETS conference and in a wide-variety of subsequent publications (Andrich, 1978; Boone et al., 2011; Choppin, 1985; Libarkin & Anderson, 2005; Linacre, 1998, 2006, 2010; Miyake et al., 2010; Tong, 2012; Mark Wilson & Adams, 1995; Wright & Stone, 1979; Wright, 1977, 1984). Rasch modeling is both norm-referenced (comparing individuals to the group) and criterion-referenced (measured according to specific standards; Siegel & Ranney, 2003). This is accomplished by calculating the difficulty of the item on a continuum, the participant's response to that item (ability), and the probability that a participant will choose a response (Boone et al., 2014). Winsteps software (Linacre & Wright, 2000) was used to compute person measures. Person measures are expressed in logits (log-odds units), which are equal-interval units that can be applied to parametric statistical tests because they avoid many of the issues of the non-linearity of rating scales as well as the non-linearity of raw survey data (Boone et al., 2014).

Logit values for person measures and survey items were used to construct Wright Maps, which are also known as person-item maps (Wilson. & Draney, 2002). Wright Maps are a visual method of displaying complex rating scale and test data (Boone et al., 2014) and place items and person measures on the same linear scale, much like a ruler. Wright maps simultaneously display both items and person measures because Winstep Rasch analysis transforms both data types into logits (Boone et al., 2014; Linacre & Wright, 2000). Wright maps allow the researcher to apply qualitative labels to quantitative values calculated in Rasch analysis because scores are shown on the same linear scale (Boone et al., 2014; Siegel & Ranney, 2003). We identified some of the highest- and lowest-rated items and those earning mean scores to provide qualitative insight into the CARS responses (Siegel & Ranney, 2003). Appendix 7 displays a simplified Wright map and shows how to identify the extreme items used to make those inferences. In order to preserve the distinction between the two instruments the logit values on the SAI II were scaled to a 0-100 scoring system while the CARS survey results are shown with their unmodified logit values.

Results

Semester

Results of for the SAI II and the CARS survey for the spring and fall semesters are plotted in Figures 2.1 and 2.2 respectively. Repeated-measure ANOVA tests were used to determine any significant changes over time, while the means of pre- and post-class SAI II surveys were compared using paired t-tests. SAI II results show statistically significant decreases for the whole class in both the spring ($t(62)=3.04$, $p=0.004$, $d=0.34$) and fall ($t(76)=5.73$, $p=0.000$, $d=0.56$) semesters (Figure 2.1). Students in the spring semester showed no significant change in their CARS score over time while fall semester students had a statistically significant improvement in their attitudes about the relevance of science over time ($F(2,128) = 7.91$, $p = 0.001$, $\eta_p^2 = 0.11$; Figure 2.2). Pearson correlations among all factors for spring and fall semesters are shown in Table 2.4. Students' attitudes about the nature of science and the relevance of science had significant ($p<0.05$) positive correlations at the beginning (pre-SAI II, CARS A) and end of the course (post-SAI II, CARS C) for both spring and fall semesters. Fall and spring semester students showed significant positive correlations among final grade, exam total, Post-SAI II, and CARS C.

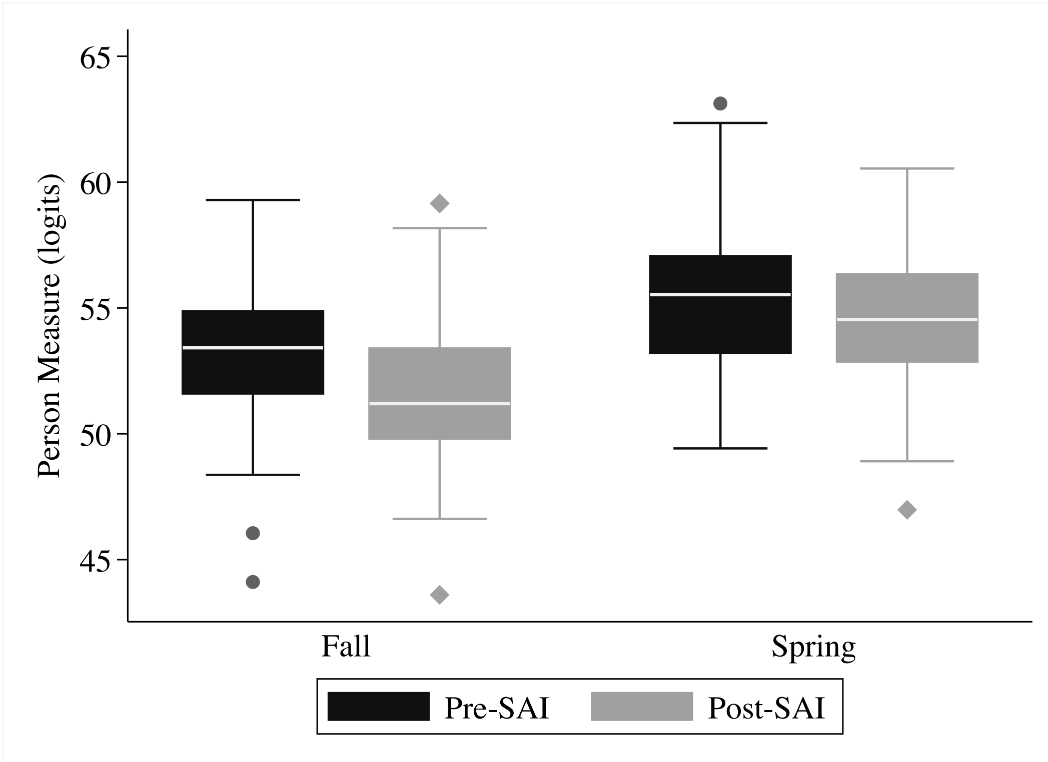


Figure 2.1: SAI II scores by semester.

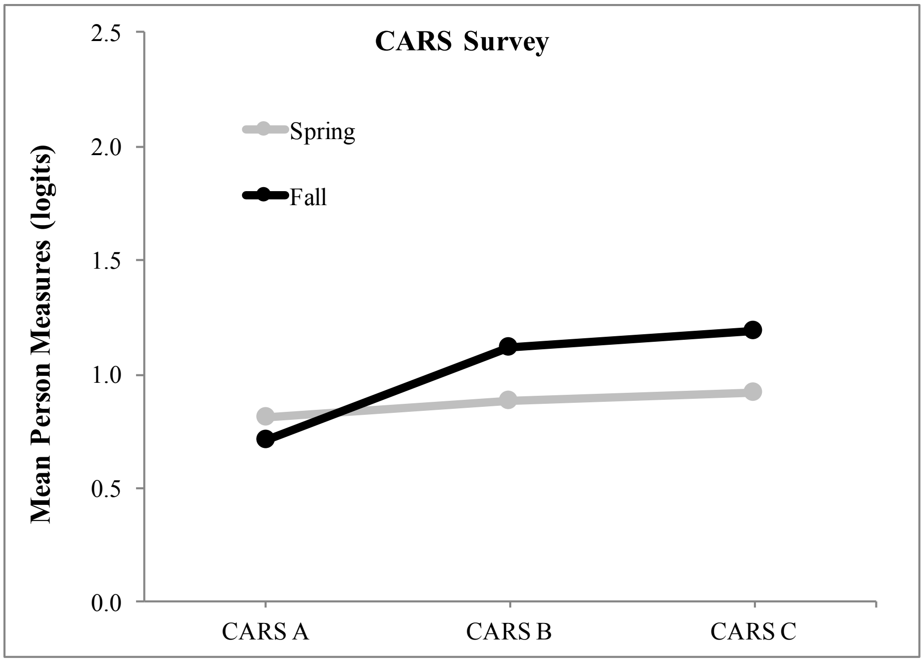


Figure 2.2: CARS survey scores by semester.

Table 2.4: Pearson correlation matrix

	1	2	3	4	5	6	7	
Spring	**0.32***	0.10	**0.28***	0.06	0.23	0.20		7. Exam Total
	0.03	0.64	0.12	-0.03	0.23			6. Final Grad
	0.48*	**0.55****	**0.56****	**0.65****				5. CARS C
	0.37*	**0.37****	**0.40****					4. CARS B
	0.57*	**0.52****						3. CARS A
	0.60*							2. Post-SAI II
								1. Pre-SAI II
Fall	0.11	**0.43****	**0.28****	0.17	**0.47****	**0.86****		7. Exam Total
	0.77	**0.32****	0.19	0.10	**0.34****			6. Final Grad
	0.57*	**0.72****	**0.67****	**0.75****				5. CARS C
	0.53*	**0.60****	**0.60****					4. CARS B
	0.67*	**0.54****						3. CARS A
	0.63*							2. Post-SAI II
								1. Pre-SAI II

Statistically significant values are highlight in bold.
$p < 0.05$ *
$p < 0.001$ **

Academic Rank

Upper class students (combined juniors and seniors) showed the most consistent improvement (while not always statistically significant) in their attitudes about the relevance of science (Figure 2.3) while change among sophomores and freshman over the course of the semester was inconsistent. There were no statistically significant changes in CARS scores among students of any academic rank during spring semester. However, upper class students during the fall semester did show a statistically significant improvement in their attitudes about the relevance of science over time ($F(2,36) = 4.24$, $p = 0.02$, $\eta_p^2 = 0.19$) but sophomores and freshmen showed no significant changes.

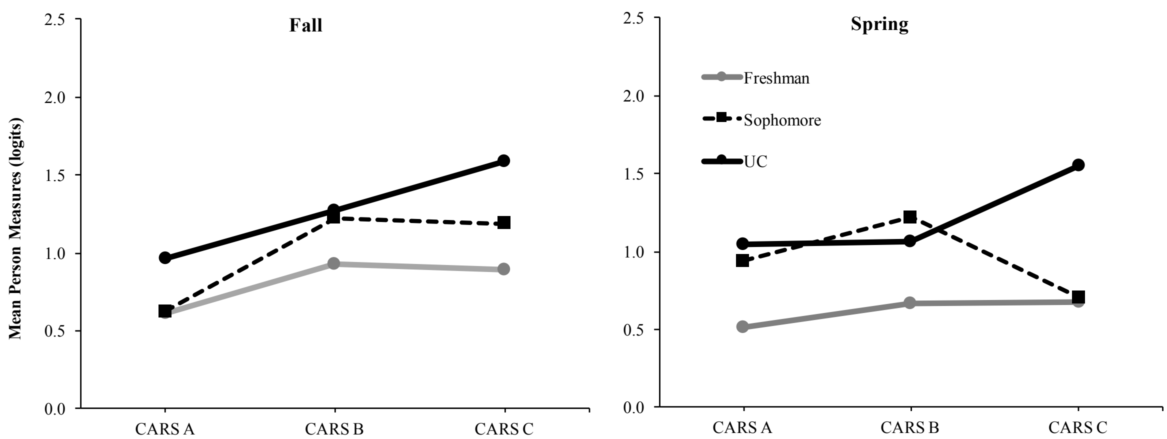


Figure 2.3: CARS survey scores stratified by academic rank.

While there was a negative overall trend in SAI II scores for each class (Figure 2.4), results among the four academic ranks show that the spring semester had higher scores than fall, and freshman recorded the lowest scores (Figure 2.4). In the spring semester sophomores showed a significant decrease ($t(13)=3.367$, $p=0.005$, $d=0.62$) in SAI II scores between pre- and post-surveys. Freshman ($t(35)=4.118$, $p=0.000$, $d=0.62$) and sophomores ($t(23)=3.868$, $p=0.001$, $d=0.55$) from the fall semester both showed significant decreases in their SAI II scores. Neither juniors nor seniors from spring or fall showed any significant changes.

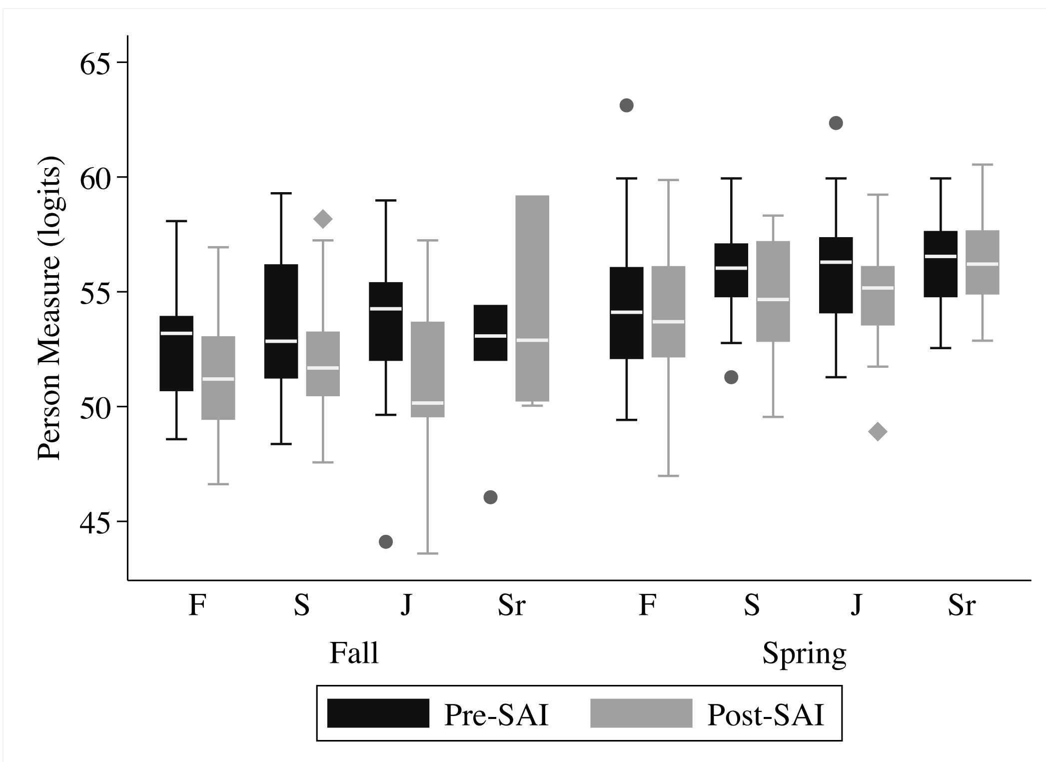


Figure 2.4: SAI II scores stratified by academic rank.

Gender

Male students consistently perceived science to be more relevant than female students in both the spring and fall semesters and both groups showed gains in scores over the semester (Figure 2.5). In the spring semester neither group recorded a statistically significant change in their attitudes toward the relevance of science. In the fall semester male and female students began the class with similar person measures but male students showed a statistically significant ($F(2,80) = 11.07$, $p = 0.000$, $\eta_p^2 = 0.22$) gain while the improvement in female students' scores was not statistically significant.

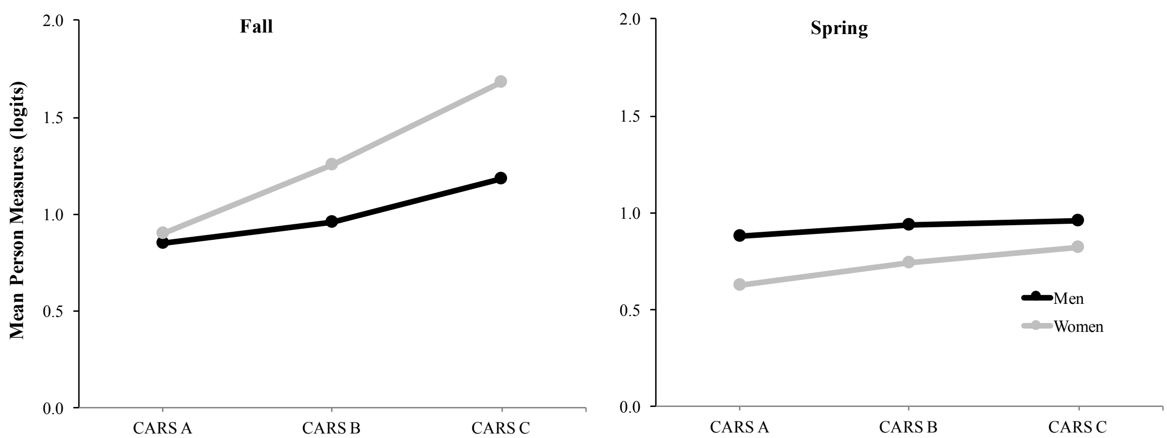


Figure 2.5: CARS survey scores stratified by gender.

SAI II scores from both semesters show a decrease for both men and women over the course of the semester (Figure 2.6). In the spring semester men showed a statistically significant decrease ($t(45)=3.01$, $p=0.004$, $d=0.41$) in their SAI II scores while women's scores showed no statistically significant change. In the fall semester both men ($t(43)=3.14$, $p=0.003$, $d=0.40$) and women ($t(32)=5.28$, $p=0.000$, $d=0.82$) showed significant decreases in their SAI II scores.

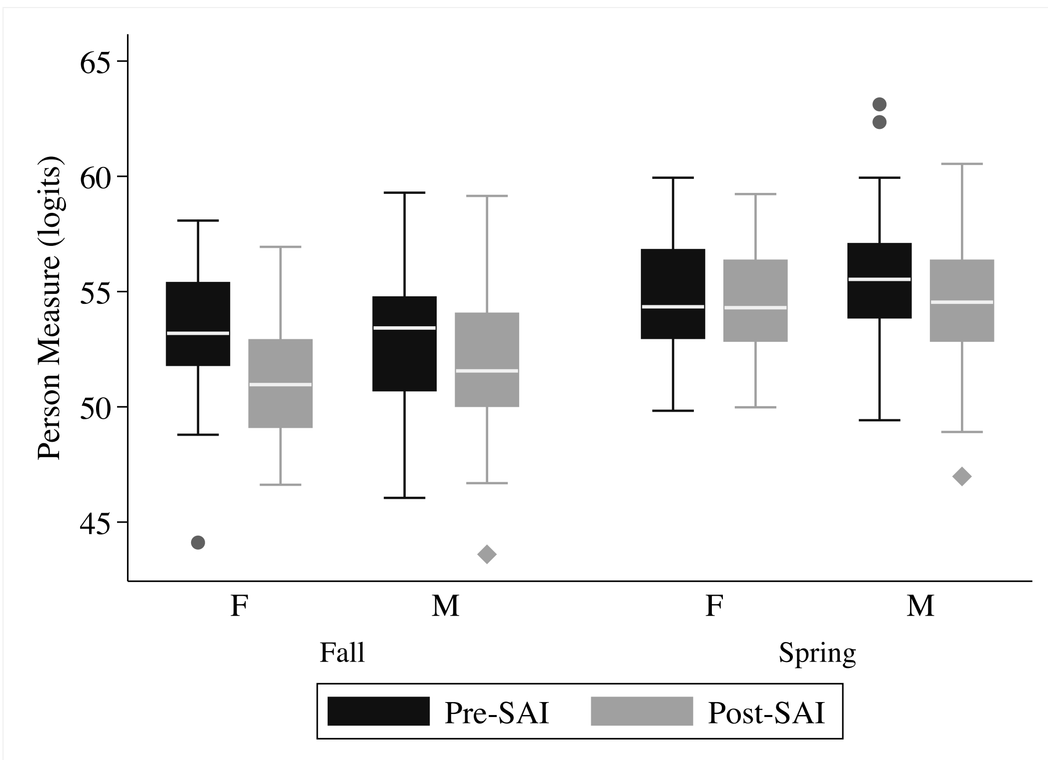


Figure 2.6: SAI II scores stratified by gender.

Degree Type

STEM majors consistently found science more relevant than non-STEM majors (Figure 2.7). In the spring semester STEM and non-STEM majors do not show any statistically significant change in attitude about the relevance of science over time. During the fall semester non-STEM majors show a significant improvement ((F(2,) = 4.97, p = 0.004, $\eta_p^2 = 0.09$) in their attitude about the relevance of science over time while STEM majors show no significant change over time.

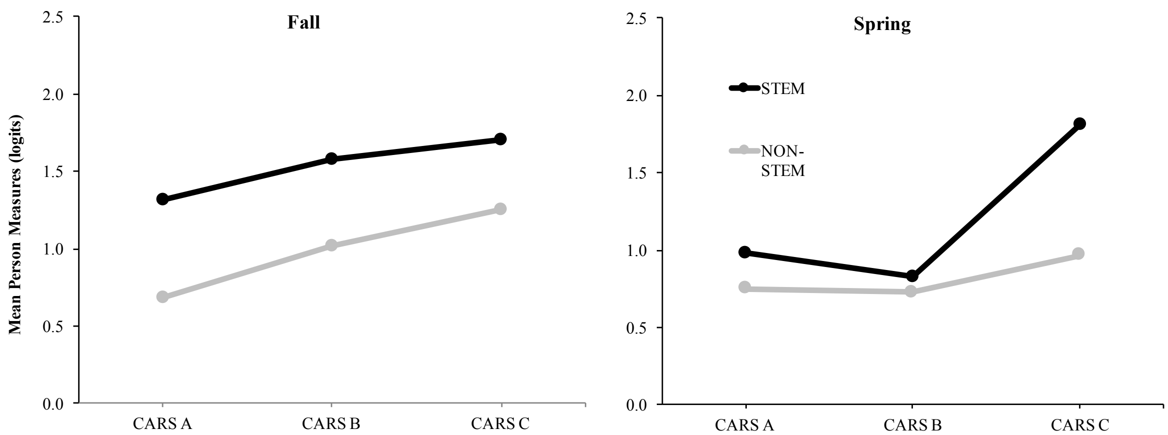


Figure 2.7: CARS survey scores stratified by degree type.

Non-STEM majors showed a significant decrease in their SAI II scores for both spring (t(45)=2.67, p=0.010, d=0.37) and fall (t(71)=5.267, p=0.001, d=0.52) semesters (Figure 2.8). In contrast, STEM majors showed no significant changes between pre- and post-class surveys during either semester (Figure 2.8).

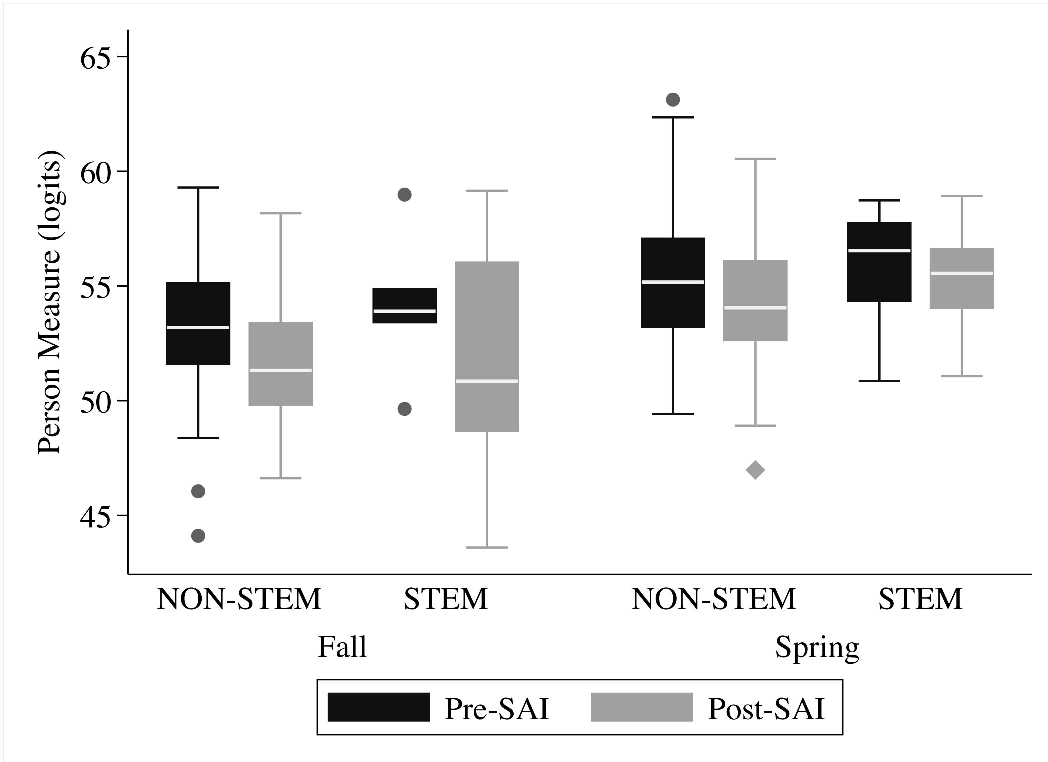


Figure 2.8: SAI II scores stratified by degree type.

Wright Maps

Wright maps were constructed for the three CARS surveys and the both iterations of the SAI II (Appendices 8 and 9 respectively) for both the spring and fall semesters. Items with a period placed between the letter "Q" and the item number were reverse scored. The following discussion will present items that plotted at the top (most difficult item/least agreement), bottom (easiest item/most agreement) and/or mean item (item representing the sample average if applicable). Items are discussed in terms of agreement because the linear scale on both the CARS survey and SAI II asked students to respond based on a categories of agreement. Wright maps provide insight into how the group as a whole responded to a series of items, however, it is also important to understand the differences between high, average, and low performing students' attitudes with in the CARS and SAI II surveys. To show this we identified representative responses in order to provide a more robust picture on what it means to go up and down the logit scale for the CARS and SAI II surveys. A table of person measures, interpretations of the attitudes, and items aligned with that attitude are shown in Table 2.5.

Table 2.5: Interpretations of the attitudes of high scoring, average, and low scoring participants for the CARS and SAI II surveys.

	CARS Survey		SAI II Survey		
Person Measure	**Attitude**	**Representative Items**	**Person Measure**	**Attitude**	**Representative Items**
>2.0	Feels that science is relevant to their personal lives	CARS A Q13 CARS B Q19 CARS C Q22	>60	Feel that science is more about discovery than creating things for people.	Q31 Q9 Q24
0.5	Sees the value of science in academics and in voting	CARS A Q17 CARS B Q18 CARS C Q7	54	Science is challenging and may or may not be an interesting career. Unsure if being a scientist requires creativity	Q14 Q20 Q13
<-2.0	Feel that science has little impact on their personal lives	CARS A Q13 CARS B Q19 CARS C Q22	<46	Feel that science is only useful when it creates things that can help people	Q31 Q9 Q24

CARS Survey

The order and characteristics of items in each survey changes, i.e. statement Q1 in CARS A is not the same as Q1 in CARS B or CARS C. During the first iteration of the survey (CARS A) spring semester students almost completely agreed that science helps them understand the environment (Q5) but also felt that science had nothing to do with buying things such as cars and computers (Q13). In the fall semester students entered the course (CARS A) collectively agreed that science can help them understand about the environment (Q5) and most also agreed that science involves local issues (Q1). By mid-semester (CARS B) spring semester students still mostly agreed that science can involve local issues, such as waste from a factory (Q1), however, students also mostly agreed that personal feelings were important for making scientific choices (Q19). By mid-semester (CARS B) for fall semester students maintained their agreement that science involves local issues (Q1; reverse scored) and that science involves the real world (Q8; reverse scored) but, on average students agreed that they only take STEM courses if they are required to do so (Q23; reverse scored). At the end of the spring semester (CARS C) a large majority of students agreed that science is tied to real life (Q5; reverse scored). Throughout the spring semester, on all iterations of the CARS survey, students had consistent attitudes about geology class. Items such as “geology class will help me prepare for a career (Q17, CARS A)”, “I only take geology because it is a required course (Q18, CARS B; reverse scored)”, and “Learning geology will have an effect on the way I vote in elections (Q7, CARS C)” correlated with the average person measures for the students in the spring semester. By the end of the fall semester (CARS C) students agreed that making decisions are difficult when they don’t understand the choices (Q11) and showed agreement that careers in science are valuable (Q5; reverse scored). Students also show high agreement that the scientific method cannot help them make decisions (Q22).

SAI II

Spring semester students had high agreement that not all scientists agreed with each other (Q5; Figure 2.11). Students agreed that anything we need to know can be found out through science (Q2; reverse scored) and students also agreed that a major purpose of science is to help people live better (Q31; reverse scored). At the end of the semester, spring students' levels of agreement stayed the same for Q5 and Q2, but students agreed that electronics are examples of really valuable products of science (Q31 reverse scored).

At the beginning of the fall semester students' had high agreement that not all scientists agree with each other (Q5). They also agreed that a major purpose of science is to help people live better (Q31; reverse scored), and on average students agreed that scientific laws have been proven beyond all possible doubt (Q35), and that they considered scientific work too be difficult (Q14). At the end of the semester students continued to believe that not all scientists agree with each other (Q5) and that a major purpose of science is to help people live better (Q31; reverse scored). However, the students also generally agreed that scientific work would be hard for them (Q14; reverse scored) in addition to the idea that scientific laws have been proven beyond all possible doubt (Q35; reverse scored).

Discussion

Students' attitudes about the nature of science decline as they take a physical geology course, while scores for their attitudes about the relevance of science rise. Participation in a physical geology course that discusses a relatively traditional curriculum in a classroom setting featuring active learning practices does not appear to impact students' attitudes about the nature of science. Alternatively, students' negative pre-conceived attitudes about the nature of science may be already be well-developed by the time of course enrollment and thus difficult to change. The decrease is SAI II score suggests that the typical content of a physical geology course could be unintentionally reinforcing negative attitudes. Wright map are interpreted to suggest that students' preconceptions about the nature of science may be difficult to change over the course of a single semester as the proportion of students agreeing with key survey items did not significantly change from pre-to post-class surveys. For example, students consistently agreed that helping make people lives better was the major purpose of science, that science must "serve" us to be useful (SAI II Q31). However, a more conventional view is that the major purpose of science is the discovery of knowledge. Despite the persistence of misconceptions, students from both semesters agree that scientific information does not necessarily represent a consensus view and that 'part' of the scientific process in not necessarily to generate a consensus (Q5; SAI II). Students from both semesters agreed with the Q11 on the CARS C by the end of the course suggesting that while students may not see the advantages of the scientific method in their own lives, they agreed with the importance of understanding the choices when making important science-related decisions.

Students' attitudes toward geoscience information and geoscience careers measure by the CARS surveys were consistently neutral at the beginning of both semesters and showed little change after taking an introductory geology course. This is a crucial area where our introductory geology courses should be making a more positive impact on students. This reinforces the link between students' attitudes and their interest in geology as a major, and this connection is supported by research in other STEM fields (McMillan & May, 1979; Osborne et al., 2003; M. D. Piburn & Baker, 1993). These data could also indicate that this particular introductory geology course is not confronting the contrasting perceptions students have about science in society and science in the classroom because they are not finding course content relevant outside of its more obvious connections with society.

When students are able to find material relevant in a curriculum or academic projects they are more likely to be open to new learning experiences and modify previously held beliefs and attitudes about geoscience (Huntoon & Lane, 2008). Students taking our introductory geology course may not see how geology could potentially fit into their lives because our data shows inconsistent change in their attitudes about the relevance of science upon completing the course. Also, our results show that students consistently agree with items stating that geoscience and science in general does not impact their personal lives. The diversity of students entering a discipline is linked to how relevant those students find course content (Baber, Pifer, Colbeck, & Furman, 2010). While gender diversity in the geosciences is approaching greater parity with the overall population; the proportion of underrepresented students still fall well below national benchmarks (Levine, Gonzalez, Cole, Fuhrman, & Floch, 2008), and a perceived lack of relevance could provide one explanation why the geosciences are not as diverse as other STEM disciplines (Huntoon & Lane, 2008).

Misconceptions about the nature of science and the importance of scientific information persist among students despite taking an introductory physical geology class (e.g. SAI II Q9 & Q31). This is a recurring theme found in other studies (Abd-El-Khalick, 2004; Abell & Smith, 1994; McComas, Almazroa, & Clough, 1998) and our data show some consistency about what misconceptions students display in their perceptions of the nature of science. Students felt that the most important purpose of science is the development of products, or to help people live better lives. The nature of science is a multifaceted concept

(Nadelson & Viskupic, 2010) but several common themes have emerged form research attempting to define its characteristics (Abd-El-Khalick, 2004; Abell & Smith, 1994; Bell, Blair, Crawford, & Lederman, 2003; McComas et al., 1998; Moore, Richard W., Suntman, 1970; Ryder, Leach, & Driver, 1997). Students' high frequency of agreement with certain items on the SAI II suggests that they misunderstand this concept to mean that science is *only* useful if it produces something that makes our lives better. This highly pragmatic view of science could also suggest that students perceive purely academic pursuits, often referred to as basic science, as not being a valuable or relevant. Much of the information in an introductory geology can be classified as basic science from the viewpoint of a non-scientist. This could explain the negative impact on students' attitudes about the nature of science, as well as a decrease in their emotional perceptions of the nature of science (Nadelson & Viskupic, 2010). Students in both semesters agreed that the scientific method does not contribute to making personal decisions. This can could be interpreted that students do not relate how the scientific process can be applied to their personal lives but it also highlights another misconception about the nature of science. The nature of science is defined by the idea that science is not a linear process and that there is not a 'single' scientific method (Nadelson & Viskupic, 2010). Students from both semesters consistently agreed with Q13 (CARS A) and Q22 (CARS C) suggesting that they perceived the scientific process as only applying to science or represented a process that could only be performed in a controlled environment such as a lab.

A potential reason for the lack of positive impact of a standard physical geology course on student attitudes about the nature of science could be the limited duration of a single course. Ryder et al. (1997) showed that students can show changes in the understanding of the nature of science after about 5-8 months of related experiences. If time is a major influence, then this places an even greater emphasis on introductory geoscience courses because they will need to be designed around much more effective strategies at making positive impacts on the majority of students' attitudes.

In contrast to the relative uniformity of students' attitudes about the nature of science, students' attitudes toward the relevance of science amongst class year and degree type show some differences. These data are interpreted to show that students who have had more experience with STEM courses find science more relevant than non-STEM majors. Students' attitudes about the relevance of science showed some variation among gender in this study while attitudes about the nature of science were consistently more negative over the course of the semester for both genders. Female students consistently (although not statistically) found science less relevant than their male peers. This parallels the results of previous studies (Breakwell & Beardsell, 1992; Erickson & Erickson, 1984; Harding, 1983; Harvey & Edwards, 1980; Hendley et al., 1996; Johnson, 1987; Jovanovic & King, 1998; Kahle & Lakes, 1983; Robertson, 1987; Smail & Kelly, 1984). However, while this variation was evident in the results from the CARS data it is not reflected in the SAI II.

Previous studies have connected attitudes with class performance (Papanastasiou & Zembylas, 2004; Tuan et al., 2005). Both the SAI II and CARS showed positive significant correlations at the end of the semester with final grades and exam totals (Table 2.4). The final grades and exam scores of women were similar to those of men in both semesters (Figure 2.9) despite finding science less relevant. This observation holds true even when looking at the stratified results among degree type and academic rank where there was variation in attitudes but no comparable variation in class performance. Other researchers have also observed no significant relationship between relevance and achievement (Frymier & Shulman, 1995; Newby, 1991; Warren, 2011). Factors such as student motivations to earn a certain grade may be providing a greater control on class performance. Lukes & McConnell (2014) investigated college student's motivation to study for exam and found that all the participants were motivated by mastery goals tied to course grades and exam scores. Because our study population was demographically similar to their study population, it is reasonable to assume that our participants, regardless of gender, degree type, and academic year are driven by the similar motivations. These data could have broad implications for future research investigating the connections between students' attitudes about the relevance of science and their academic performance.

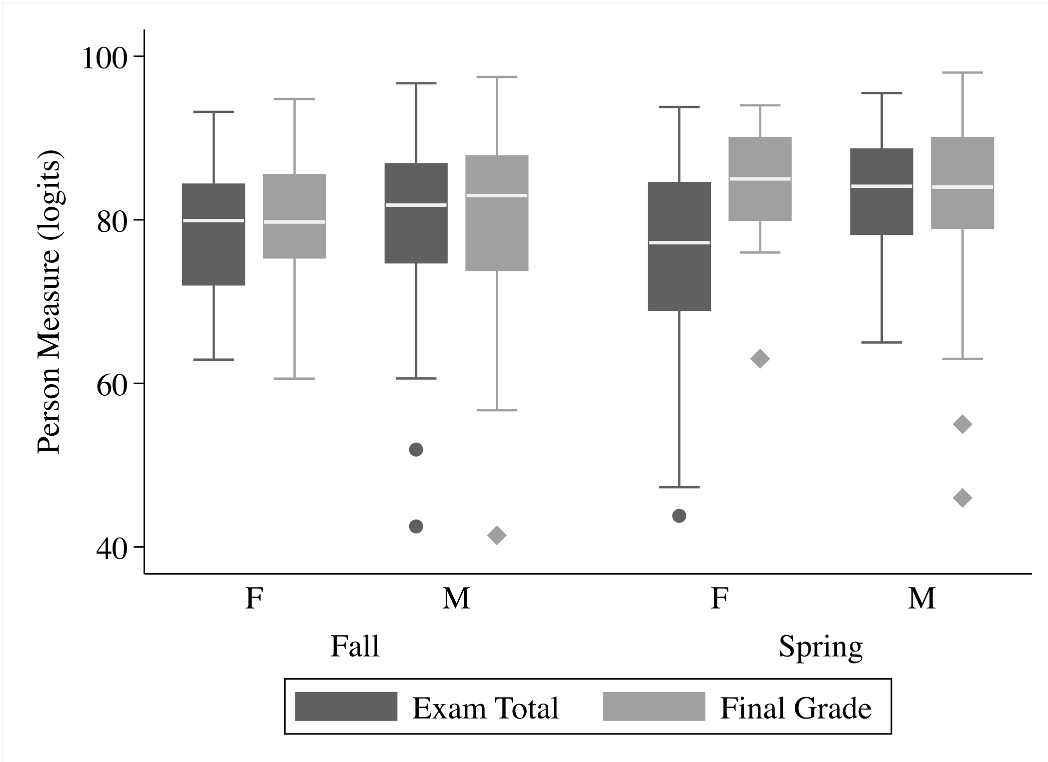


Figure 2.9: Class Performance stratified by gender: M=male, F=female. Exam total is the average of the four class exams.

Study Limitations

The data and interpretations presented in this study are limited by the quantitative nature of its design. The use of quantitative instruments was advantageous because their ease of implementation allowed us to examine and summarize the attitudes of many students. Qualitative methods, which admittedly would have provided a more in-depth analysis of students' attitudes, would not have allowed us to gain as broad a picture of students' attitudes as was necessary to be able to inform our research questions. Qualitative methods, such as student interviews, would have been more susceptible to volunteer bias because IRB restrictions would have required participants to volunteer for lengthy interviews multiple times per semester that also would have needed to be done outside of regular class time. Consequently, the results of this study should be viewed as an initial survey of attitudes about science of students taking a physical geology course.

Conclusion

Students' attitudes about the nature of science are negatively influenced by taking a physical geology course and their attitudes toward relevance show inconsistent changes. Data from two semesters of the same class show that students have several incorrect preconceptions about the nature of science that persist despite completing a physical geology course. STEM majors consistently found science more relevant than non-STEM majors, and male students generally found science more relevant than female students. Consequently, our data suggest that introductory geology classes may not be using the most effective strategies for reaching populations that are traditionally underrepresented in STEM. This study provides a baseline that can be used to further investigate the impacts of our introductory geoscience courses on students' attitudes about the nature of science and their perceptions about its relevance. Further research is necessary to identify effective strategies that can make positive influences on students' attitudes. Identifying instructional strategies that can make positive impacts on the attitudes of students may be necessary if we seek to attract a greater diversity of students to the geosciences.

CHAPTER 3: On the road toward science literacy: Determining the impact of socioscientific issues on the students' attitudes about the nature of science, relevance of science, and their descriptions of sustainability.

Abstract

There is consensus among industrialized nations that it is important for its citizens and leaders to be scientifically literate. Therefore, it is important for the educational system to provide students with pertinent scientific knowledge, an understanding of the scientific processes, and the ability to evaluate scientific claims. Students' attitudes toward science and its relevance are important aspects of science literacy. Some researchers have proposed that we can improve students' attitudes towards science and foster science literacy by emphasizing the connections between science and society. We sought to determine if the repeated and explicit exposure to socioscientific issues through the use of InTeGrate course materials would result in positive changes to students' attitudes about the nature and relevance of science. We collected data on student attitudes using the revised Scientific Attitude Inventory and the Changes in Attitude about the Relevance of Science survey in a quasi-experimental design over four semesters of an introductory physical geology course. Results show that the emphasis of socioscientific issues can influence both students' attitudes about the nature of science and their perceptions on the relevance of science. Changes were observed in data from both STEM and non-STEM majors. These findings have implications about how we select content for introductory science courses in general, and proves the utility of designing geoscience lessons based around socioscientific issues.

Introduction

Anthropogenic influences on global life support systems have reached a magnitude unprecedented in human history and currently approach levels that have the potential to jeopardize the well-being as a species (Jerneck et al., 2011; UNDP 2007, 2009; World Bank, 2007, 2009). Human beings are an integral part of the earth system and since the Industrial Revolution our society has developed to a point where we have the capability of significantly impacting our environment (Vitousek et al., 1997; Goldstein, 2002; Barnosky et al., 2014). A stable future will require the sustainable and ethical use of our natural resources and environmental systems (American Geophysical Union, 1994; National Science Foundation, 1996; National Research Council, 1997, 2001; National Science Board, 2003, 2003). The United States must build robust educational pathways for its citizenry to develop the global perspective, cultural sensitivity, economic wisdom, and scientific acumen to inform decisions regarding these environmental challenges. A geoscience literate population is necessary to make thoughtful decisions that will sustain human beings in an evolving earth system.

There is broad consensus among the United States, and other industrialized nations, that it is important for their citizens and leaders to be scientifically literate (J. D. Miller, 2004; Tewksbury, Manduca, Mogk, & Macdonald, 2013). The proportion of US adults who can be considered scientifically literate has steadily increased since the 1950s (Miller, 2004). However, this number remains low enough to be problematic for a democratic society that values citizens understanding major national policies, and requires their participation in the resolution of important environmental debates (Miller, 2004). The term “science literacy” is thought to have been first used in the late 1950s in an article entitled *Science Literacy: Its*

Meaning for American Schools (Hurd, 1958; see reference in DeBoer, 1991; Laugksch, 2000; Roberts, 1983). A national interest in science emerged during the late 1950s response to the launch of the Soviet Satellite Sputnik in 1957 (Laugksch, 2000). Since that time there have been multiple attempts to define and assess science literacy (DeBoer, 1991; Feinstein, 2011; Laugksch, 2000; J. D. Miller, 2004; Sadler, 2014; Shen, 2014). A key challenge in defining science literacy is the evolving and iterative nature of science itself and the fact that scientific information changes over time. Shen (1975) argued that science literacy should be viewed as a series of separate metrics and proposed different literacy standards for citizens with different roles (e.g., consumer, government representative, expert). Miller (2004) describes a science literate citizen as someone who can understand the science section of the *New York Times*. This definition of science literacy emphasizes a persons' ability to read and comprehend scientific information written at a certain educational level. However, many would view science literacy as a more complex construct that goes beyond understanding content. The Program for International Student Assessment (PISA 2006; Bybee & McCrae, 2011) outlined four interrelated characteristics that defined science literacy: (1) recognizing life situations involving science and technology, (2) understanding the natural world, (3) demonstrating competencies that include identifying scientific questions, and (4) responding with an interest in science, support for scientific inquiry and motivation to act responsibly towards natural resources and environments. The fourth item in the PISA 2006 definition of science literacy is clearly attitudinal in orientation (Bybee & McCrae, 2011).

This drive towards citizen science literacy has become a crucial and nebulous goal of science education. We live in a republican society that constitutes an environment in which science interacts with other societal forces (Nowtony, Scott, & Gibbons, 2001). In this view, it becomes important for the educational system to provide student with an understanding of processes of science, and the ability to evaluate scientific claims (Colucci-Gray, Camino, Barbiero, & Gray, 2006). The InTeGrate project (Interdisciplinary Teaching about Earth for a Sustainable Future) represents an effort to meet this requirement in introductory college geoscience courses. The InTeGrate project was designed and implemented by a team of geoscience researchers and is hosted by the Science Education and Research Center (SERC). InTeGrate seeks to promote geoscience literacy in the context of socioscientific issues through the development of innovative teaching materials featuring a systems thinking approach, grounded in the earth science literacy documents, and focused on highlighting the connections between geoscience and society. The process of developing materials for the InTeGrate project with teams of geoscience faculty is detailed in Chapter 1.

Dealing with socioscientific issues arising from the complex interactions of science and society represents an integral component of scientific literacy and citizenship education (Colucci-Gray et al., 2006). Sadler and Zeidler (2005) argued that students with a better understanding of the content tied to socioscientific issues make more rational and informed decisions. In the domain of teaching socioscientific issues the model that places greater emphasis on the cognitive aspects, and is based on the assumption that there is a strong connection between the knowledge of the topic and the solution of the problem will lead to greater learning gains (Sadler & Zeidler, 2005). The limited research on the impact of socioscientific issues suggests that an emphasis on the connections between content and society should help student gain greater insight into the use and impact of scientific information (Zo`bi, 2014).

Conceptual Framework

Research has shown that student's attitudes and certain aspects of science literacy are connected (Alsop & Watts, 2003; Koballa & Glynn, 2007; Zeidler et al., 2003). However, further research is needed to accurately characterize the connections between students' attitudes and their responses to science-related life situations (Bybee & McCrae, 2011). Personal attitudes play a significant role in an individual's interest, attention, and response to science and technology and therefore contribute to the development of student's science literacy (Bybee & McCrae, 2011). PISA 2006 primarily investigated the attitudes of younger adolescents (15-year olds from over 50 countries). Few studies have researched college students' attitudes toward science and even fewer have examined how attitudes are influenced by the introduction of a specific emphasis on socioscientific issues throughout a course.

Science literacy is connected to a student's attitude toward science, their perceptions of its relevance to their lives and their ability to evaluate and understand issues dealing with sustainability (Holbrook & Rannikmae, 2009). For instance, in their review of the PISA 2006 results, Bybee & McCrae (2011) found that students were most interested in science issues related to health and safety and they had the least interest in topics that they did not find relevant. This observation may seem obvious to educators, however, with regards to geoscience education many of those topics that students did not find relevant were tied to the physical sciences. Consequently, any teaching strategy seeking so improve and influence students level of science literacy in college classrooms needs to take into account students' attitudes toward the nature of science and its relevance. Van der Leeuw, Wiek, Harlow, and Buizer (2012) noted that future generations will face a myriad of sustainability challenges at a range of scales. These issues have given rise to the scientific cross-discipline of sustainability science. Sustainability science is an attempt to bridge the natural and social sciences intent on seeking creative solutions to complex problems such as, but not limited to; biodiversity loss, deforestation, climate change, land degradation, and water scarcity (Jerneck et al., 2011). Many of the issues related to sustainability science are deeply rooted within the geosciences (Jerneck et al., 2011; Kajikawa, 2008; van der Leeuw et al., 2012). Overlapping sustainability-related issues tied to the geosciences have recently gained greater prominence due in part to community sourced literacy documents (e.g. Wysession et al., 2012), published calls for discipline-wide action (e.g. Tewksbury, Manduca, Mogk, & Macdonald, 2013), and through community-wide efforts to improve geoscience literacy and sustainability education by developing new instructional materials (e.g. InTeGrate project).

Emphasizing the connections between science and society has the potential to enhances students understanding of key environmental issues such as sustainability. Further, it provides an opportunity to influence student attitudes about science that may help drive learning and recruit majors. However, research on the links between socioscientific issues, science literacy, and students' attitudes is lacking. Therefore, this study sought to determine if the repeated and explicit emphasis of socioscientific issues throughout an introductory geology class can positively influence students' attitudes toward the nature of science, their perceptions on the relevance of science, and their understanding of sustainability in the geosciences. We investigated the following research questions:

1. How are students' attitudes about the nature of science and its relevance influenced by the repeated emphasis of socioscientific issues in an introductory geology class?
2. Does an emphasis of socioscientific issues help students characterize sustainability in the geosciences?
3. Does an emphasis on socioscientific topics results in change in student learning of class content?

We hypothesize that the repeated emphasis of socioscientific issues during an introductory geology course will result in positive changes in students' attitudes toward the nature of science and the relevance of science. Additionally, we posit that this explicit emphasis on socioscientific issues will better prepare students to identify concepts dealing with sustainability and to help them gain a more sophisticated understanding of the relationship between geoscience and sustainability.

Study Participants

This study took place over four consecutive semesters and involved 310 students (Table 3.1) enrolled in a physical geology course at North Carolina State University (NCSU). Each participant was over the age of 18 and provided their written consent to be included in this study. Our sample primarily consisted of freshman and sophomores most of whom were male non-STEM (science, technology, engineering, and mathematics) majors. STEM majors were identified using information from the university registrar system that is available to instructors. A STEM major is defined as any student who had declared their area of study as one of the following at the time of enrollment: e.g. biology, chemistry, engineering, mathematics, physics, geology, or computer science. The distribution of students among the four academic levels were similar for all four semesters. The academic and demographic distribution of students in this study is typical of introductory physical geology courses at NCSU. Instruction in the physical geology classes involved an active learning format where students frequently participated in think-pair-share activities, responded to conceptests (conceptual multiple choice questions) using clickers, and collaborated in small groups to make predictions or discuss interpretations of data. The course is sub-divided into eight modules (see Table 2.2) that incorporates a selection of topics presented in almost every Physical Geology textbook. The course includes four exams, each testing students on material from two content modules. In addition, students completed online ‘learning journals’ prior to each class meeting and answered questions on an online quiz for each module. The learning journals consist of a series of multiple choice and open-ended questions answered on the basis of reading assignments or short videos that are linked to upcoming lecture topics. Student grades were calculated on the basis of exams (60%), in-class clicker questions (conceptests, 10%), online quizzes (10%) and learning journals (20%). Classes from all four semesters lasted 75 minutes and met twice per week. The room layout was auditorium style with non-movable chairs arranged in rows facing the front of the class.

Table 3.1: Study participants.

	Control Semester 1 (n=79)	Control Semester 2 (n=87)	Treatment Semester 1 (n=74)	Treatment Semester 2 (n=70)
Men	58	53	51	46
Women	21	34	23	24
Freshman	31	37	33	37
Sophomore	19	26	23	11
Junior	19	16	12	19
Senior	10	8	6	3
STEM	20	9	20	17
Non-STEM	59	78	59	54

Instructional Reliability

The principal investigator of this study taught class for all four semesters. Using the same instructor minimized variability in instructional practice and provided an enhanced level of internal consistency and confidence. Controlling for instructional practice is especially important when measuring student's attitudes because research has shown that certain teaching strategies as well as the instructor's personality and interactions with students have a significant influence on attitudes about a course (Osborne et al., 2003). Additionally, training an instructor (or several instructors) to use the InTeGrate teaching materials could have introduced a myriad of additional variables that would have been difficult to control for. A formally trained, independent observer employed the Reformed Teaching Observation Protocol (RTOP; Sawada et al., 2002) to score three non-consecutive classes per semester in order to provide a quantitative gauge on the consistency of instruction. The RTOP instrument is designed to be used to identify key dimensions of active learning, , including the design and implementation of the lesson, the content (propositional) knowledge of the instructor, the nature of the tasks the students were asked to complete (procedural knowledge), communication among students, and the relationship between the instructor and students (Sawada et al., 2002). InTeGrate materials used in this study are designed to feature a higher degree of active learning, therefore, observations made during the treatment semesters was conducted during classes where only standard materials were taught. This was again used to provide greater confidence that the style of instruction was consistent among all semesters and the ensure that the RTOP scores obtained were comparable amongst semesters. The total scores, as well the sub-category totals, for all six observations are shown in Table 3.2 and these scores did not significantly vary as a function of semester ($F(2,9)=2.294$, $p=0.157$, $\eta_p^2=0.34$).

Table 3.2: Breakdown of RTOP observation scores

Observation	**CS1**			**CS2**			**TS1**			**TS2**		
	1	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**	**1**	**2**	**3**
Lesson Design	8	9	11	9	10	11	8	11	9	9	10	11
Propositional Knowledge	18	19	15	18	18	18	18	19	18	18	17	16
Procedural Knowledge	7	8	4	6	8	8	6	5	7	7	6	6
Communicative Interaction	12	9	11	10	11	13	9	11	9	10	9	9
Student/Teacher Relationships	7	8	9	8	10	7	8	7	10	9	10	9
Total	53	53	50	51	57	57	49	53	53	53	52	51
Semester Mean	**52**			**55**			**52**			**52**		

Methods

We employed a quasi-experimental design (Anderson-Cook, 2005) consisting to two control semesters (CS1 and CS2) and two treatment semesters (TS1 and TS2). For the purposes of this study, a quasi-experiment is defined as a study seeking to determine causal explanation for an observation that lacks the random assignment of participants to experimental and control groups (Shadish, W.R., Cook, T.D., & Campbell, 2002). Treatments semesters had 60% of the standard course material for physical geology replaced by materials developed for the InTeGrate project (Table 3.3). InTeGrate materials are different when compared to standard course content because they are designed to explicitly connect geoscience content with its importance to society. InTeGrate materials are based off of the earth science literacy documents and designed around research-validated instructional strategies shown to improve student engagement and learning. Each module and its subsequent units are designed around measurable learning objectives, a variety of formative and summative assessments and student-centered activities. Additionally, these materials are designed to be modular. This means that InTeGrate materials can be broken down and implemented into a course at a variety of scales. For example, and instructor may choose an entire two-week module, a series of single class units, or a variety of individual units to integrate into their classes. We sought to determine if this change in curricular materials resulted in any differences among control and treatment groups in attitudes about the nature of science, its relevance, and their understanding of sustainability in the geosciences. This chapter employs the same attitude surveys and methods that were discussed in chapter 2 but with the addition of a short answer response designed to inform us about the level of understanding students have about sustainability in the geosciences.

Table 3.3: Content modules for control and treatment semesters. Modules in gray during the treatment semester are InTeGrate modules

Module	**Control**	**Treatment**
1	Introduction/Scientific Method	Introduction/Scientific Method
2	Plate Tectonics	Plate Tectonics
3	Rocks and Minerals	Humans' Dependence on Earth Mineral Resources
4	Earthquakes and Volcanoes	Living on the Edge: Building Resilient Societies on Active Plate Margins
5	Geologic Time and Dating	Geologic Time and Dating
6	Earths Past Climate	Climate of Change
7	Water and Society	Environmental Justice and Freshwater Resources
8	Energy Resources and Climate Change	Energy Resources and Climate Change

Students' attitudes about the nature of science were measured using revised Scientific Attitude Inventory (SAI II). The SAI II was first developed by Moore and Suntman (1970) and was subsequently the subject of a thorough re-evaluation by Munby (1983) which highlighted several issues with the surveys validity. This review by Munby (1983) led to the revision of the SAI by Moore & Foy (1997) which addressed the validation issues and created an improved instrument which was re-named the SAI II. The SAI II is a forty item survey featuring a five-point Likert scale where respondents choose an option between strongly agree and strongly disagree. The validity of the SAI II has also been questioned by Lichtenstein et al. (2008). In the re-evaluation of the SAI II they found the instrument to be unreliable for a convenience sample of adolescents. However, despite this criticism Nadelson and Viskupic (2010) found the instrument valid and reliable for use with college students which they attributed to the greater maturity level of college students compared to adolescents. We administered the SAI II on the first day of class to all participants and then again near the end of the semester. The actual day for the post-survey varied due to situational factors such available class time.

Attitudes about the relevance of science were measured using the Changes in Attitude about the Relevance of Science (CARS) survey. The CARS survey was developed by Siegel & Ranney (2003) and is composed of three 25-item surveys. Students respond to each item using a five-point Likert scale where respondents choose an option between strongly agree and strongly disagree. The CARS surveys allows the researcher to gain greater insight about how students' attitudes about the relevance of science changes overs time (Siegel & Ranney, 2003). Additionally, the CARS survey is designed to minimize the impacts of repeated testing bias by dividing 59 items among the three different versions, each version consisting twenty-five items (seventeen version-unique items and eight repeated items). The first of the three versions were administered on the first day of class, the second version during the middle of the semester, and third version was administered at the end of the semester with the post-class iteration of the SAI II. Several of the items from the original CARS survey have been re-worded to make them more geology-specific. For example, item 17 on the first version of the CARS survey was initially worded as follows: "Science class will help prepare me for a career." The word 'science' was replaced with 'geology' and now reads as "Geology class will help me prepare for a career." Only minor wording changes were made to items in order to maintain the instruments original validity and reliability. Similar changes have been made to other attitude instruments, for example, Adams, Wieman, Perkins, & Barbera (2008) modified the Colorado Learning Attitudes about Science Survey (CLASS) from a physics-focused instrument to one that could be used in a chemistry class. Their revisions were much more robust than the changes detailed here yet still maintained the instruments original validity and reliability (Adams et al., 2008).

Sustainability Short Answer Question

Measuring students understanding of sustainability was more difficult because there were neither any quantitative instruments that adequately measured such a construct, nor were there any qualitative instruments that fit into our study design. This led us to develop an easily administered short answer question that students would respond to during their final exam. The design of this question was guided by the framework sustainability science and earth science literacy documents and asked students to perform three tasks: (1) define sustainability in the geosciences, (2) identify topics that deal with sustainability, and (3) describe why one of those topics would be important to know to be scientifically literate. To evaluate students' responses a rubric was developed that scored each of the three parts (Table 3.4) and then was added up to create a total score.

Table 3.4: Sustainability Question Rubric

Question: *In your own words define sustainability as it applies to the geosciences and make a list of topics covered in class that discuss issues related to sustainability. From your list, select one topic that you consider to be the most important topics for the general public to understand in order to be a science literate citizen and elaborate why.*

Part 1: Definition of Sustainability				
Student's definition should describe a balance between humans and the earth systems, state the importance for future generations, and show a consideration beyond self.	0 No definition provided, definition is not logical, or definition is not relevant to the geosciences.	1 Definition incorporates some description of a balance between humans and natural systems but does not cite a central socioscientific issue.	2 Describes a balances between humans and the earth system by citing a socioscientific issues but does not elaborate on the importance for future generations.	3 Definition used in the context of one or more socioscientific issues, or one or more socioscientific issues are used to emphasize the importance of sustainability for future generations.
Part 2: Topic List				
Student will need to make a list of topics discussed in this course that are related to sustainability.	0 No list or only one topic in the list is related to a prominent earth science-related socioscientific issue.	1 Two topics in the list are related to both the course and prominent socioscientific issues.	2 Three topics in the list are related to both the course and the socioscientific issues.	3 All fours topics dealing with the earth science literacy documents and socioscientific issues (i.e. climate change, energy resources, mineral resources, and water) are listed.
Part 3: Science Literate Citizen				
Student must choose a topic from their list. Their elaboration must be logical, tied to earth science literacy, and linked to their definition of sustainability.	0 Discussion not related to their list, is incoherent, or not reasonably answering the question.	1 Discussion not well supported, not tied to socioscientific issues, or only one line of evidence is cited.	2 Students choose a topic from their list tied to the literacy documents. Elaboration is logical using at least two lines of evidence.	3 Topic is tied to the literacy documents and aligned with their definition of sustainability. Elaboration is well-supported, cites correct information, and evidence explicitly states why information is important to society.

Development, validation, and grading of the rubric

The first part of the student's response is intended to gauge whether or not students understand that sustainability in the geoscience has important implications for future generations and that it requires a balance between human society and the ecosystems that we exploit. The field of sustainability science has commonly framed sustainability within a three pillar model; and pillars are represented by the economy, environment, and society (Kajikawa, 2008; Kastenhofer & Rammel, 2005). The three-pillar model stresses balance between human and natural components. The idea of balance is crucial when attempting to achieve and define sustainability. Hay & Mimura (2006) argued that sustainability can only be achieved by conserving both resources and the capacity of the environment to absorb the multiple stresses induced by human activities. Another crucial aspect to defining sustainability is the temporal aspect of resource use and ecosystem viability (Kajikawa, 2008). Sustainability is frequently characterized as a social choice about what to develop, what to sustain, and for how long (Parris & Kates, 2003). This temporal aspect also forces a person to consider the tradeoffs and implications between short-term and long-term processes (Kajikawa, 2008). Consequently, we required that students explicitly state the importance of sustainability for ensuring the prosperity of future generations. The second part of the student's response was designed to quantify their proficiency at identifying course topics that involve issues of sustainability. Sustainability science identifies several important socioscientific issues related to earth science literacy documents. Those issues are climate change, energy and other natural resources, and water scarcity (Kajikawa, 2008), and we required that student's lists focus on these four broad topics. The third part of the question and rubric were designed to elucidate why and how a topic of their choice is an important socioscientific issue, and we wanted students to tie socioscientific issues with lines of evidence tied to geology class. Students could earn three points for each part of the rubric with a maximum possible score of 9 points.

Validation of the rubric was done using two independent raters unaffiliated with the current study. Several meetings were conducted with the raters to ensure the rubric had sufficient face and content validity. Once a sufficient level of agreement was achieved the rubric was piloted using a random sample of 50 sustainability responses from both control and treatment semesters. Students' responses were anonymized; graders could not see the students' name, gender, or which semester the response was from. This preliminary examination lead to an excellent level of agreement (K_w>0.80, p=0.001) on the total score, however, examination of the three individual parts of the scores revealed that the third part did not have a sufficient level of agreement. This lead to the second round of rubric revision following which the the raters scored a different set of 50 responses. Phase two of the rubric validation revealed excellent (K_w=0.81, p=0.001) inter-rater reliability amongst total scores and all three individual parts.

Sustainability questions were graded using the same individuals that participated in the rubric validation phases. Unaffiliated graders were necessary to eliminate any potential for grading bias because the primary investigator also taught the classes and had numerous personal interactions with students over the course of the semester. Grading was completed in two different phases due to time constraints from situational factors and the availability of the two graders. Phase one grading consisted of the first three semesters (CS1, CS2, and

TS1). Phase two grading consisted of responses from the final study semester (TS2). Additionally, unknown to the graders, a random sample of responses from the three previous study semesters that consisted of 20% of the total number of the responses were mixed in with the final set of responses in phase two. This was done determine any ‘drift’ from graders over time. This was also done so to re-confirm the inter-rater reliability between the graders for the TS2 responses and for their previous graded responses. Final analysis showed that the inter-rater reliability for the TS2 responses was (K_w=0.83, p=0.001) and the re-analysis of the sample of previously scored responses showed no significant shift in scoring.

Quantitative scores are only a small part of the dataset created from student’s responses to the sustainability questions. To investigate their responses in more detail we sought to carry out a supplementary qualitative analysis of their answers. We examined their responses in more detail to determine if there were any variations in responses when defining sustainability in the geosciences and to examine how students discussed topics related to sustainability. This qualitative analysis was guided by simplified version of grounded theory (Glaser & Strauss, 1967) as well as the major topics defined by sustainability science. Coding of individual responses and subsequent data analyses was performed using Atlas.ti software in the method described by Smit (2002).

Data Analysis

Data analysis of the SAI II and CARS surveys was carried out using the same methodology described in chapter 2. Participants identified as Juniors and Seniors were combined into a single group labeled Upper Class (UC) for analysis of the CARS survey results. When these data were stratified by academic rank the resulting numbers of juniors and seniors are significantly smaller than freshman and sophomores. Repeated-measures ANOVA tests are used to analyze the CARS data and consequently the small sample sizes of juniors and seniors make them more susceptible to violate the assumption of sphericity and result in type I or II errors.

Survey scores from the SAI II and the CARS surveys were converted to person measures using Rasch modeling and analysis. Rasch analysis was developed by George Rasch (1960) and represents a one-way probabilistic approach based on Item Response Theory (IRT). The application of Rasch analysis and measurement in the social sciences has been most notably discussed in the 1967 Invitational ETS conference and in a wide-variety of subsequent publications (e.g. Andrich, 1978; Boone et al., 2011; Choppin, 1985; Libarkin & Anderson, 2005; Linacre, 1998, 2006, 2010; Miyake et al., 2010; Tong, 2012; Mark Wilson & Adams, 1995; Wright & Stone, 1979; Wright, 1977, 1984). Rasch modeling is both norm-referenced (comparing individuals to the group) and criterion-referenced (measured according to specific standards; Siegel & Ranney, 2003). This is accomplished by considering the difficulty of the item on a continuum, the participant's response to that item (ability), and the probability that a participant will choose a response (Boone et al., 2014). Winsteps software (Linacre & Wright, 2000) was used to compute person measures. Person measures are expressed in logits (log-odds units), which are equal-interval units that can be applied to parametric statistical tests because they avoid many of the issues of the non-linearity of rating scales as well as the non-linearity of raw survey data (Boone et al., 2014).

Logit values for person measures and items were used to construct Wright Maps, which are also known as person-item maps (Wilson. & Draney, 2002). Wright Maps are a visual method of displaying complex rating scale and test data (Boone et al., 2014) and places items and person measures on the same linear scale, much like a ruler. Wright maps simultaneously display both items and person measures because Winstep Rasch analysis transforms both data types into logits (Boone et al., 2014; Linacre & Wright, 2000). Wright maps also allow the researcher to apply qualitative labels to quantitative values calculated in Rasch analysis because scores are shown on the same linear scale (Boone et al., 2014). For example, we identified the most extreme and average items in order to provide qualitative insight into what our study sample most frequently and least frequently agreed with (Siegel & Ranney, 2003). Appendix 7 displays a simplified Wright map and shows how to identify the extreme items used to make those inferences. In order to preserve the distinction between the two instruments the logit values on the SAI II were scaled to a 0-100 scoring system while the CARS survey results are shown with their unmodified logit values.

Results

The results from the CARS survey and SAI II are show in Figures 3.1 and 3.2. CARS data show no significant change over time during CS1 while CS2 did show a significant improvement in students' attitudes toward the relevance of science over time ($F(2,128)=7.913$, $p=0.001$, $\eta^2=0.11$). Students' attitudes toward the relevance of science show a significant improvement over time for both TS1($F(2,52)=10.189$, $p=0.001$, $\eta^2=0.28$) and TS2 ($F(2,84)=3.399$, $p=0.038$, $\eta^2=0.07$). Data from the CARS C (version administered at the end of the semester) can provide insight into the absolute change observed between study groups. CARS C means for CS1 and CS2 do not significantly vary and neither do the scores from TS1 and TS2. Investigating differences grouped by spring and fall semester between CARS C scores show that the mean CARS C score for TS1 (Spring) was significantly higher than in CS1 (Spring; $t(99)=1.982$, $p=0.05$, $d=0.40$). TS2 (Fall) and CS2 (Fall) did not significantly vary. SAI II results show statistically significant decreases in both CS1 ($t(62)=3.04$, $p=0.004$, $d=0.34$), CS2 ($t(76)=5.73$, $p=0.000$, $d=0.56$) and in TS1 ($t(40)=2.85$, $p=0.007$, $d=0.36$). TS2 showed no change between pre- and post-class SAI II surveys. Additionally, all four study groups show no significant differences between their attitudes toward the nature of science at the beginning and end of the semesters. The Pearson correlations for instruments administered at the end of the semester are shown in Table 3.5

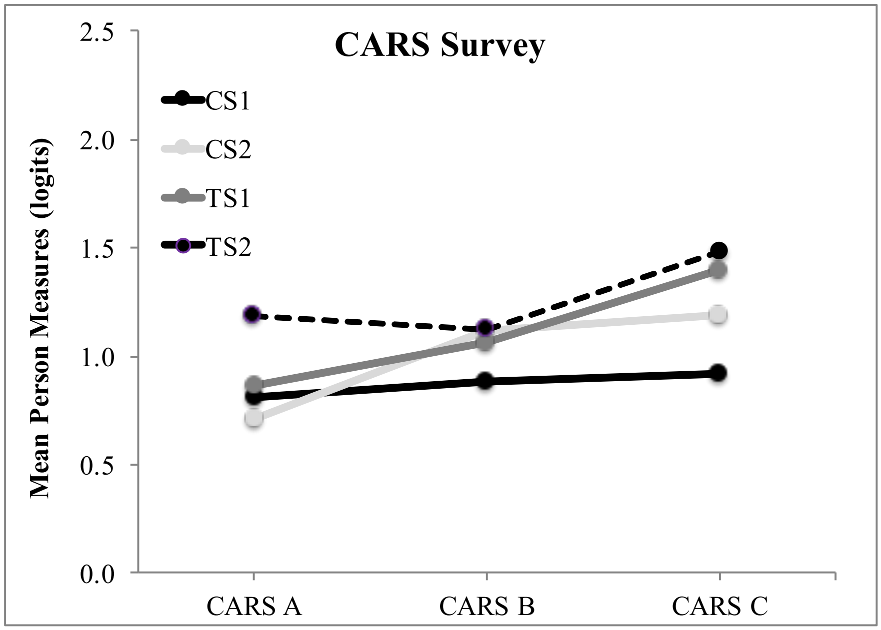


Figure 3.1: CARS survey results for all four study semesters.

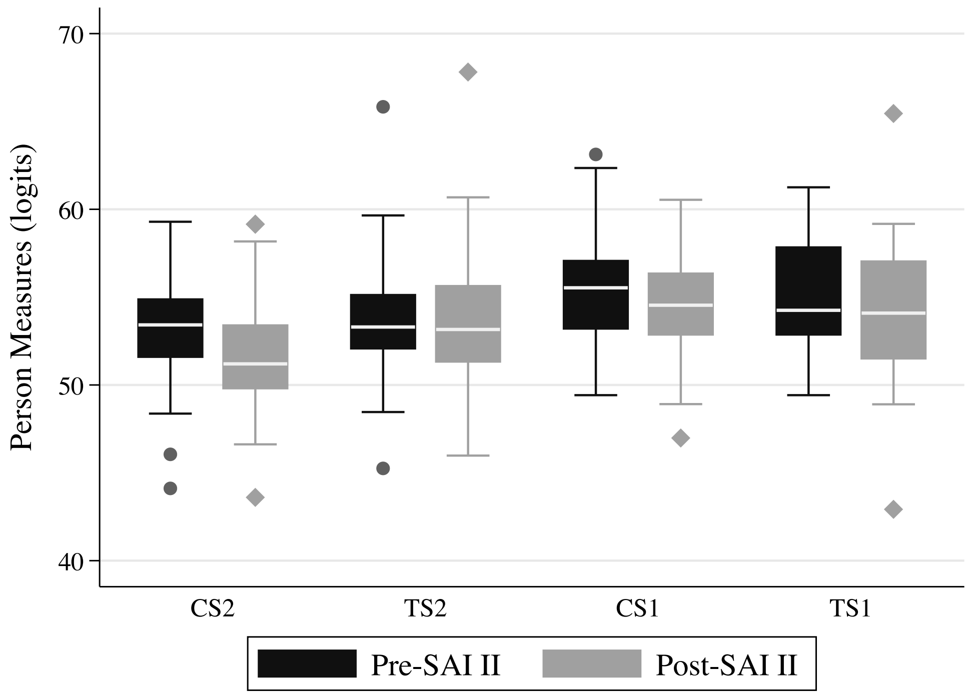


Figure 3.2: Results from the SAI II for all four study semesters.

Table 3.5: Pearson correlation matrix for all post-semester instruments from all four study semesters.

	1	2	3	4
CS1				
1. Post-SAI II				
2. CARS C	0.53			
3. Sust. Quest.	0.33	0.20		
4. Exam Total	0.04	0.02	0.13	
5. Final Grade	0.05	0.07	0.23	0.42
CS2				
1. Post-SAI II				
2. CARS C	0.71			
3. Sust. Quest.	0.11	0.18		
4. Exam Total	-0.19	-0.14	0.08	
5. Final Grade	0.35	0.31	0.17	-0.12
TS1				
1. Post-SAI II				
2. CARS C	0.20			
3. Sust. Quest.	0.27	0.19		
4. Exam Total	-0.15	0.03	-0.01	
5. Final Grade	0.26	0.16	0.19	-0.40
TS2				
1. Post-SAI II				
2. CARS C	0.72			
3. Sust. Quest.	0.14	0.21		
4. Exam Total	-0.11	-0.23	-0.08	
5. Final Grade	0.14	-0.01	0.21	-0.32

Average final grades and exam totals (mean score of the four course exams) for all four semesters are shown in Figure 3.3 and are used as a way to measure students' learning gains and are collectively referred to as class performance. There is no statistically significant variation in exam total amongst the four study groups. An analysis of variance amongst final grade does show a significant difference between semesters ($F(3,305)=8.61$, $p=0.001$), and a Bonferroni post-hoc analysis indicated that much of that variation is attributed to TS2.

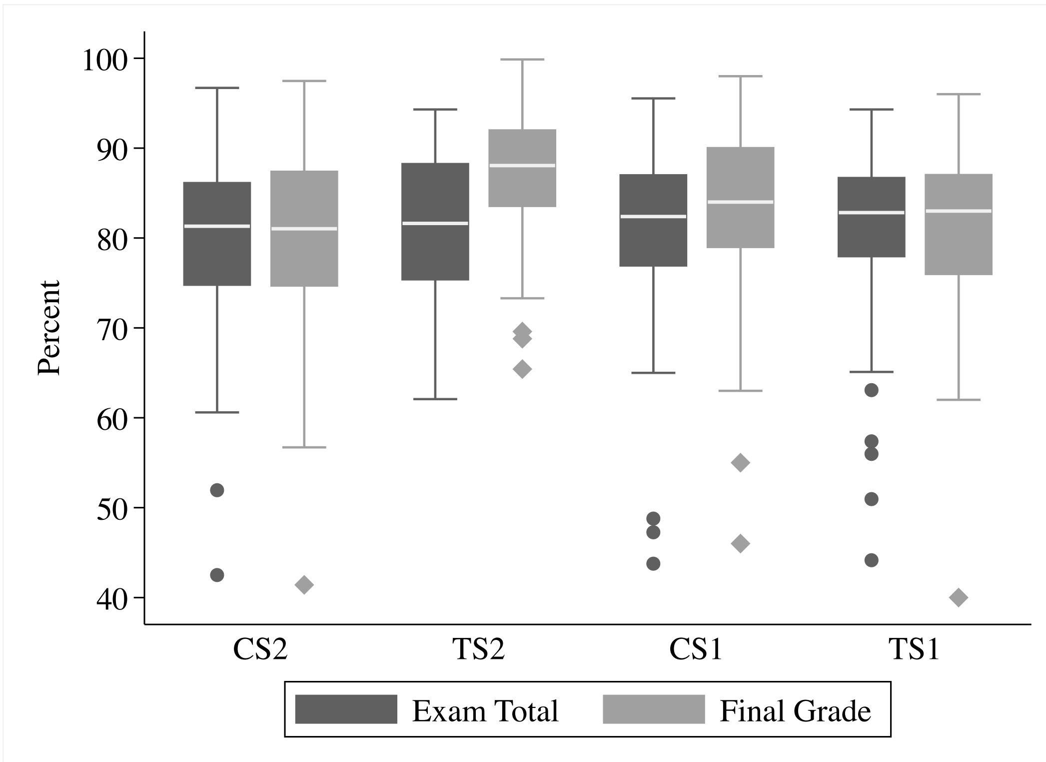


Figure 3.3: Final grade and exam totals for all four study semesters. Exam total is the average score for all four of the course exams.

Sustainability Question

Results from the grading of the sustainability responses for all four study semesters are shown in Figure 3.4. A one-way analysis of variance did not show any significant differences. Comparisons of control and treatment grouped by fall or spring semester showed a statistically significant different between CS2 (Fall) and TS2 (Fall) ($t(151)=2.20$, $p=0.03$, $d=0.36$).

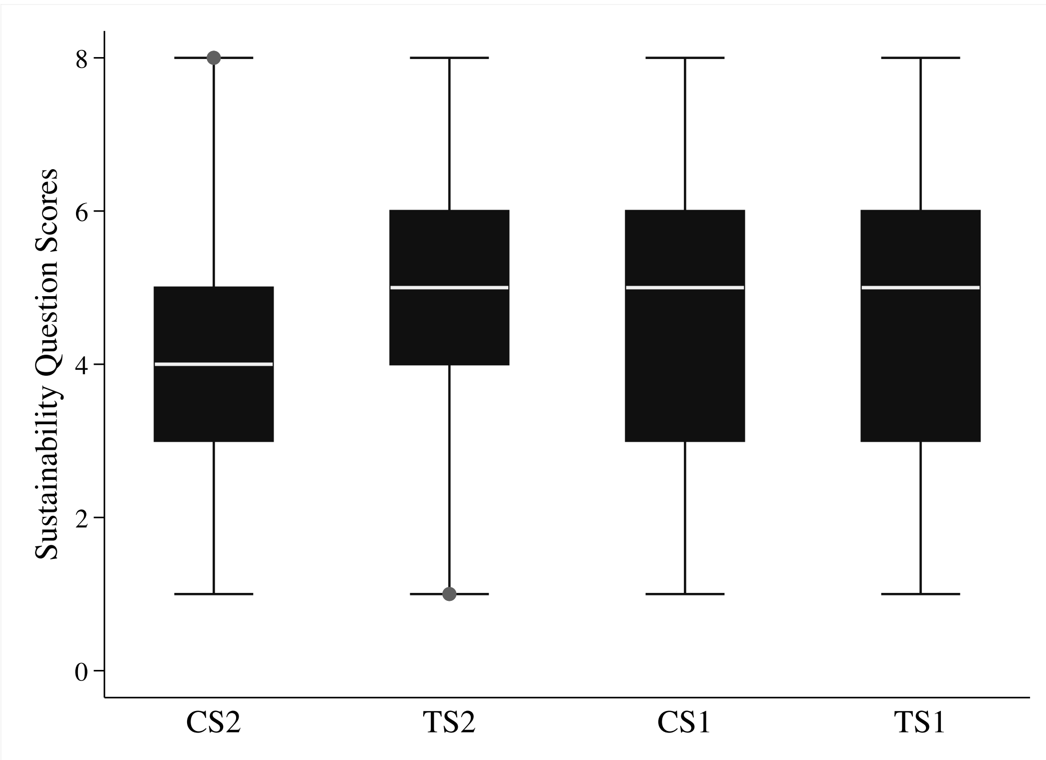


Figure 3.4: Sustainability scores for all four study semesters.

Academic Rank

CARS survey results stratified by a students' academic rank are shown in Figure 3.5. The numbers of juniors and seniors are significantly smaller than freshman and sophomores. Therefore, in order to minimize the potential for Type I and II errors, data for juniors and seniors were combined into one group labeled upper class (UC). Freshman and sophomores for both CS1 and CS2 show no statistically significant improvement in their attitudes toward the relevance of science over time. UC students for CS1 show no improvement over time while students for CS2 show a statistically significant improvement in their attitudes over time ($F(2,26)=3.259$, $p=0.022$, $\eta^2=0.20$). Freshman from both TS1 ($F(2,16)=6.757$, $p=0.007$, $\eta^2=0.46$) and TS2 ($F(2,16)=6.757$, $p=0.007$, $\eta^2=0.46$) show a statistically significant improvement over time. Sophomores from both TS1 and TS2 show no improvement in their attitudes over time. UC students from TS1 show no statistically significant change in attitude over time while UC students from TS2 do improve over time ($F(2,28)=4.946$, $p=0.014$, $\eta^2=0.26$)

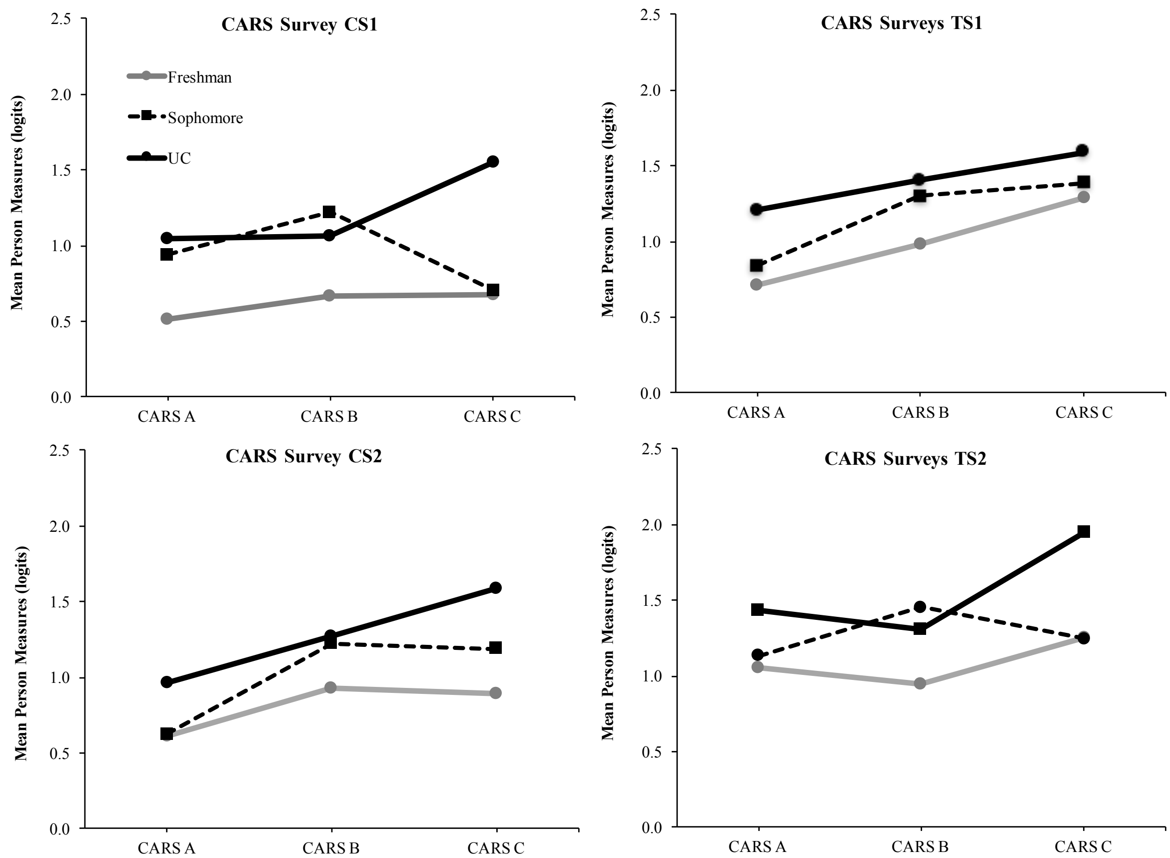


Figure 3.5: CARS survey results stratified by academic rank for all four study semesters.

SAI II results stratified by students' academic rank are shown in Figure 3.6. In TS1 only two of the six seniors for the semester took the post-SAI II and each of those students got the same score. There is only minimal variation amongst the four semesters. An analysis of variance for both the pre- and post-class iterations of the SAI II showed no significant differences amongst academic rank. In CS1 sophomores showed a significant decrease ($t(13)=3.367$, $p=0.005$, $d=0.62$) in SAI II scores over the course of the semester. Freshman ($t(35)=4.118$, $p=0.000$, $d=0.62$) and sophomores ($t(23)=3.868$, $p=0.001$, $d=0.55$) from CS2 both showed significant decreases in their SAI II scores. With the exception of freshman from TS1, the other three academic ranks showed no significant changes during treatment semesters. Freshman in TS1 showed a statistically significant decrease ($t(17)=2.71$, $p=0.015$, $d=0.30$) in SAI II score.

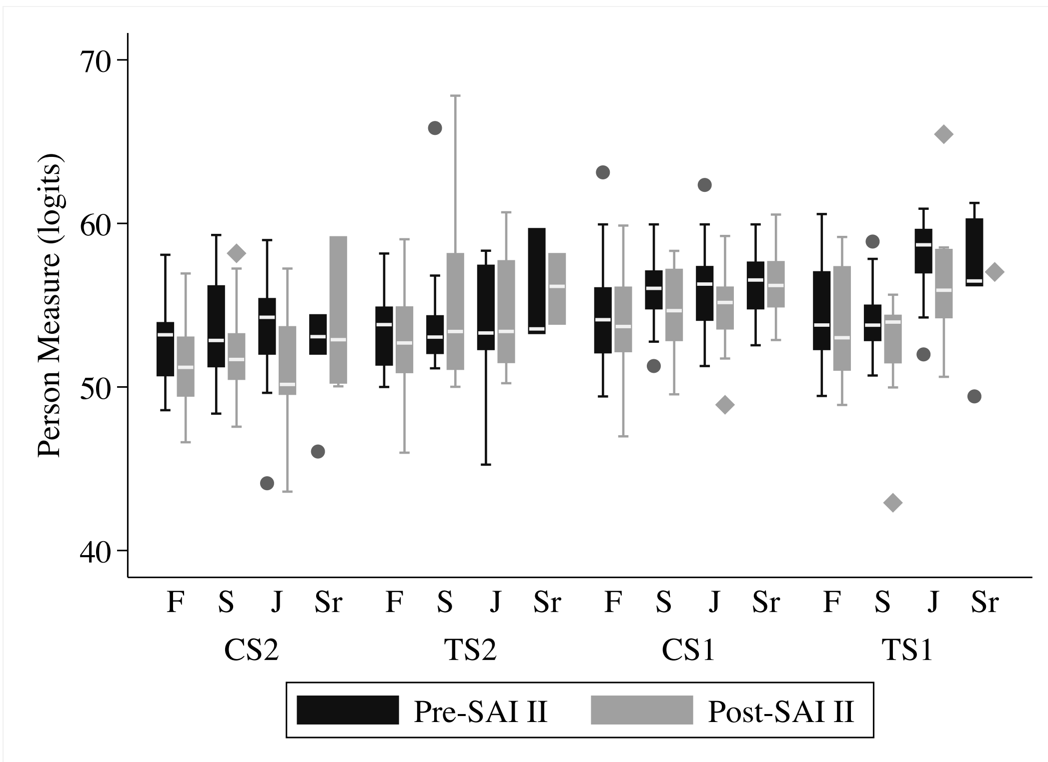


Figure 3.6: SAI II scores stratified by academic rank for all four study semesters. F=freshman, S=sophomore, J=junior, Sr=senior.

Sustainability scores stratified by academic rank are shown in Figure 3.7. In general, most students scored between three and six on this exercise. Freshmen, sophomores, and seniors show no significant variation as a function of study group. Juniors showed a statistically significant ($F(3,58)=2.85$, $p=0.038$, $\eta^2=0.13$) variation as a function of study group, and a Bonferroni post-hoc analysis revealed that much of that variation can be attributed to scores from CS1. Juniors scored lower than the other three academic ranks during CS1, but scored highest during CS2. Freshman, sophomore, and seniors do not display much difference amongst study groups. TS2 displays a semi-linear increase in mean and median scores from the Freshman (lowest) to Senior (highest).

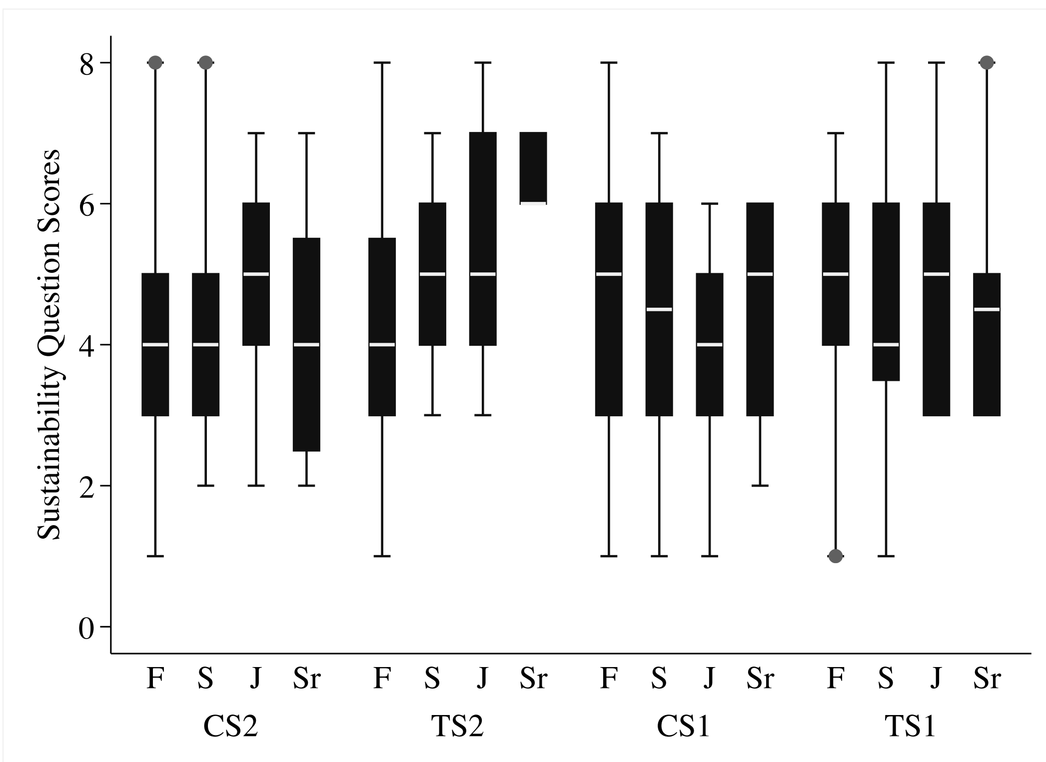


Figure 3.7: Sustainability scores for all four study groups stratified by academic rank.

Students' class performance stratified by academic rank for all four study groups is shown in Figure 3.8. Most of the semesters' final grade and exam totals show a small increase from freshman and sophomore to juniors and senior. TS2 shows a distinct pattern. Freshman and sophomores performed similarly to juniors and better than seniors on their final grades. TS2 is also the only semester with no final grades or exam totals below 60%. Final grade displays statistically significant variability on the basis of academic rank ($F(3,305)=2.63$, $p=0.05$, $\eta^2=0.03$).

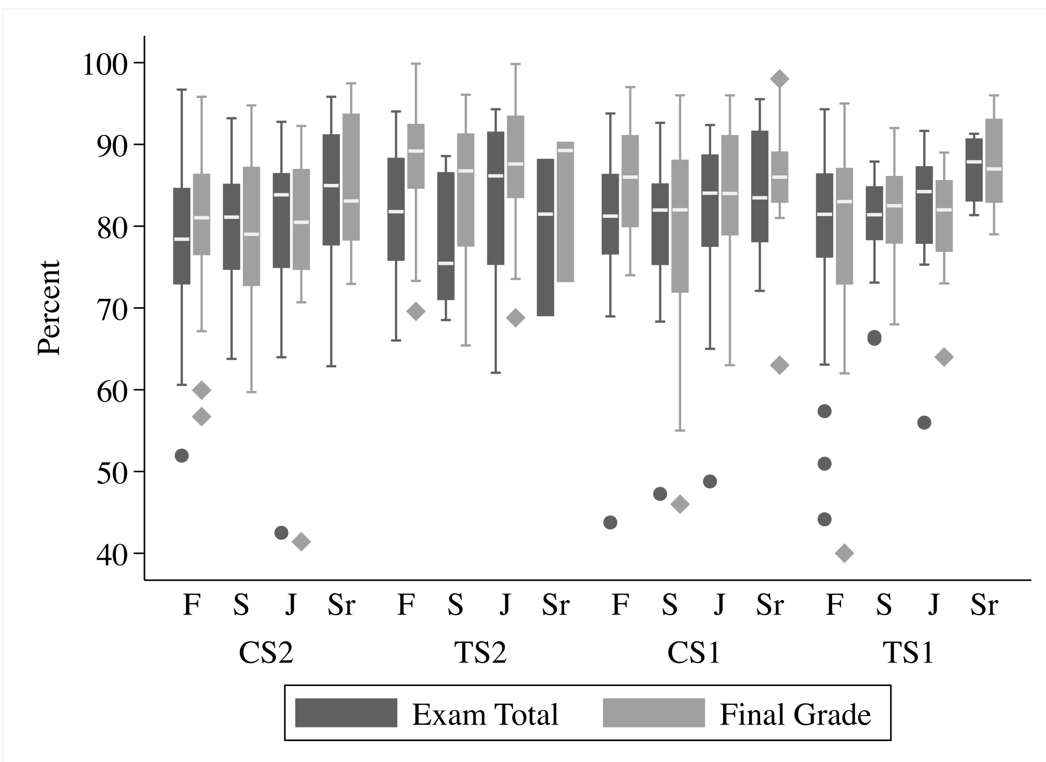


Figure 3.8: Final grad and exam totals for all four study semesters.

Gender

CARS survey results stratified by gender for all four study groups are shown in Figure 3.9. Neither men nor women show a significant improvement in their attitudes toward the relevance of science in CS1. In CS2 men showed a statistically significant improvement in their attitudes ($F(2,80)=11.066$, $p=0.001$, $\eta^2=0.22$) and women showed no change over time. In TS1 both women ($F(2,32)=5.188$, $p=0.011$, $\eta^2=0.24$) and men ($F(2,18)=5.424$, $p=0.014$, $\eta^2=0.38$) showed statistical significant improvement over time with women finishing the semester with a statistically higher mean CARS C survey score than men. Women did not show a significant improvement during TS2 but men again showed a statistically significant improvement in their attitudes over time ($F(2,52)=4.431$, $p=0.017$, $\eta^2=0.15$). All four study semesters revealed that both men and women started the semester with similar scores on the first CARS survey (CARS A) and then attitudes of men and women diverged over the course of the semester (with the exception of CS1).

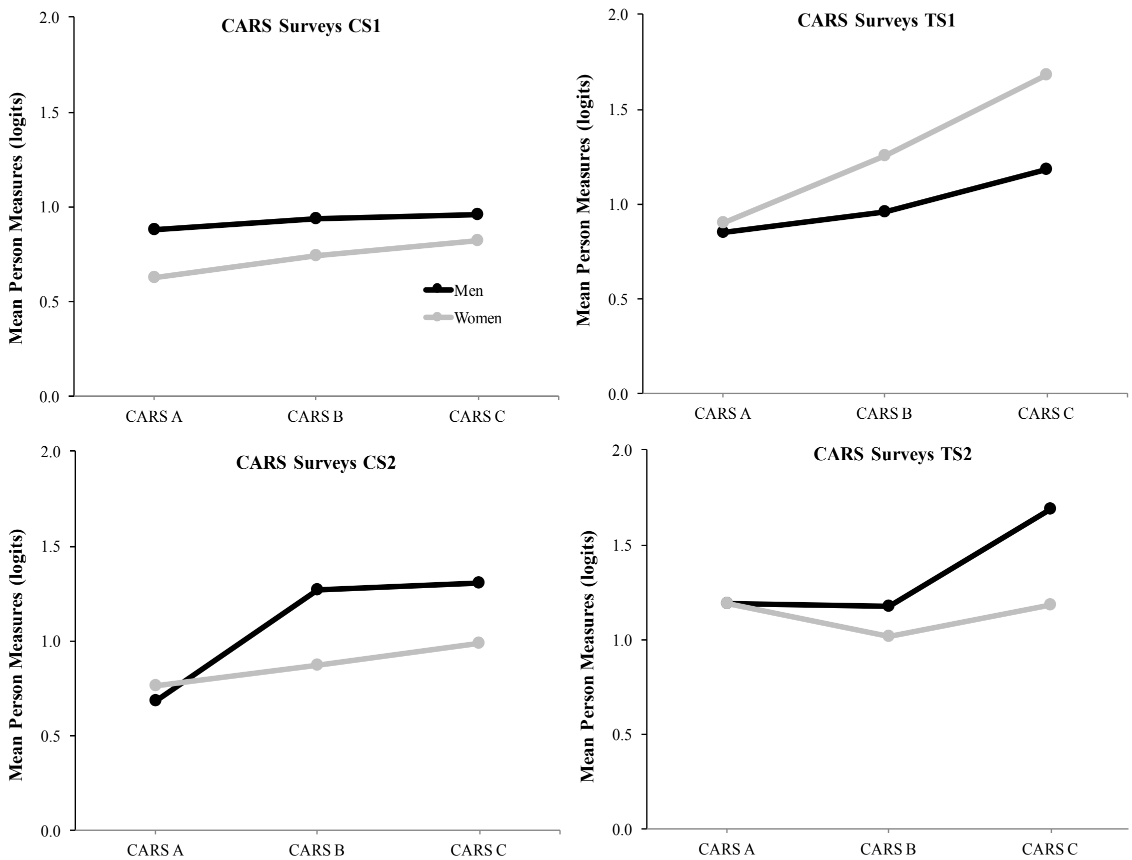


Figure 3.9: CARS survey results stratified by gender for all four study semesters.

SAI II results stratified by gender for all four study semesters are shown in Figure 3.10. Results grouped on gender still show minimal variability. Men and women for all four study semesters have slightly positive to neutral attitudes toward the nature of science at the beginning of the course. An analysis of variance showed that no significant variability exists for either the pre-SAI II or post-SAI II scores amongst gender. Men ($t(45)=3.01$, $p=0.004$, $d=0.41$) in CS1 had a significant decrease in SAI II score while women showed no change. In CS2 both men ($t(43)=3.14$, $p=0.003$, $d=0.40$) and women ($t(32)=5.28$, $p=0.000$, $d=0.82$)

show a statistically significant decrease in SAI II score over the semester. In TS1 neith women nor men showed a significant decrease in SAI II score. Neither men nor women showed any significant change during TS2.

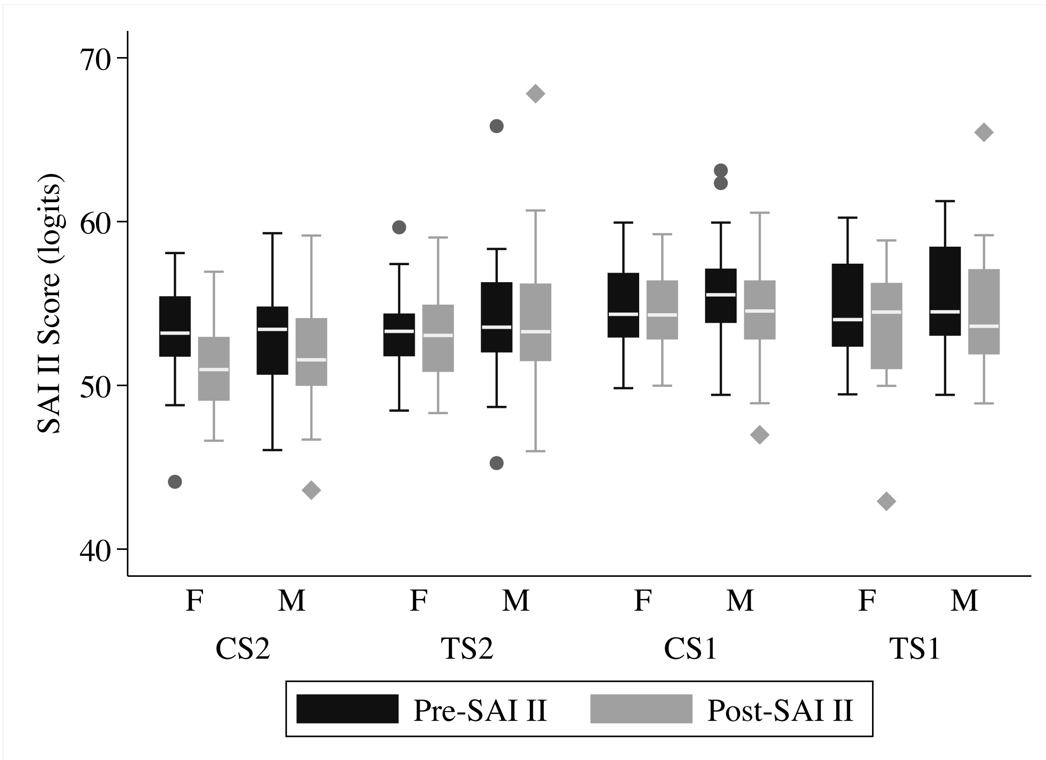


Figure 3.10: SAI II scores for stratified by gender for all four study semesters. M=male, F=female.

Sustainability scores stratified by gender for all four study semesters are shown in Figure 3.11. No significant differences were calculated between men and women in CS1, TS1, or TS2; and the ranges for those semesters and medians are similar. Women in CS2 scored significantly higher on the sustainability question than men ($t(82)=2.906$, $p=0.005$, $d=0.65$).

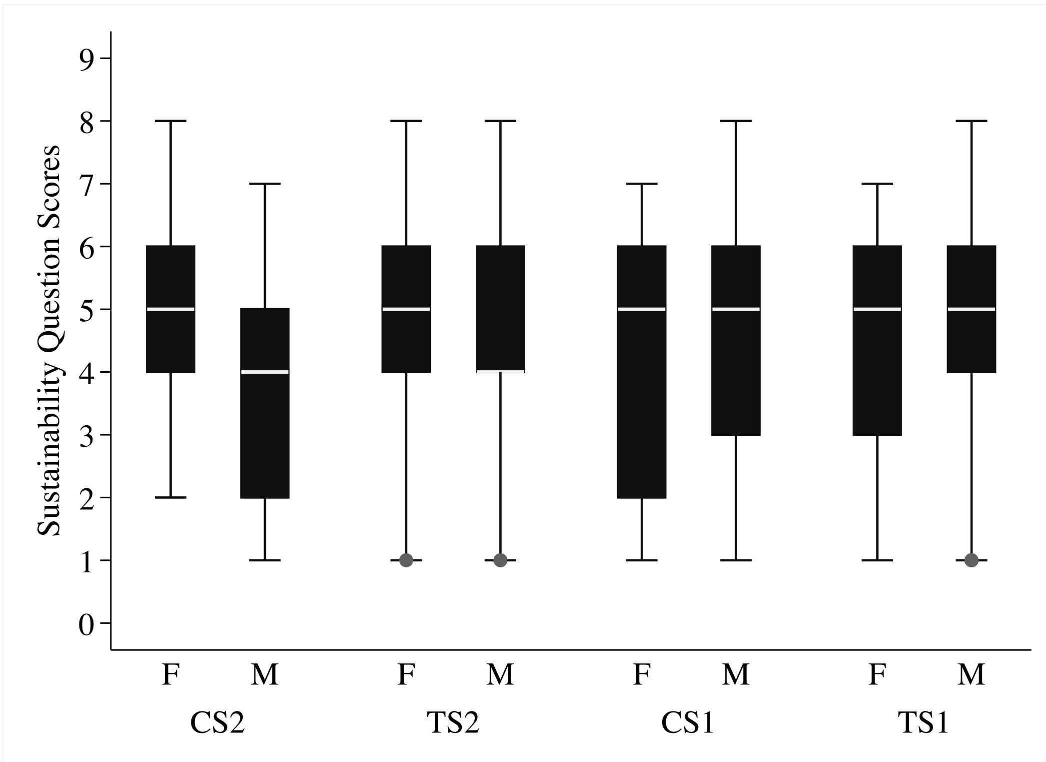


Figure 3.11: Sustainability scores stratified by gender for all four study semesters. F=female. M=male.

Class performance stratified by gender is shown in Figure 3.12. There is no significant variation among exam total or final grades when data are grouped by gender. TS2 is the again the only semester where none of the exam total averages or final grades fell below 60%. The other three semester show a much greater distribution of scores for both men and woman ranging on average from 97% to 40%.

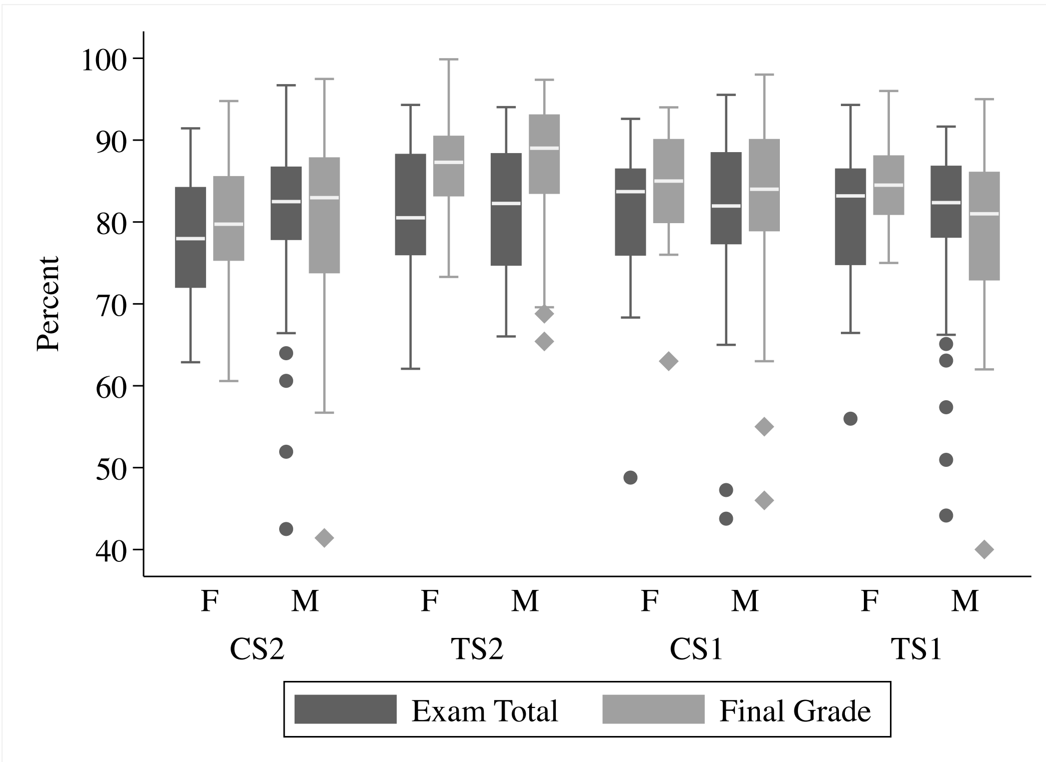


Figure 3.12: Class performance stratified by gender for all four study semesters. F=female. M=male

Degree Type

CARS survey results for STEM and non-STEM major for all four study semesters are shown in Figure 3.13. STEM majors consistently perceived science more relevant than non-STEM majors at the beginning and end of the semester. In CS1 both STEM and non-STEM majors showed no significant difference over time in their attitudes toward the relevance of science. STEM majors also showed no significant changes in CS2 but non-STEM majors did show a significant positive change over time in their attitudes ($F(2,116)=5.672$, $p=0.004$, $\eta^2=0.41$). In TS1 neither STEM nor non-STEM majors showed a significant change over time. In contrast, STEM majors showed a significant positive change over time in TS2 ($F(2,22)=5.02$, $p=0.016$, $\eta^2=0.31$).

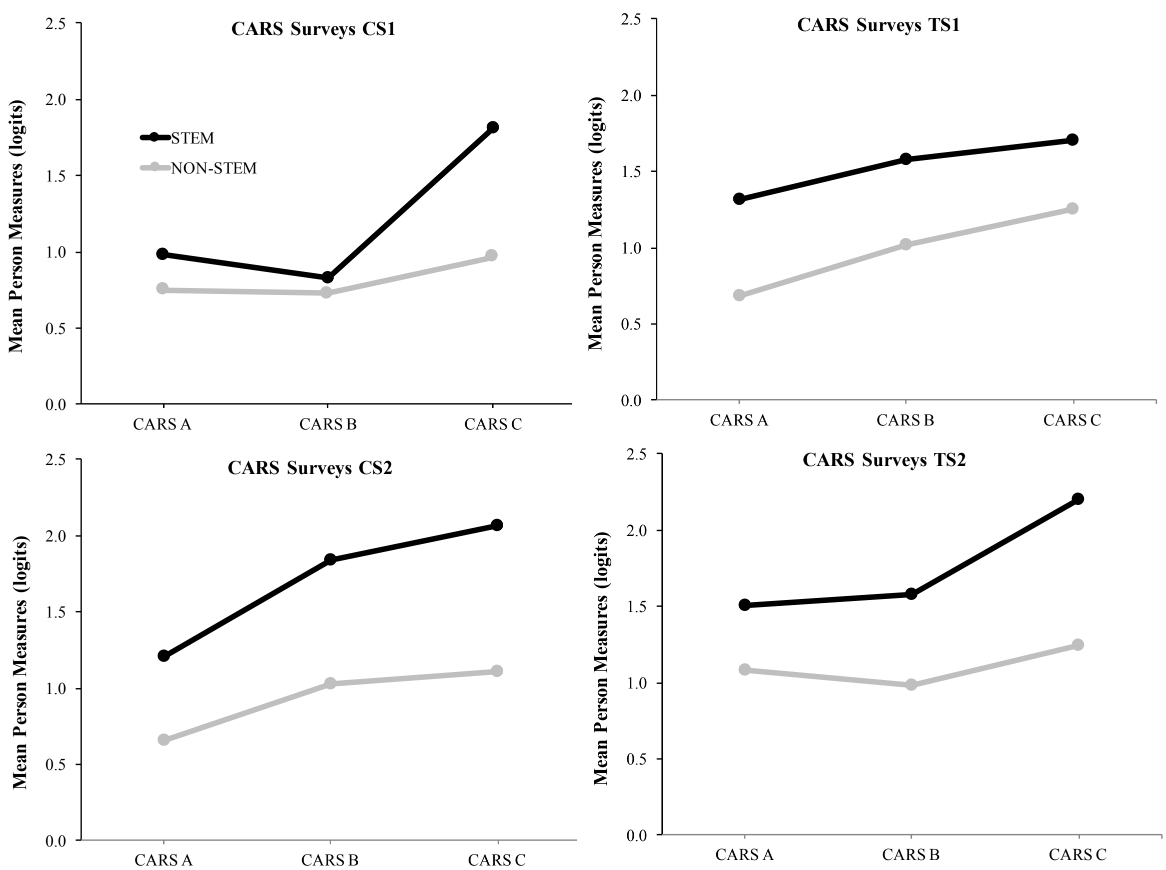


Figure 3.13: CARS survey scores stratified by degree type from all four study semesters.

SAI II scores stratified by degree type are shown in Figure 3.14. Analyses of variance show significant variability amongst pre-SAI II scores for non-STEM majors ($F(2,233)=6.86$, $p=0.0002$, $\eta^2=0.06$) and for STEM majors ($F(3,56)=2.77$, $p=0.05$, $\eta^2=0.13$). Post-SAI II scores exhibited significant variability for non-STEM majors ($F(3,195)=7.70$, $p=0.0001$, $\eta^2=0.03$) but not for STEM majors. In CS1 non-STEM showed a significant decrease in their SAI II score ($t(45)=-2.67$, $p=0.010$, $d=0.37$) but STEM majors showed no significant changes. In CS2 non-STEM majors also showed a significant decrease in SAI II score ($t(71)=-5.267$, $p=0.001$, $d=0.52$) while STEM majors showed no differences between pre- and post-class surveys. In TS1 non-STEM majors showed a decrease in scores ($t(31)=-3.41$, $p=0.002$, $d=0.37$) but STEM majors again showed no changes. In TS2 non-STEM majors showed no significant changes while STEM majors showed a significant improvement in their attitudes toward the nature of science ($t(10)=2.28$, $p=0.045$, $d=0.34$).

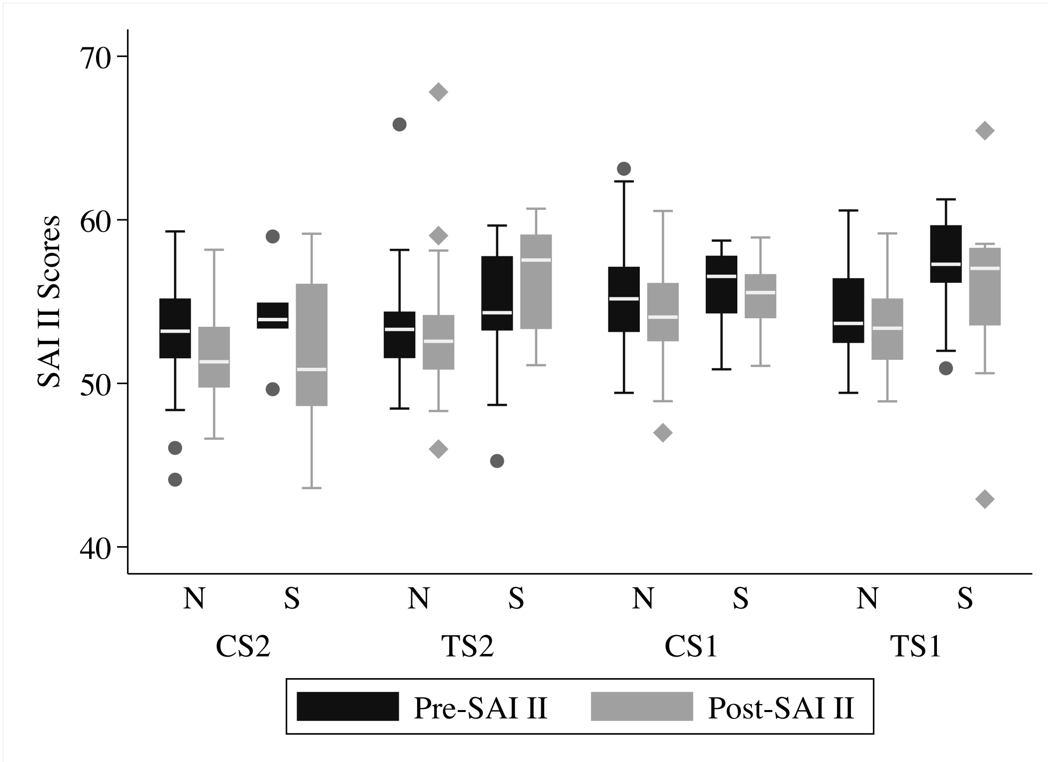


Figure 3.14: SAI II scores stratified by degree type from all four study semesters. N=non-STEM majors and S=STEM majors.

Sustainability questions scores stratified by degree type are shown in Figure 3.15. Scores for both groups were consistent amongst all four semesters except for TS2 where STEM majors scored statistically higher than non-STEM majors ($t(67)=2.733$, $p=0.008$, $d=0.14$).

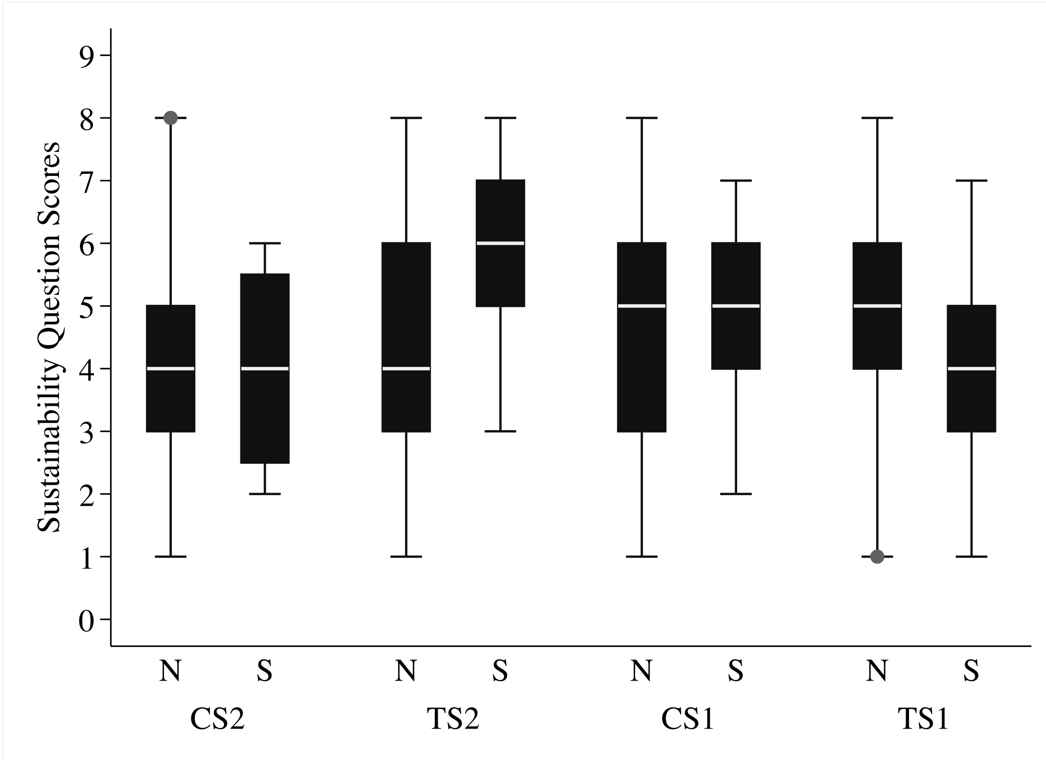


Figure 3.15: Sustainability questions scores stratified by degree type for all four study semesters. N=non-STEM, S=STEM.

Class performance stratified by degree type for all four study semesters are shown in Figure 3.16. No significant variability was determined amongst students' mean exam totals when stratified by degree type. Analysis of final grade showed statistically significant variable among non-STEM students ($F(3,239)=6.10$, $p=0.001$, $\eta^2=0.05$).

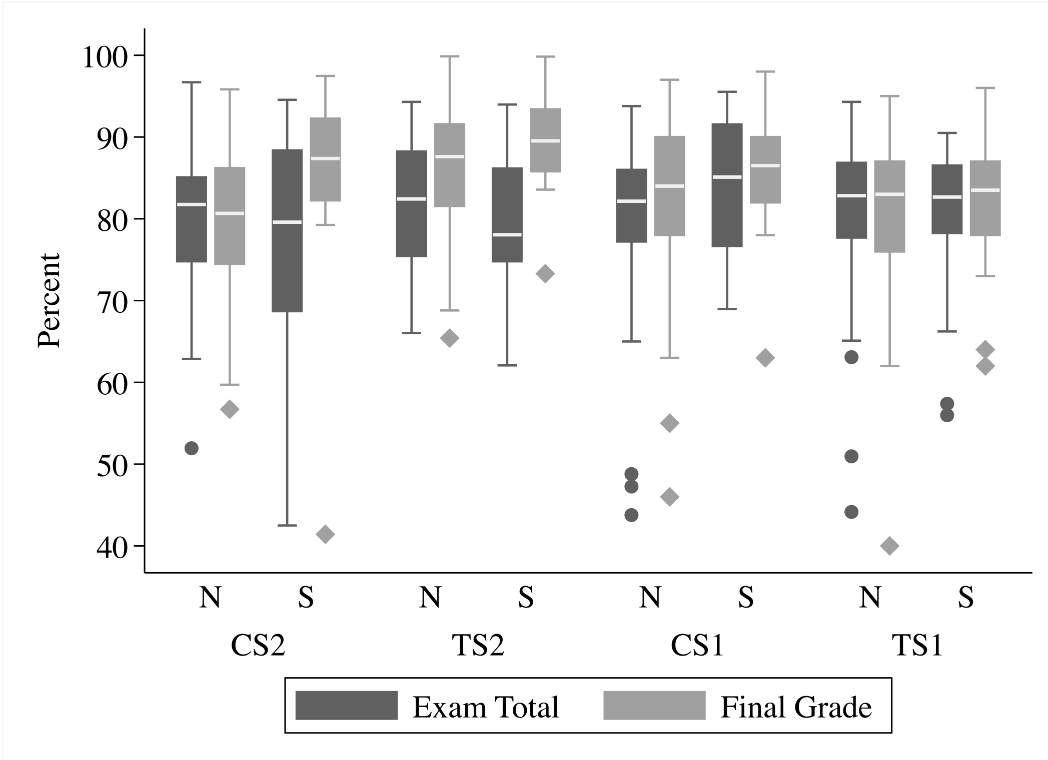


Figure 3.16: Class performance stratified by degree type for all four study semesters. N=non-STEM, S=STEM.

Qualitative Analysis of Sustainability Questions Responses

Analysis of students' responses was done using a modified version of grounded theory (Glaser & Strauss, 1967) as well as the major topics defined by sustainability science. Coding of individual responses and subsequent data analyses was performed using Atlas.ti software in similar method as described by Smit (2002). Iterative comparison of the responses led to the development of six codes. Two codes emerged from students' attempts to define sustainability in the geosciences: 1) Sustainability (Future), which encompasses occurrences where students explicitly describe the importance of conservation resources for future generations; 2) Sustainability (Balance), which encompasses occurrences where students explicitly describe the importance of a balance between humanity and natural systems. The other four codes identified occurrences where students discussed one of the following major topics tied to both the geosciences and sustainability science: 1) Climate change, 2) Energy Resources, 3) Water, and 4) Mining and Minerals.

Results of the coding show differences between the control and treatment groups percentage of coded statements (Table 3.6) and the calculated cooccurrence values (Table 3.7). Results of our qualitative analysis revealed a significant difference among coded statements describing a balance between humans and natural systems between control and treatment semesters. There is also a large discrepancy in the number of coded statements for Mining and Minerals between control and treatment groups.

Table 3.6: Percent of coded statements from Control and Treatment groups.

	Def. of Sust.-Balance	**Def. of Sust.-Future**	**Climate Change**	**Energy Resources**	**Water**	**Mining and Minerals**
Control	2%	7%	21%	29%	38%	4%
Treatment	0.4%	6%	24%	24%	26%	19%

Table 3.7: Code cooccurrences for combined Control and Treatment groups.

CONTROL	Climate Change	Energy Resources	Mining & Minerals
Energy Resources	0.23		
Mining & Minerals	0.07	0.06	
Water Resources	0.16	0.16	0.05
TREATMENT	Climate Change	Energy Resources	Mining & Minerals
Energy Resources	0.30		
Mining & Minerals	0.21	0.23	
Water Resources	0.17	0.18	0.22

Discussion

The impact of an explicit and repeated emphasis of socioscientific issues in a physical geology class show variable positive changes between control and treatment semesters. Both treatment semesters (TS1 & TS2) showed significant improvements in students' attitudes about the relevance of science over time, but the second control semester (CS2) also showed a significant improvement over time with a larger effect size than TS2. This could suggest that something other than the InTeGrate materials may be capable of influencing attitudes about relevance that was not controlled for in this study However, students in TS2 started out the semester with the highest CARS survey scores measured in this study, suggesting some degree of diminishing returns on the possible changes to students' attitudes over the course of a fifteen-week semester.

Our interpretation of the data collected from women in this study suggests that the socioscientific focus provided by the InTeGrate materials may be a more effective method of reaching underrepresented groups within the geosciences. Amongst all four study semesters UC, male, STEM majors had the highest CARS C scores and therefore found geoscience more relevant than freshman, sophomores, non-STEM majors, and women. However, during the treatment semesters both men and women showed a significant improvement in their attitudes about the relevance of science. This suggests that the socioscientific focus provided by the InTeGrate materials may influence men and women equally, however, men still had higher CARS C scores than women during the treatment semesters.

Emphasis on socioscientific issues in a physical geology class reduced the negative impact of a physical geology course on students' attitudes toward the nature of science. Freshman and sophomores showed significant decreases in the attitudes toward the nature of science after taking a physical geology class during the control semesters (Figure 3.6). This observation was true for both men and women during the control semesters (Figure 3.10). In the treatment semesters the stratified data for gender and academic rank do not consistently show these statistically significant decrease in attitudes. Conversely, in treatment semesters (with the exception of STEM majors) the majority of the data show that students' attitudes do not significantly decrease after taking a physical geology course (Figure 3.14). Additionally, STEM majors in TS2 showed a statistical improvement toward more positive attitudes about the nature of science. This result was not seen in TS1 where no difference was observed in

STEM majors' attitude scores. These data show that the explicit connection to socioscientific issues in class can make positive change to students' attitudes. Unfortunately, these data can also be interpreted to infer that a standard physical geology course without a pervasive emphasis on socioscientific issues may be turning students' attitudes about the nature of science more negative.

This study shows that attitudes about the nature of science and the relevance of science can be influenced by a single course, and this is contrary to the results of similar studies investigating the science attitudes of undergraduates. Our results are supported by the findings of Zo'bi (2014) and Sadler and Zeidler (2005) who both show that explicitly displaying the connections between science and society will have positive impacts on students' perceptions of science. We propose a connection between student perceptions of course relevance with attitudes about the nature of science. Our data does show a high significant correlation between CARS C scores and SAI II scores during treatment semesters (Table 3.5). There are several conceptually similar items between the CARS C and the SAI II, but they are not similar enough to account for this correlation. A student's perception of the relevance of science is closely tied to the emotional domain (Holbrook & Rannikmae, 2009; Keller, 1983) and there is a portion of a student's attitudes toward the nature of science derived from an emotional perspective (Moore & Foy, 1997). Therefore, because of this overlap amongst the emotional domain of both the nature of science and the perceived relevance of science, class content designed to influence a students' attitudes toward the relevance of science may also be an effective method of altering their attitudes on the nature of science.

The emphasis of socioscientific issues did not have a significant negative or positive impact on students' class performance. Relevance has been connected to learner empowerment and motivation to study (Keller, 1983). However, our data do not support this statement because our results show non-significant to negative correlations between CARS C scores and students final grades and exam totals (Table 3.5). Instead, our results parallel the findings of Warren (2011) who showed no connection between content relevance and student cognitive learning. Warren suggested that the lack of correlation between relevance and performance could be due to the fact that multiple instructors were used to teach classes. Our data provides evidence contrary to that hypothesis because we employed the same instructor for all four study groups and still found no statistically significant link between students' class performance and how relevant they found scientific information. In TS2 students CARS C scores negatively correlated with both final grade and exam total. These mixed findings could suggest that there is no significant relationship between achievement and students' perceptions of the relevance of science or it could could be more complex than suggested in Keller (1983).

Data from the sustainability question scores show minimal differences between control and treatment semesters. Students in TS2 did show a significant improvement in scores over CS1 and CS2. In TS2 much of that improvement was from students identified as STEM majors while non-STEM majors show no difference between control and treatment groups. This suggests that the emphasis on socioscientific issues in a physical geology class can positively influence students understanding of sustainability if they already decided to major in STEM discipline. As educators this observation should seem obvious but it raises the question of how can we help non-majors build a better understanding of sustainability, especially if our goal is to help our students develop science literacy?

Results of the coding of student responses to the sustainability question show differences between the control and treatment groups. Students from both control and treatment groups do not typically define sustainability by referring to both the need for balance between humans and natural systems and its importance for ensuring the prosperity of future generations (Table 3.6). Students who accurately define sustainability primarily discuss its importance in the context of ensuring the prosperity of future generations. This could suggest that despite the focus on socioscientific issues in the treatment semesters students do not internalize the importance of these two central sustainability-related concepts. These data could also indicate a need for physical geology courses to find more effective strategies to emphasize the importance a balance between society and the earth system.

Cooccurrence values show quantitatively how often certain codes are found within the same coded segments (in our analyses a coded segment was a sentence). For example, a code cooccurrence value of 1 would indicate that two codes were always found in the same statements. Code cooccurrence values (Table 3.7) for the four topical codes (climate change, water, energy resources, mining and minerals) were higher in the treatment semesters than in the control semesters. These data can be interpreted to suggest that students in the treatment semesters are discussing sustainability in a more integrated manner than students from control semesters. It could also be used to infer that students in the treatment semesters are gaining a more nuanced understanding of sustainability from systems perspective by presenting several discrete topics within the same statements.

The frequency of coded statements and cooccurrence values for Mining and Minerals are significantly higher during treatment semesters. During the control semesters minerals were covered in a highly traditional manner, and mining and its impacts were rarely discussed explicitly. For example, the focus of minerals in a traditional geology course in on topics such as the definition of a mineral, classes of minerals, rock cycle and the three different types of rocks (e.g. igneous, sedimentary & metamorphic) formed from one or more minerals. Mining and its impacts to society are a central component of the Human's Dependence on Earth's Mineral Resources module used in the treatment semesters. Therefore, this stark contrast between control and treatments semester could explain the discrepancy in coded data. Analyzing the coded results from the control semesters revealed that out of approximately 19,000 words from the responses of 166 students the word "minerals" was mentioned four times. The way we use minerals and their rates of consumption are an important aspect of both earth science literacy (Wysession et al., 2012) and sustainability (Jerneck et al., 2011; Kajikawa, 2008). An introductory geology course that deals with mining-related processes relatively briefly may not make sufficient impressions on students to have them recognize their role in a sustainable future. Few other courses will provide an opportunity to convey the importance of our mineral-based resources is an area where physical geology should be more effective. In contrast, student responses from the treatment semesters show a significant increase in the mention of minerals, mining and their importance. Not only were those topics mentioned more frequently but the order of magnitude increase in cooccurrence values with all three of the other codes suggests that students are integrating the use of minerals and mining in with other topics, such as climate change and energy resources, when discussing sustainability. This is evidence supporting of the effectiveness of InTeGrate materials to present geoscience information in a way that changes how students think about a complex topic such as sustainability.

Study Limitations

This study has many of the same limitations as were discussed in Chapter 2. The quantitative nature of our research design limited the depth to which we could analyze differences in students' attitudes between control and treatment semesters. However, the use of qualitative methods could have dramatically reduced the numbers of students included in our study and could have introduced a high level of volunteer bias. To qualitatively characterize how students' attitudes would change over the course of the semester would have required students to volunteer for multiple lengthy interviews outside of class. This would have been difficult because we did not have any method in place to incentivize participants.

Results during TS2 were more consistent than in TS1. This could be the result of unforeseen and unavoidable difficulties that accompany the first implementation of new course materials. TS1 was the first time the instructor had included so many InTeGrate units into their course. Consequently, this led to difficulties determining how to adapt materials to physical geology at NCSU. Most importantly, the instructor noted that the course had a very inconsistent pacing with the quantity and difficulty of worked needed to be completed by students. This created an increased level of frustration amongst some students during the semester and that was also evident in the course evaluations. Throughout the first semester the instructor made careful note about what worked and what did not work. These lessons learned, along with feedback provided by students during course evaluations, led to a round of revision regarding how the InTeGrate materials were incorporated into the course. These revisions were piloted during a much small summer session and were again tuned to better fit the large class taught during the upcoming fall semester (TS2). The revised method of implementing the InTeGrate materials provided a much more organized and efficient experience both for the instructor and students. Students' perceptions of a course do have an impact on their attitudes toward the nature of science and its relevance. Therefore, it is reasonable to assume that the frustrations voiced by students in TS1 may have had some influence on their attitudes. Student feedback and evaluations for TS2 were more positive.

Conclusion

The repeated and explicit emphasis of socioscientific issues provided through the InTeGrate course materials reduced the negative impacts of a standard physical geology class on students' attitudes about the nature of science; and it helped students to perceive geoscience content to be more relevant. The emphasis on the intersections between science and society also provided students with a more integrated view of how course topics are related to each other and how they fit within the context of sustainability in the geosciences. It is important to note that improvement was observed in both STEM and non-STEM majors. These findings have broad implications not only for introductory geosciences courses but for other introductory courses within the physical sciences that struggle with improving the perceived relevance of course content to a broader audience. The results of this study may be interest to programs and classes seeking to increase the diversity of students majoring in the geosciences as content relevance is linked to a student's interest in a particular discipline.

REFERENCES

Abd-El-Khalick, F. (2004). Over and Over Again: College Students' Views of Nature of Science. *Scientific Inquiry and Nature of Science*, 389–425.

Abell, S. K., & Smith, D. C. (1994). What is science? Preservice elementary teachers' conceptions of the nature of science. *International Journal of Science Education*, *16*(4), 475–487.

Adams, W. K., Wieman, C. E., Perkins, K. K., & Barbera, J. (2008). Modifying and Validating the Colorado Learning Attitudes about Science Survey for Use in Chemistry. *Journal of Chemical Education*, *85*(10), 1435.

Alsop, S., & Watts, M. (2003). Science education and affect. *International Journal of Science Education*, *25*(9), 1043–1047.

Ambrose, R., & Clement, L. (2004). Assessing Prospective Elementary School Teachers' Beliefs About Mathematics and Mathematics Learning: Rationale and Development of a Constructed-Response-. *... and Mathematics*, 56–69.

Amrein-Beardsley, A., & Popp, S. E. O. (2012). Peer observations among faculty in a college of education: Investigating the summative and formative uses of the Reformed Teaching Observation Protocol (RTOP). *Educational Assessment, Evaluation and Accountability*, *24*, 5–24.

Anderson-Cook, C. M. (2005). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. *Journal of the American Statistical Association* (Vol. 100).

Andrich, D. (1978). Relationships Between the Thurstone and Rasch Approaches to Item Scaling. *Applied Psychological Measurement*.

Avalos, B. (2011). Teacher professional development in Teaching and Teacher Education over ten years. *Teaching and Teacher Education*, *27*(1), 10–20.

Baber, L. D., Pifer, M. J., Colbeck, C., & Furman, T. (2010). Increasing Diversity in the Geosciences: Recruitment Programs and Student Self-Efficacy. *Journal of Geoscience Education*, *58*, 32.

Barak, M., & Shakhman, L. (2008). Reform-Based Science Teaching : Teachers ' Instructional Practices and Conceptions, *4*(1), 11–20.

Bell, R. L., Blair, L. M., Crawford, B. a., & Lederman, N. G. (2003). Just do it? Impact of a science apprenticeship program on high school students' understandings of the nature of science and scientific inquiry. *Journal of Research in Science Teaching*, *40*(5), 487–509.

Bezzi, A. (1999). What is this thing called geoscience? Epistemological dimensions elicited with the repertory grid and their implications for scientific literacy. *Science Education*, 675–700.

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch Model: Fundamental Measurement in the Human Sciences*. *Journal of Educational Measurement* (Vol. 2).

Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer.

Boone, W. J., Townsend, S. J., & Staver, J. (2011). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform efforts: An exemplar utilizing STEBI self-efficacy data. *Science Education*, *95*(2), 258–280.

Bralower, T. J., Feiss, P. G., & Manduca, C. A. (2008). Preparing a New Generation of Face Earth ' s Future. *Liberal Education*, *94*(2).

Breakwell, G. M., & Beardsell, S. (1992). Gender, parental and peer influences upon science attitudes and activities. *Public Understanding of Science* , *1* (2), 183–197.

Brown, J. L. (2004). *Making the most of Understanding by Design*. Washington D.C.

Burbank, M. D., & Kauchak, D. (2003). An alternative model for professional development : investigations into effective collaboration. *Teaching and Teacher Education*, *19*, 499–514.

Bybee, R., & McCrae, B. (2011). Scientific Literacy and Student Attitudes: Perspectives from PISA 2006 science. *International Journal of Science Education*, *33*(1), 7–26.

Choppin, B. H. L. (1985). Lessons for psychmetrics from thermometry. *Evaluation in Education*, *9(1)*, 9–12.

Choy, S. P., Chen, X., Bugarin, R., National Center for Educational Statistics., & Institute of Education Sciences (U.S.). (2006). Teacher professional development in 1999-2000 what teachers, principals, and district staff report. *Statistical Analysis Report*.

Cicchetti, D. V, & Sparrow, S. A. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, *86*, 127–137.

Clarke, D., & Hollingsworth, H. (2002). Elaborating a model of teacher professional growth. *Teaching and Teacher Education*, *18*(8), 947–967.

Colucci-Gray, L., Camino, E., Barbiero, G., & Gray, D. (2006). From scientific literacy to sustainability literacy: An ecological framework for education. *Science Education*, *90*(2), 227–252.

Corcoran, T. B. (1995). Transforming professional development for teachers: A guide for state policy makers. *Washington D. C. National Governor's Association*.

Corcoran, T. B., Shields, P. M., Zucker, A. A. (1998). *Evaluation of NSF's Statewide Systematic Initiatives (SSI) Program: The SSIs and professional development for teachers*. Menlo Park, CA: SRI International.

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology*, *78*(1), 98–104.

Creswell, J. W., & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. *Applied Linguistics* (Vol. 2nd).

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

D'Avanzo, C. (2013). Post-Vision and Change: Do we know how to change? *CBE Life Sciences Education*, *12*(3), 373–382.

Dall'Alba, G., & Sandberg, J. (2006). Unveiling Professional Development: A Critical Review of Stage Models. *Review of Educational Research*, *76*(3), 383–412.

Dancy, M., & Henderson, C. (2010). Pedagogical practices and instructional change of physics faculty. *American Journal of Physics*, *78*(10), 1056.

DeBoer, G. E. (1991). *A history of ideas in science education: Implications for practice*. New York, NY: Teachers College Press.

Desimone, L. M. (2009). Improving Impact Studies of Teachers ' Professional Development : Toward Better ... *Educational Researcher*, *38*, 181–199.

Ebenezer, J. V, & Zoller, U. (1993). Grade 10 students' perceptions of and attitudes toward science teaching and school science. *Journal of Research in Science Teaching*, *30*, 175–186.

Ebert-May, D., Derting, T. L., Hodder, J., Momsen, J. L., Long, T. M., & Jardeleza, S. E. (2011). What We Say Is Not What We Do: Effective Evaluation of Faculty Professional Development Programs. *BioScience*, *61*(7), 550–558.

Eddy, S. L., & Hogan, K. a. (2014). Getting Under the Hood: How and for Whom Does Increasing Course Structure Work? *Cell Biology Education*, *13*(3), 453–468.

Erickson, G. L., & Erickson, L. J. (1984). Females and science achievement: Evidence, explanations, and implications. *Science Education*, *68*(2), 63–89.

Fairweather, J. (2010). Linking Evidence and Promising Practices in STEM Undergraduate Education.

Fang, Z. (1996). A review of research on teacher beliefs and practices. *Educational Research*, *38*(1), 46–65.

Feiman-Nemser, S. (2001). From Preparation to Practice: Designing a Continuum to Strengthen and Sustain Teaching. *Teachers College Record*, *103*, 1013–1055.

Feinstein, N. (2011). Salvaging science literacy. *Science Education*.

Fishbein, M. and Ajzen, I. (1975). *Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley Publishing.

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(23), 8410–5.

Garet et al. (2001). What Makes Professional Development Effective? Results From a National Sample of Teachers. *American Educational Research Journal*, *38*(4), 915–945.

Gess-Newsome, J., Southerland, S. a., Johnston, a., & Woodbury, S. (2003). Educational Reform, Personal Practical Theories, and Dissatisfaction: The Anatomy of Change in College Science Teaching. *American Educational Research Journal*, *40*(3), 731–767.

Glaser, B. G., & Strauss, A. L. (1967). The discovery of grounded theory. *International Journal of Qualitative Methods*, *5*, 1–10.

Gregorc, A. F. (1973). Developing Plans for Professional Growth. *NASSP Bulletin*, *57*(377), 1–8.

Guskey, T. (1986). Staff Development and the Process of Teacher Change. *Educational Researcher*, *15*(5), 5–12.

Guskey, T. (2002). Professional development and teacher change. *Teachers and Teaching: Theory and Practice*, *8*(3).

Haak, D. C., HilleRisLambers, J., Pitre, E., & Freeman, S. (2011). Increased structure and active learning reduce the achievement gap in introductory biology. *Science (New York, N.Y.)*, *332*(6034), 1213–1216.

Hake, R. R. (1998). Interactive-engagement versus traditional methods : A six-thousand-student survey of mechanics test data for introductory physics courses, (May 1996), 64–74.

Haladyna, T., Olsen, R., & Shaughnessy, J. (1982). Relations of student, teacher, and learning environment variables to attitudes toward science. *Science Education*, *66*(5), 671–687.

Handelsman, J., Ebert-may, D., Beichner, R., Bruns, P., Chang, A., Dehaan, R., … Wood, W. B. (2004). Scientific Teaching. *Science*, *304*, 521–522.

Harding, J. (1983). Switched off: The science education of girls. *London Longman for Schools Council*.

Hargreaves, A., & Dawe, R. (1990). Paths of professional development: Contrived collegiality, collaborative culture, and the case of peer coaching. *Teaching and Teacher Education*.

Harvey, T. J., & Edwards, P. (1980). Childrens expecations and realisations of science. *British Journal of Educational Psychology*, *50*(1), 74–76.

Hashweh, M. Z. (1996). Effects of science teachers' epistemological beliefs in teaching. *Journal of Research in Science Teaching*, *33*(1), 47–63.

Hay, J., & Mimura, N. (2006). Supporting climate change vulnerability and adaptation assessments in the Asia-Pacific region: an example of sustainability science. *Sustainability Science*, *1*(1), 23–35.

Henderson, C., Beach, A., & Finkelstein, N. (2011). Facilitating change in undergraduate STEM instructional practices: An analytic review of the literature. *Journal of Research in Science Teaching*, *48*(8), 952–984.

Henderson, C., & Dancy, M. (2007). Barriers to the use of research-based instructional strategies: The influence of both individual and situational characteristics. *Physical Review Special Topics - Physics Education Research*, *3*(2), 020102.

Henderson, C., Dancy, M., & Niewiadomska-Bugaj, M. (2012). Use of research-based instructional strategies in introductory physics: Where do faculty leave the innovation-decision process? *Physical Review Special Topics - Physics Education Research*, *8*(2), 020104.

Hendley, D., Stables, A., Parkinson, J., & Tanner, H. (1996). Pupils' attitudes to technology in key stage 3 of the national curriculum: A study of pupils in South Wales. *International Journal of Technology and Design Education*, *6*(1), 15–29.

Hoaglin, D. C., & Iglewicz, B. (1987). Fine-Tuning Some Resistant Rules for Outlier Labeling. *Journal of the American Statistical Association*, *82*(400), 1147–1149.

Holbrook, J., & Rannikmae, M. (2009). The Meaning of Scientific Literacy. ... *Journal of Environmental and Science Education*, *4*(3), 275–288.

Hora, M. T. (2012). Organizational Factors and Instructional Decision-Making: A Cognitive Perspective. *The Review of Higher Education*, *35*(2), 207–235.

Hora, M. T., Oleson, A., & Ferrare, J. J. (2013). Teaching dimensions observation protocol (TDOP) user's manual. *Tdop.Wceruw.Org*, 28.

Huberman, M. (1989). The professional life cycle of teachers. *The Teachers College Record*, *91*, 31–57.

Huntoon, J. E., & Lane, M. J. (2008). Diversity in the geosciences and successful strategies for increasing diversity. *Journal of Geoscience Education*, *55*, 447–457.

Huntoon, J. E., Peach, C., & Hopkins, J. (2005). *Geoscience Education and Diversity : Vision for the Future and Strategies for Success Report of the Second Geoscience Education*. Washington D.C.

Hurd, P. D. (1958). Science Literacy: Its meaning for American Schools. *Educational Leadership*, *16(1)*, 13–16.

Jerneck, A., Olsson, L., Ness, B., Anderberg, S., Baier, M., Clark, E., … Persson, J. (2011). Structuring sustainability science. *Sustainability Science*, *6*(1), 69–82.

Johnson, R. B., & Onwuegbuzie, A. J. (2004). Mixed Methods Research: A Research Paradigm Whose Time Has Come. *Educational Researcher*, *33*(7), 14–26.

Johnson, S. (1987). Gender differences in science: parallels in interest, experience and performance. *International Journal of Science Education*, *9*(4), 467–481.

Jones, M. G., Carter, G. (2007). Science teacher attitudes and beliefs. In *Handbook of research on science education* (pp. 1067–1104).

Jovanovic, J., & King, S. S. (1998). Boys and Girls in the Performance-Based Science Classroom: Who's Doing the Performing? *American Educational Research Journal*, *35*(3), 477–496.

Kagan, D. (1992). Professional growth among preservice and beginning teachers. *Review of Educational Research*, *62*(2), 129–169.

Kahle, J. B., & Lakes, M. K. (1983). The myth of equality in science classrooms. *Journal of Research in Science Teaching*, *20*(2), 131–140.

Kajikawa, Y. (2008). Research core and framework of sustainability science. *Sustainability Science*, *3*(2), 215–239.

Kang, N.-H., & Wallace, C. S. (2005). Secondary science teachers' use of laboratory activities: Linking epistemological beliefs, goals, and practices. *Science Education*, *89*(1), 140–165.

Kastenhofer, K., & Rammel, C. (2005). Obstacles to and potentials of the societal implementation of sustainable development: a comparative analysis of two case studies. *Sustainability: Science, Practice, and Policy*, *1*, 5–13.

Keller, J. (1983). Motivational design of instruction. In *Instructional -design theories and models* (pp. 386–434).

Keys, C., & Bryan, L. (2001). Co-constructing inquiry-based science with teachers: Essential research for lasting reform. *Journal of Research in Science Teaching*, *38*(6), 631–645.

Koballa, T. R. J., & Glynn, S. M. (2007). Attitudinal and Motivational Constructs in Science Learning. In *Handbook of research in science education* (pp. 75–102).

Kober, N. (2015). *Reaching Students: What Research Says About Effective Instruction in Undergraduate Science and Engineering*. National Academies Press.

Laugksch, R. C. (2000). Scientific Literacy : A Conceptual Overview. *Science Education*, *84*(1), 71–94.

Levine, R., Gonzalez, R., Cole, S., Fuhrman, M., & Floch, K. C. L. (2008). The geoscience pipeline: a conceptual framework. *Journal of Geoscience Education*, *55*, 458–468.

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

Lichtenstein, M. J., Owen, S. V., Blalock, C. L., Liu, Y., Ramirez, K. A., Pruski, L. A., … Toepperwein, M. A. (2008). Psychometric reevaluation of the scientific attitude inventory-revised (SAI-II). *Journal of Research in Science Teaching*, *45*(5), 600–616.

Linacre, J. M. (1998). Detecting multidimensionality: Which residual data-type works best? *Journal of Outcome Measurement*, *2(3)*, 266–283.

Linacre, J. M. (2006). Rasch analysis of rank-ordered data. *Journal of Applied Measurement*, *7*, 129–139.

Linacre, J. M. (2010). Predicting responses from rasch measures. *Journal of Applied Measurement*.

Linacre, J. M., & Wright, B. D. (2000). *Winsteps [computer software]*. Chicago: Mesa Press.

Loucks-Horsley, S., & Matsumoto, C. (1999). Research on professional development for teachers of mathematics and science: The state of the scene. *School Science and …*, *99*, 258–271.

Luft, J., & Roehrig, G. (2007). Capturing science teachers' epistemological beliefs: The development of the teacher beliefs interview. *Electronic Journal of Science …*, *11*(2).

Lukes, L. a., & McConnell, D. a. (2014). What Motivates Introductory Geology Students to Study for an Exam? *Journal of Geoscience Education*, *62*, 725–735.

MacIsaac, D., & Falconer, K. (2002). Reforming Physics Instruction Via RTOP. *The Physics Teacher*, *40*(November), 479–485.

Mao, J.-Y., Vredenburg, K., Smith, P. W., & Carey, T. (2005). The state of user-centered design practice. *Communications of the ACM*.

McComas, W. F., Almazroa, H., & Clough, M. P. (1998). The Nature of Science in Science Education: An Introduction. *Science and Education*, *7*(6), 511–532.

McMillan, J. H., & May, M. J. (1979). A study of factors influencing attitudes toward science of junior high school students. *Journal of Research in Science Teaching*, *16*(3), 217–222.

Miller, J. D. (2004). Public Understanding of, and Attitudes toward, Scientific Research: What We Know and What We Need to Know. *Public Understanding of Science*, *13*(3), 273–294.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. a. (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science (New York, N.Y.)*, *330*(6008), 1234–7.

Moore, R. W., & Foy, R. L. H. (1997). The scientific attitude inventory: A revision (SAI II). *Journal of Research in Science Teaching*, *34*(4), 327–336.

Moore, Richard W., Suntman, F. X. (1970). The development , field test, and validation of an inventory of science attitudes. *Jounrnal of Research in Science Teaching*, *7*, 85–94.

Munby, H. (1982). The place of teachers' beliefs in research on teacher thinking and decision making, and an alternative methodology. *Instructional Science*, *11*(3), 201–225.

Munby, H. (1983). Scientific Attitude Inventory: What confidence can we have in this instrument?, *20*(2), 141–162.

Myers, R. E., & Fouts, J. T. (1992). A cluster analysis of high school science classroom environments and attitude toward science. *Journal of Research in Science Teaching*, *29*(9), 929–937.

Nadelson, L. S., & Viskupic, K. (2010). Perceptions of the Nature of Science by Geoscience Students Experiencing Two Different Courses of Study. *Journal of Geoscience Education*, *58*(5), 275.

Nespor, J. (1987). The Role of Beliefs in the Practive of Teaching. *Journal of Curriculum Studies*, *19*, 317–328.

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079.

Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, *62*, 307–332.

Papanastasiou, E. C., & Zembylas, M. (2004). Differential effects of science attitudes and science achievement in Australia, Cyprus, and the USA. *International Journal of Science Education*.

Parise, L. M., & Spillane, J. P. (2010). Teacher Learning and Instructional Change: How Formal and On-the-Job Learning Opportunities Predict Change in Elementary School Teachers' Practice. *The Elementary School Journal*.

Parris, T. M., & Kates, R. W. (2003). Characterizing a sustainability transition: goals, targets, trends, and driving forces. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(14), 8068–8073.

Penuel, W. R., Fishman, B. J., Yamaguchi, R., & Gallagher, L. P. (2007). What Makes Professional Development Effective? Strategies That Foster Curriculum Implementation. *American Educational Research Journal*, *44*(4), 921–958.

Piburn, M. D., & Baker, D. R. (1993). If I were the teacher … qualitative study of attitude toward science. *Science Education*, *77*(4), 393–406.

Piburn, M., Sawada, D., & Turley, J. (2000). Reformed teaching observation protocol (RTOP) reference manual. *... of Teachers*, 1–41.

Putnam, R. T., & Borko, H. (2000). What Do New Views of Knowledge and Thinking Have to Say About Research on Teacher Learning? *Educational Researcher*.

Richardson, V. (1996). The role of attitudes and beliefs in learning to teach. In *The Handbook of Research in Teacher Education* (pp. 102–119). New York: Macmillan.

Richter, D., Kunter, M., Klusmann, U., Lüdtke, O., & Baumert, J. (2011). Professional development across the teaching career: Teachers' uptake of formal and informal learning opportunities. *Teaching and Teacher Education*, *27*, 116–126.

Roberts, D. A. (1983). *Science Literacy: Towards Balance in Setting Goals for School Science Programs.* Ottawa, Ontario: Publications Office, Science Council of Canada.

Robertson, I. J. (1987). Girls and boys and practical science. *International Journal of Science Education*, *9*(5), 505–518.

Rokeach, M. (1968). Beliefs, attitudes, and values: A theory of organization and change.

Ryder, J., Leach, J., & Driver, R. (1997). Undergraduate science students' omages of the nature of science. *Symposium "New Perspectives on Conceptual Change in Science and Mathematics Learning", American Educational Research Association Annual Conference, 24th-28th March 1997*.

Ryker, K., Mcconnell, D. A., Bruckner, M. A., & Manduca, C. A. (2013). Teaching is Believing: Pedagogical Beliefs, Practices, and Professional Development. In *GSA Annual Meeting in Denver*.

Sadler, T. D. (2004). Informal reasoning regarding socioscientific issues: A critical review of research. *Journal of Research in Science Teaching*, *41*(5), 513–536.

Sadler, T. D. (2014). Assessing Schools for Generation R (Responsibility), *41*, 101–113.

Sadler, T. D., & Zeidler, D. L. (2005). The significance of content knowledge for informal reasoning regarding socioscientific issues: Applying genetics knowledge to genetic engineering issues. *Science Education*, *89*(1), 71–93.

Sampson, V., & Grooms, J. (2013). Development and Initial Validation of the Beliefs About Reformed Science Teaching and Learning (BARSTL) Questionnaire. *School Science and Mathematics*, *113*, 3–15.

Sawada, D., Piburn, M. D., Judson, E., Turley, J., Falconer, K., Benford, R., & Bloom, I. (2002). Measuring Reform Practices in Science and Mathematics Classrooms : The Reformed Teaching Observation Protocol, *102*(October).

Schuler, D., & Namioka, A. (1993). Participatory Design: Principles and Practices. *System*, 319.

Shadish, W.R., Cook, T.D., & Campbell, D. T. (2002). Experimental and Quasi-Experimental Designs. In *Experimental and Quasi-Experimental Designs for Generalized Causal Inference.* (pp. 171–206).

Shen, B. S. P. (1975). Science Literacy and the Public Understanding of Science. In *Communication of Scientific Information* (pp. 44–52). S. Karger AG.

Shen, B. S. P. (2014). Sigma Xi , The Scientific Research Society developing and industrialized countries alike Benjamin s . p . Shen Science Literacy Views Public understanding of science is becoming vitally needed in developing and industrialized countries alike. *American Scientist*, *63*(3), 265–268.

Siegel, M. a., & Ranney, M. a. (2003). Developing the changes in attitude about the relevance of science (CARS) questionnaire and assessing two high school science classes. *Journal of Research in Science Teaching*, *40*(8), 757–775.

Sikes, P. J., Measor, L., & Woods, P. (1985). *Teacher careers: crisis and continuity*.

Simpson, R. D., & Oliver, J. S. (1985). Attitude toward science and achievement motivation profiles of male and female science students in grades six through ten. *Science Education*, *69*(4), 511–525.

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012a). *Discipline-Based Education Research : Understanding and Improving Learning in Undergraduate Science and Engineering*. *Social Sciences*.

Singer, S. R., Nielsen, N. R., & Schweingruber, H. A. (2012b). *Understanding and Improving Learning in*.

Singer, S., & Smith, K. a. (2013). Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering. *Journal of Engineering Education*, *102*(4), 468–471.

Smail, B., & Kelly, A. (1984). Sex Differences in Science and Technology Among 11-year-old Schoolchildren: II--affective. *Research in Science & Technological Education*, *2*(2), 87–106.

Smith, M. K., Vinson, E. L., Smith, J. a., Lewin, J. D., & Stetzer, M. R. (2014). A Campus-Wide Study of STEM Courses: New Perspectives on Teaching Practices and Perceptions. *Cell Biology Education*, *13*(4), 624–635.

Smith, R. M. (1991). The Distributional Properties of Rasch Item Fit Statistics. *Educational and Psychological Measurement*.

Smylie, M. A. (1995). Teacher learning in the workplace: Implications for school reform. In *Professional development in education: new paradigms and practices* (pp. 92–113).

Snyder, T. D., & Dillow, S. A. (2012). Digest of education statistics 2011. *National Center for Education, Institute of Education Sciences. United States Department of Education Statistics*, (NCES 2012-001), Chapter 2.

Sunal, D. W., Hodges, J., Sunal, C. S., & Whitaker, K. W. (2001). Teaching science in higher education : Faculty professional development and b ..., (May).

Sundberg, M. D., Dini, M. L., & Li, E. (1994). Decreasing course content improves student comprehension of science and attitudes towards science in freshman biology. *Journal of Research in Science Teaching*, *31*(6), 679–693.

Talton, E. L., & Simpson, R. D. (1985). Relationships between peer and individual attitudes toward science among adolescent students. *Science Education*, *69*(1), 19–24.

Tewksbury, B. J., Manduca, C. A., Mogk, D. W., & Macdonald, R. H. (2013). Geoscience education for the Anthropocene. *The Geological Society of America Special Paper*, *2501*(08), 189–201.

Tong, V. (2012). Geoscience Research and Education.

Tuan, H.-L., Chin, C.-C., & Shieh, S.-H. (2005). The Development of a Questionnaire to Measure Students' Motivation Towards Science Learning. Research Report. *International Journal of Science Education*, *27*(6), 639–654.

Unruh, A., & Turner, H. (1970). *Supervision of change and innovation*. Boston: Houghton Mifflen.

van der Leeuw, S., Wiek, A., Harlow, J., & Buizer, J. (2012). How much time do we have? Urgency and rhetoric in sustainability science. *Sustainability Science*, *7*(S1), 115–120.

Veenman, S. (1984). Perceived Problems of Beginning Teachers. *Review of Educational Research*, *54*(2), 143–178.

Wieman, C., Perkins, K., & Gilbert, S. (2010). Transforming Science Education at Large Research Universities: A Case Study in Progressxs. *Change: The Magazine of Higher Learning*, *42*(2), 6–14.

Wilson, M., & Adams, R. J. (1995). Rasch models for item bundles. *Psychometrika*, *60*, 181–198.

Wilson., M., & Draney, K. (2002). A technique for setting standards and maintaining them over time. *Measurement and Multivariate Analysis*, 325–332.

Wright, B. D. (1977). Solving Measurement Problems With the Rasch Model. *Journal of Educational Measurement*, *14*, pp. 97–116.

Wright, B. D. (1984). Despair and hope for educational measurement. *Contemporary Education Review*, *3(1)*, 281–288.

Wright, B. D., & Stone, M. H. (1979). *Best test design*. Chicago: Mesa Press.

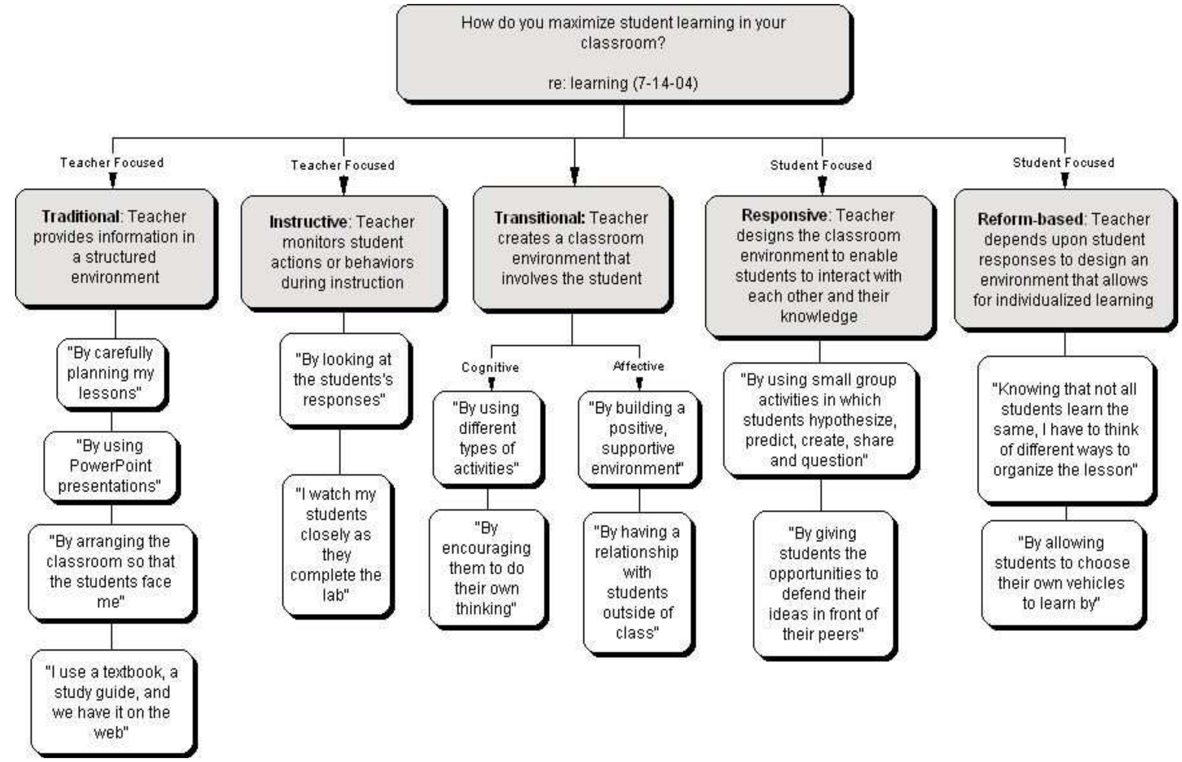
Wysession, M. E., LaDue, N., Budd, D. a., Campbell, K., Conklin, M., Kappel, E., … Tuddenham, P. (2012). Developing and Applying a Set of Earth Science Literacy Principles. *Journal of Geoscience Education*, *60*, 95–99.

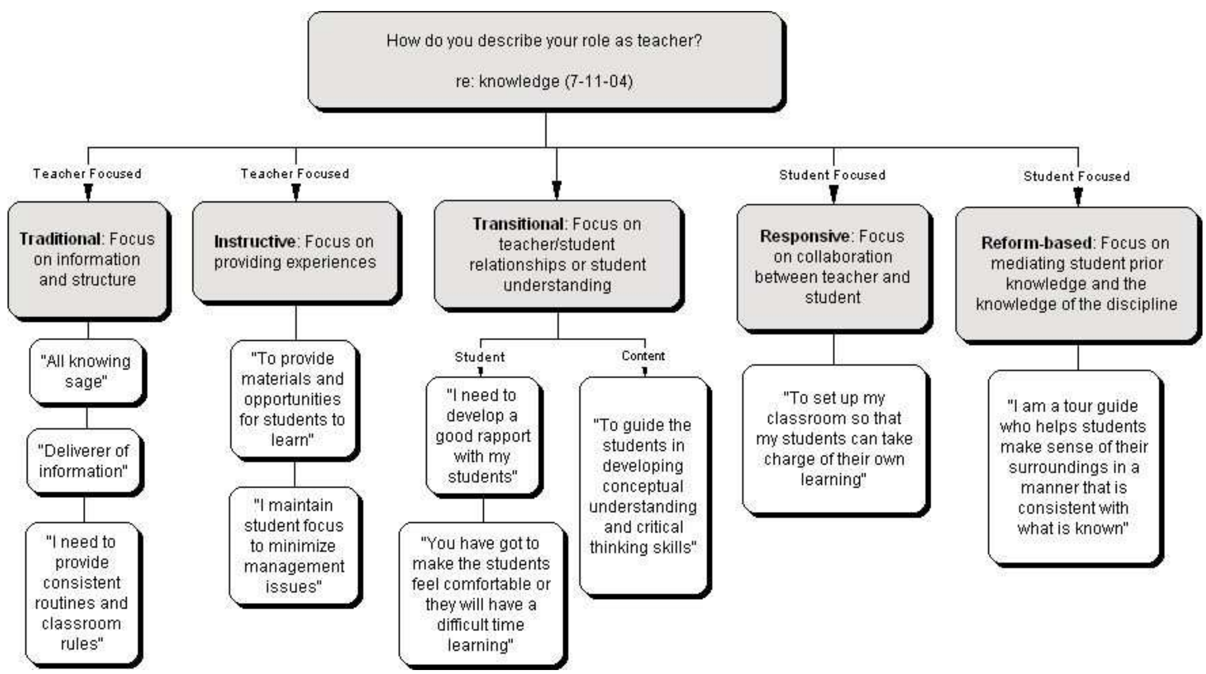
Zeidler, D. L., Osborne, J., Erduran, S., Simon, S., & Monk, M. (2003). The Role of Argument During Discourse About Socioscientific Issues. In *The Role of Moral Reasoning on Socioscientific Issues and Discourse in Science Education* (pp. 97–116).

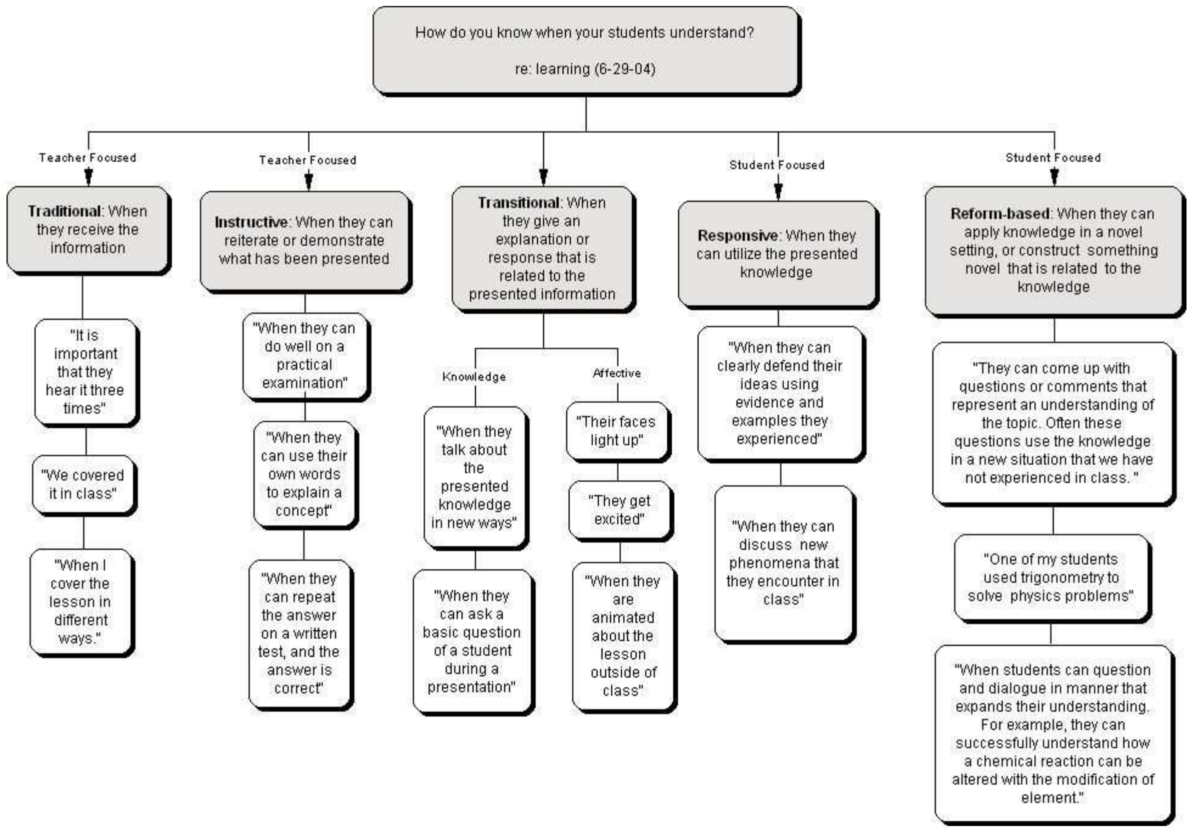
Zo`bi, A. S. (2014). The Effect of Using Socio-Scientific Issues Approach in Teaching Environmental Issues on Improving the Students' Ability of Making Appropriate Decisions Towards These Issues. *International Education Studies*, *7*(8), 113–123.

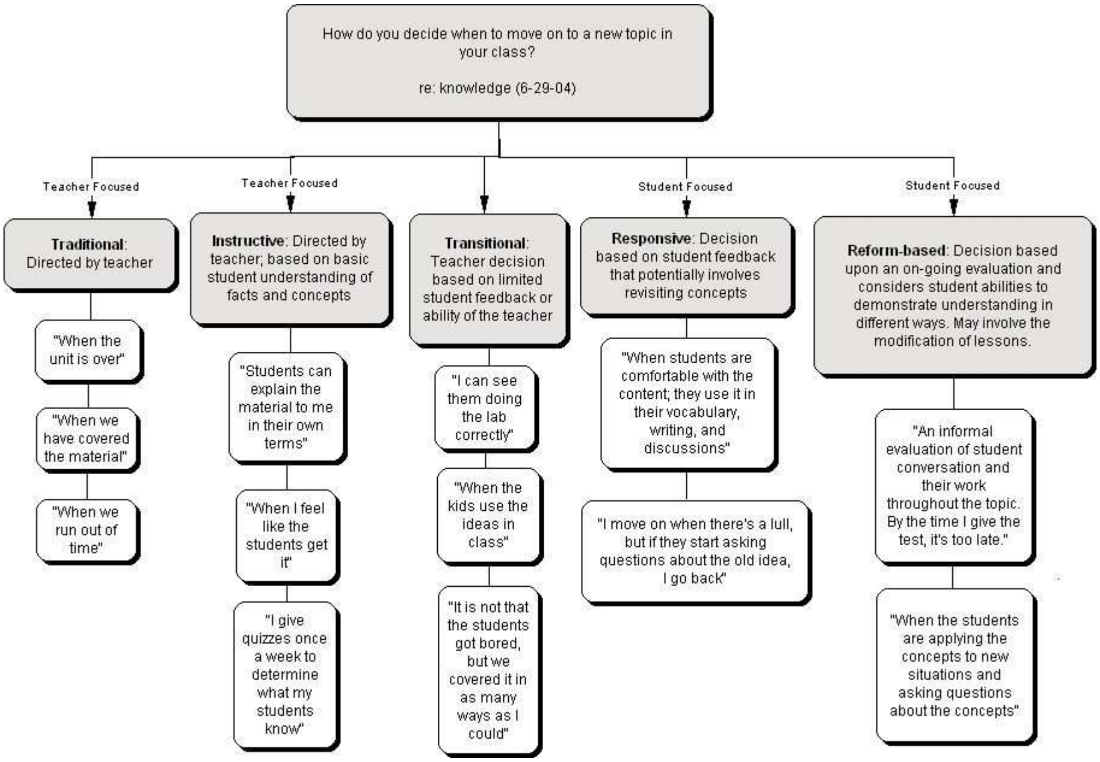
APPENDICES

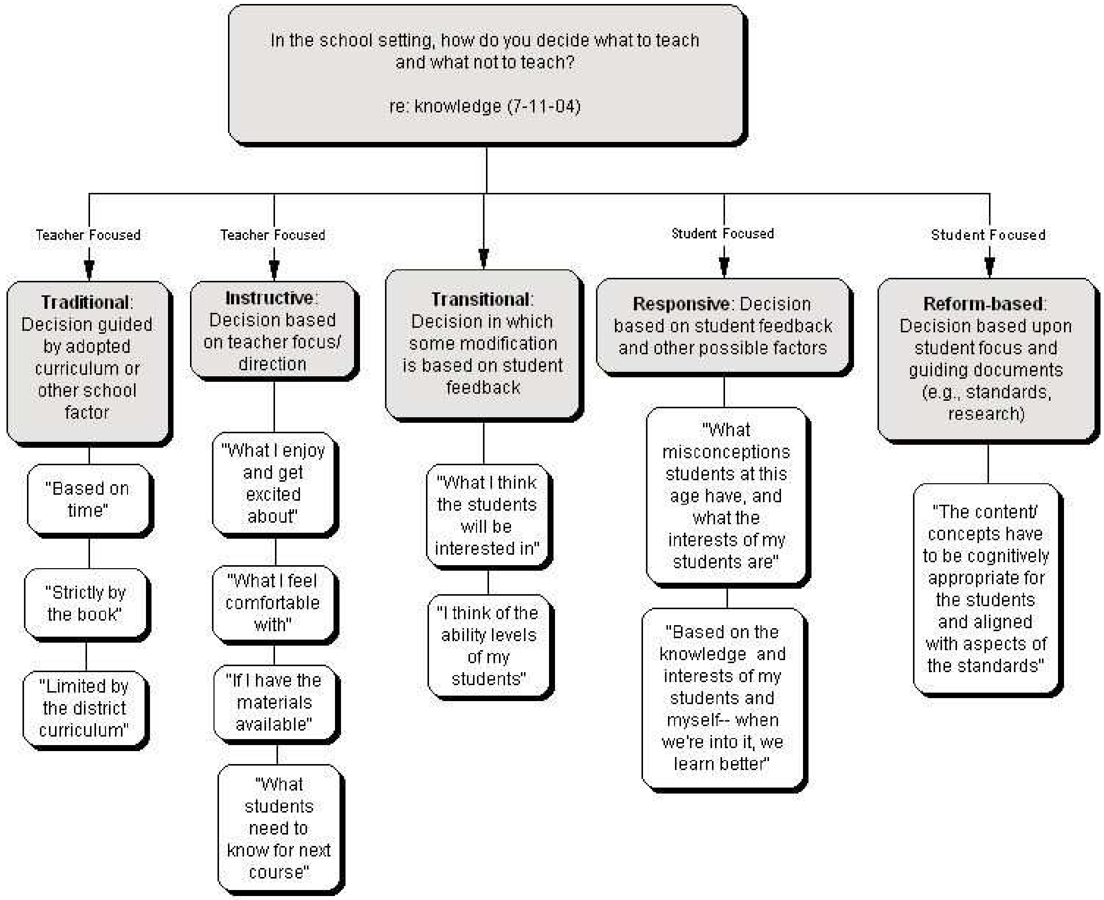
Appendix 1: Coding protocol for all seven questions of the TBI.

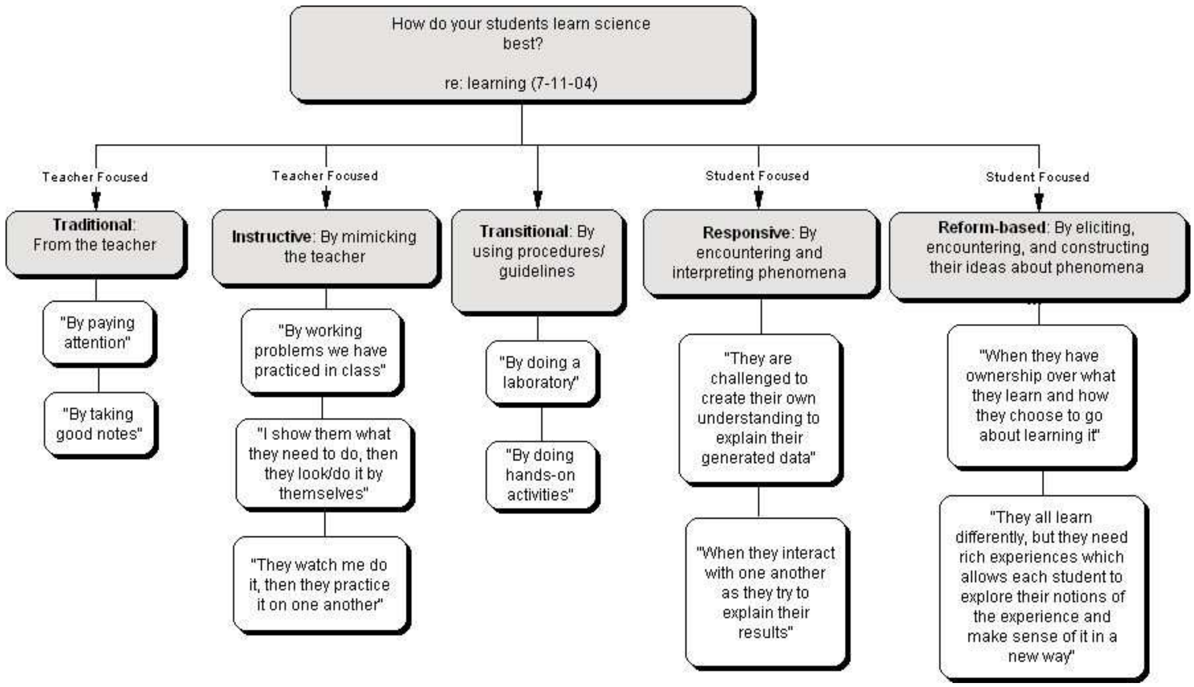


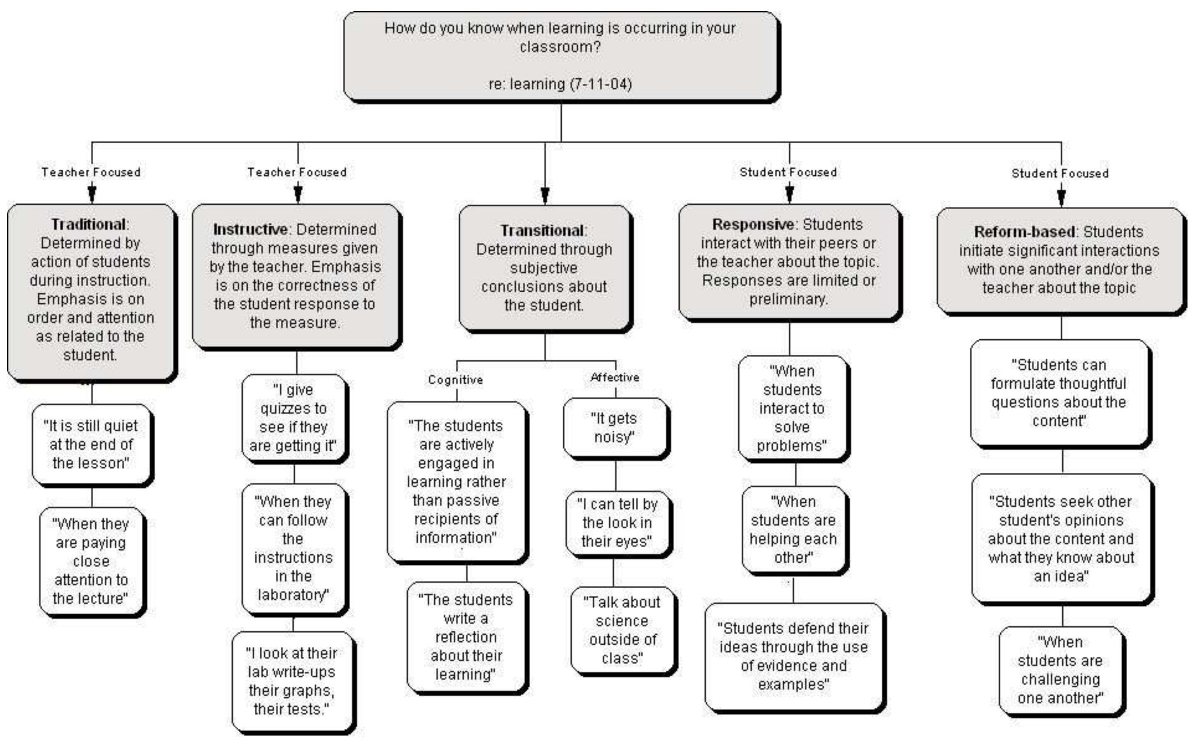



How do you know when your students understand?
re: learning (6-29-04)
Teacher Focused
Traditional: When they receive the information
"It is important that they hear it three times"
"We covered it in class"
"When I cover the lesson in different ways."
Teacher Focused
Instructive: When they can reiterate or demonstrate what has been presented
"When they can do well on a practical examination"
"When they can use their own words to explain a concept"
"When they can repeat the answer on a written test, and the answer is correct"
Transitional: When they give an explanation or response that is related to the presented information
Knowledge
"When they talk about the presented knowledge in new ways"
"When they can ask a basic question of a student during a presentation"
Affective
"Their faces light up"
"They get excited"
"When they are animated about the lesson outside of class"
Student Focused
Responsive: When they can utilize the presented knowledge
"When they can clearly defend their ideas using evidence and examples they experienced"
"When they can discuss new phenomena that they encounter in class"
Student Focused
Reform-based: When they can apply knowledge in a novel setting, or construct something novel that is related to the knowledge
"They can come up with questions or comments that represent an understanding of the topic. Often these questions use the knowledge in a new situation that we have not experienced in class."
"One of my students used trigonometry to solve physics problems"
"When students can question and dialogue in manner that expands their understanding. For example, they can successfully understand how a chemical reaction can be altered with the modification of element."


How do you decide when to move on to a new topic in your class?
re: knowledge (6-29-04)
Teacher Focused
Teacher Focused
Student Focused
Student Focused
Traditional: Directed by teacher
"When the unit is over"
"When we have covered the material"
"When we run out of time"
Instructive: Directed by teacher; based on basic student understanding of facts and concepts
"Students can explain the material to me in their own terms"
"When I feel like the students get it"
"I give quizzes once a week to determine what my students know"
Transitional: Teacher decision based on limited student feedback or ability of the teacher
"I can see them doing the lab correctly"
"When the kids use the ideas in class"
"It is not that the students got bored, but we covered it in as many ways as I could"
Responsive: Decision based on student feedback that potentially involves revisiting concepts
"When students are comfortable with the content; they use it in their vocabulary, writing, and discussions"
"I move on when there's a lull, but if they start asking questions about the old idea, I go back"
Reform-based: Decision based upon an on-going evaluation and considers student abilities to demonstrate understanding in different ways. May involve the modification of lessons.
"An informal evaluation of student conversation and their work throughout the topic. By the time I give the test, it's too late."
"When the students are applying the concepts to new situations and asking questions about the concepts"


In the school setting, how do you decide what to teach and what not to teach?
re: knowledge (7-11-04)
Teacher Focused
Teacher Focused
Student Focused
Student Focused
Traditional: Decision guided by adopted curriculum or other school factor
Instructive: Decision based on teacher focus/ direction
Transitional: Decision in which some modification is based on student feedback
Responsive: Decision based on student feedback and other possible factors
Reform-based: Decision based upon student focus and guiding documents (e.g., standards, research)
"Based on time"
"Strictly by the book"
"Limited by the district curriculum"
"What I enjoy and get excited about"
"What I feel comfortable with"
"If I have the materials available"
"What students need to know for next course"
"What I think the students will be interested in"
"I think of the ability levels of my students"
"What misconceptions students at this age have, and what the interests of my students are"
"Based on the knowledge and interests of my students and myself-- when we're into it, we learn better"
"The content/ concepts have to be cognitively appropriate for the students and aligned with aspects of the standards"


How do your students learn science best?
re: learning (7-11-04)
Teacher Focused
Teacher Focused
Student Focused
Student Focused
Traditional: From the teacher
"By paying attention"
"By taking good notes"
Instructive: By mimicking the teacher
"By working problems we have practiced in class"
"I show them what they need to do, then they look/do it by themselves"
"They watch me do it, then they practice it on one another"
Transitional: By using procedures/ guidelines
"By doing a laboratory"
"By doing hands-on activities"
Responsive: By encountering and interpreting phenomena
"They are challenged to create their own understanding to explain their generated data"
"When they interact with one another as they try to explain their results"
Reform-based: By eliciting, encountering, and constructing their ideas about phenomena
"When they have ownership over what they learn and how they choose to go about learning it"
"They all learn differently, but they need rich experiences which allows each student to explore their notions of the experience and make sense of it in a new way"


How do you know when learning is occurring in your classroom?
re: learning (7-11-04)
Teacher Focused
Teacher Focused
Student Focused
Student Focused
Traditional: Determined by action of students during instruction. Emphasis is on order and attention as related to the student.
"It is still quiet at the end of the lesson"
"When they are paying close attention to the lecture"
Instructive: Determined through measures given by the teacher. Emphasis is on the correctness of the student response to the measure.
"I give quizzes to see if they are getting it"
"When they can follow the instructions in the laboratory"
"I look at their lab write-ups their graphs, their tests."
Transitional: Determined through subjective conclusions about the student.
Cognitive
Affective
"The students are actively engaged in learning rather than passive recipients of information"
"The students write a reflection about their learning"
"It gets noisy"
"I can tell by the look in their eyes"
"Talk about science outside of class"
Responsive: Students interact with their peers or the teacher about the topic. Responses are limited or preliminary.
"When students interact to solve problems"
"When students are helping each other"
"Students defend their ideas through the use of evidence and examples"
Reform-based: Students initiate significant interactions with one another and/or the teacher about the topic
"Students can formulate thoughtful questions about the content"
"Students seek other student's opinions about the content and what they know about an idea"
"When students are challenging one another"

Appendix 2: BARSTL Survey Items

1. Students develop many beliefs about how the world works before they ever study about science in school.
2. Students learn in a in a disorderly fashion; they create their own knowledge by modifying their existing ideas in an effort to make sense of new and past experiences
3. People are either talented at science or they are not, therefore student achievement in science is a reflection of their natural abilities.
4. Students are more likely to understand a scientific concept if the teacher explains the concept in a way that is clear and easy to understand.
5. Frequently, students have difficulty learning scientific concepts in school because their beliefs about how the world works are often resistant to change.
6. Learning science is an orderly process; students learn by gradually accumulating more information about a topic over time
7. Students know very little about science before they learn it in school.
8. Students learn the most when they are able to test, discuss, and debate many possible answers during activities that involve social interaction.
9. During a lesson, students should explore and conduct their own experiments with hands-on materials before the teacher discusses any scientific concepts with them.
10. During a lesson, teachers should spend more time asking questions that trigger divergent ways of thinking than they do explaining the concept to students.
11. Whenever students conduct an experiment during a science lesson, the teacher should give step-by-step instructions for the students to follow in order to prevent confusion and to make sure students get the correct results.
12. Experiments should be included in lessons as a way to reinforce the scientific concepts students have already learned in class.
13. Lessons should be designed in a way that allows students to learn new concepts through inquiry instead of through a lecture, a reading or a demonstration.
14. During a lesson, students need to be given opportunities to test, debate and challenge ideas with their peers.
15. During a lesson, all of the students in the class should be encouraged to use the same approach for conducting an experiment or solving a problem.
16. Assessments in science classes should only be given after instruction is completed; that way the teacher can determine if the students have learned the material covered in class.
17. Students should do most of the talking in science classrooms.
18. Students should work independently as much as possible so they do not learn to rely on other students to do their work for them.
19. In science classrooms, students should be encouraged to challenge ideas while maintaining a climate of respect for what others have to say.
20. Teachers should allow students to help determine the direction and the focus of a lesson.
21. Students should be willing to accept the scientific ideas and theories presented to them during science class without question.
22. An excellent science teacher is someone who is really good at explaining complicated concepts clearly and simply so that everyone understands.
23. The teacher should motivate students to finish their work as quickly as possible.
24. Science teachers should primarily act as a resource person; working to support and enhance student investigations rather than explaining how things work.
25. A good science curriculum should focus on only a few scientific concepts a year, but in great detail.
26. The science curriculum should focus on the basic facts and skills of science that students will need to know later.
27. Students should know that scientific knowledge is discovered using the scientific method.

28. The science curriculum should encourage students to learn and value alternative modes of investigation or problem solving.
29. In order to prepare students for future classes, college, or a career in science the science curriculum should cover as many different topics as possible over the course of a school year.
30. The science curriculum should help students develop the reasoning skills and habits of mind necessary to do science.
31. Students should learn that all science is based on a single scientific method—a step-by-step procedure that begins with ‘define the problem’ and ends with ‘reporting the results.’
32. A good science curriculum should focus on the history and nature of science and how science affects people and societies.

Appendix 3: SAI II Survey

WHAT IS YOUR ATTITUDE TOWARD SCIENCE?
SAI II

There are some statements about science on the next three pages. Some statements are about the nature of science. Some are about how scientists work. Some of these statements describe how you might feel about science.

You may agree with some of the statements and you may disagree with others. That is exactly what you are asked to do. By doing this, you will show your attitudes toward science.

After you have carefully read a statement, decide whether or not you agree with it. If you agree, decide whether you agree mildly or strongly. If you disagree, decide whether you disagree mildly or strongly. You may decide that you are uncertain or cannot decide. Then, find the number of that statement on the answer sheet, and **CIRCLE** the:

A if you agree strongly
B if you agree mildly
C if you are uncertain or cannot decide
D if you disagree mildly
E if you disagree strongly

EXAMPLE:

I would like to have a lot of money.

A B C D E

The person who circled this example agrees strongly with the statement, "I would like to have a lot of money."

Please respond to each statement and circle only ONE letter for each statement.

1. I would enjoy studying science.

A B C D E

2. Anything we need to know can be found out through science.

A B C D E

3. It is useless to listen to a new idea unless everybody agrees with it.

A B C D E

4. Scientists are always interested in better explanations of things.

A B C D E

5. If one scientist says an idea is true, all other scientists will believe it.

A B C D E

6. Only highly trained scientists can understand science

A B C D E

7. We can always get answers to questions by asking a scientist.

A B C D E

8. Most people are not able to understand science.

A B C D E

9. Electronics are examples of the really valuable products of science.

A B C D E

10. Scientists cannot always find the answers to their questions.
A B C D E

11. When scientists have a good explanation, they do not try to make it better.
A B C D E

12. Most people can understand science.
A B C D E

13. The search for scientific knowledge would be boring.
A B C D E

14. Scientific work would be too hard for me.
A B C D E

15. Scientist discover laws which tell us exactly what is going in nature
A B C D E

16. Scientific idea can be changed.
A B C D E

17. Scientific questions are answered by observing things.
A B C D E

18. Good scientists are willing to change their ideas.
A B C D E

19. Some questions cannot be answered by science.
A B C D E

20. A scientist must have a good imagination to create new ideas.
A B C D E

21. Ideas are the important result of science.
A B C D E

22. I do not want to be a scientist.
A B C D E

23. People must understand science because it affects their lives.
A B C D E

24. A major purpose of science is to produce new drugs and save lives.
A B C D E

25. Scientists must report exactly what they observe.
A B C D E

26. If a scientist cannot answer a question, another scientist can.
A B C D E

27. I would like to work with other scientists to solve scientific problems.
A B C D E

28. Science tries to explain how things happen.
A B C D E

29. Every citizen should understand science.
A B C D E

30. I may not make great discoveries, but working in science would be fun.
A B C D E

31. A major purpose of science is to help people live better
A B C D E

32. Scientists should not criticize each other's work.
A B C D E

33. The senses are one of the most important tools a scientist has.
A B C D E

34. Scientists believe that nothing is known to be true for sure.
A B C D E

35. Scientific laws have been proven beyond all possible doubt.
A B C D E

36. I would like to be a scientist.
A B C D E

37. Scientists do not have enough time for families or for fun.
A B C D E

38. Scientific work is useful only to scientists.
A B C D E

39. Scientists have to study too much.
A B C D E

40. Working in a science laboratory would be fun.
A B C D E

Appendix 4: CARS A

Relevance of Science Questionnaire
Please read the below statements and indicate the degree to which you agree or disagree by circling 0-4.

0 = Strongly Disagree (SD) **1** = Disagree (D) **2** = Neutral(N) **3** = Agree (A) **4** = Strongly Agree (SA)

	SD	D	N	A	SA
1. My parents encourage me to continue with science.	0	1	2	3	4
2. I plan to take more science classes in college.	0	1	2	3	4
3. Science helps me to work with others to find answers.	0	1	2	3	4
4. Science class helps me to evaluate my own work.	0	1	2	3	4
5. Learning science helps me understand about the environment.	0	1	2	3	4
6. Emotion has no place in science.	0	1	2	3	4
7. Science class helps me to judge other people's points of view.	0	1	2	3	4
8. Science will help me understand more about worldwide problems.	0	1	2	3	4
9. Science has nothing to do with my life outside of school.	0	1	2	3	4
10. Experiments in science help me to learn with a group.	0	1	2	3	4
11. Science teaches me to help others make decisions.	0	1	2	3	4
12. Knowing science will not help me in sports.	0	1	2	3	4
13. Science has nothing to do with buying things, such as computers and cars.	0	1	2	3	4
14. Knowledge of science could make it easier to fix a bicycle.	0	1	2	3	4
15. Science teaches me to think less clearly than I already do.	0	1	2	3	4
16. Making a good decision is a scientific process.	0	1	2	3	4
17. Geology class will help prepare me for a career.	0	1	2	3	4
18. Much of what I learn in Geology class is useful in my everyday life today.	0	1	2	3	4
19. Learning science can help me when I choose where to buy a house.	0	1	2	3	4
20. Caring about people is part of making a scientific choice, such as whether to use pesticides on plants.	0	1	2	3	4

21. Science helps me to make sensible decisions.	0	1	2	3	4
22. The things I do in science have nothing to do with the real world.	0	1	2	3	4
23. Science helps me to make decisions that could affect my natural environment.	0	1	2	3	4
24. Learning science will have an effect on the way I vote in elections.	0	1	2	3	4
25. Making decisions can be difficult without reliable evidence.	0	1	2	3	4

Appendix 5: CARS B

Relevance of Science Questionnaire
Please read the below statements and indicate the degree to which you agree or disagree by circling 0-4.

0 = Strongly Disagree (SD) **1** = Disagree (D) **2** = Neutral(N) **3** = Agree (A) **4** = Strongly Agree (SA)

	SD	D	N	A	SA
1. Science has nothing to do with local issues, such as waste from nearby factories.	0	1	2	3	4
2. Science can help me make better decisions about what I buy.	0	1	2	3	4
3. Science experiments can help me to better understand the world.	0	1	2	3	4
4. Much of what I learn in Geology class is useful in my everyday life today.	0	1	2	3	4
5. Learning science can help me when I choose where to buy a house.	0	1	2	3	4
6. Caring about people is part of making a scientific choice, such as whether to use pesticides on plants.	0	1	2	3	4
7. Science helps me to make sensible decisions.	0	1	2	3	4
8. The things I do in science have nothing to do with the real world.	0	1	2	3	4
9. Science helps me to make decisions that could affect my natural environment.	0	1	2	3	4
10. Learning science will have an effect on the way I vote in elections.	0	1	2	3	4
11. Making decisions can be difficult without reliable evidence.	0	1	2	3	4
12. Science class helps me to work with others to make decisions.	0	1	2	3	4
13. I am interested in learning more about computer technology.	0	1	2	3	4
14. I would like to learn more about strategies for thinking in my science class.	0	1	2	3	4
15. Knowledge of science helps me to avoid natural hazards.	0	1	2	3	4
16. Using scientific methods helps me make environmental decisions.	0	1	2	3	4
17. Learning science is not important for my future success.	0	1	2	3	4
18. I only take science because it is a required course.	0	1	2	3	4
19. In most cases, personal feelings are important for making choices in science.	0	1	2	3	4
20. Knowing science can help me to make better choices about energy issues.	0	1	2	3	4

21. Collecting evidence is an important part of making a decision.	0	1	2	3	4
22. Geology class will help prepare me for major decisions in my future.	0	1	2	3	4
23. I will only take math classes for as long as I have to.	0	1	2	3	4
24. Learning science enables me to explain my thoughts better to others.	0	1	2	3	4
25. Knowledge of science will help me protect the environment.	0	1	2	3	4

Appendix 6: CARS C

Relevance of Science Questionnaire
Please read the below statements and indicate the degree to which you agree or disagree by circling 0-4.

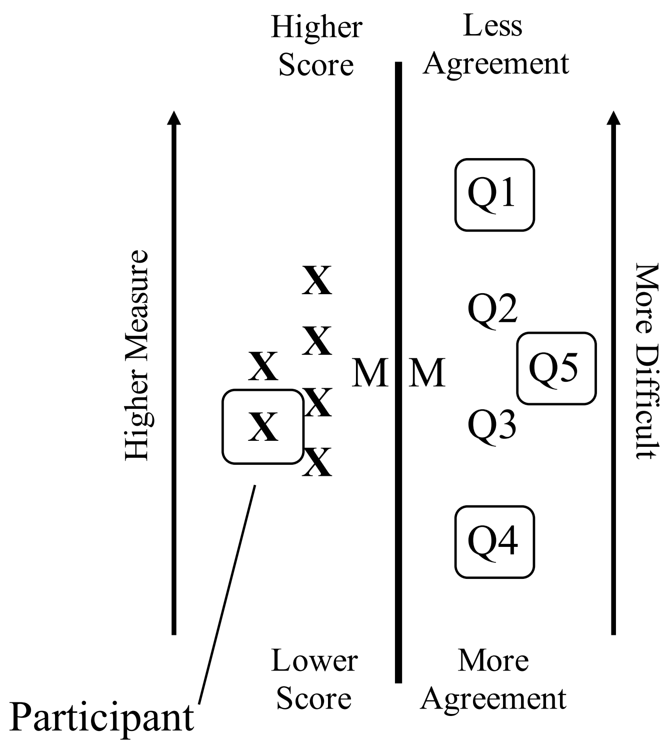
0 = Strongly Disagree (SD) **1** = Disagree (D) **2** = Neutral(N) **3** = Agree (A) **4** = Strongly Agree (SA)

	SD	D	N	A	SA
1. Much of what I learn in Geology class is useful in my everyday life today.	0	1	2	3	4
2. Learning science can help me when I choose where to buy a house.	0	1	2	3	4
3. Caring about people is part of making a scientific choice, such as whether to use pesticides on plants.	0	1	2	3	4
4. Science helps me to make sensible decisions.	0	1	2	3	4
5. The things I do in science have nothing to do with the real world.	0	1	2	3	4
6. Science helps me to make decisions that could affect my natural environment.	0	1	2	3	4
7. Learning science will have an effect on the way I vote in elections.	0	1	2	3	4
8. Making decisions can be difficult without reliable evidence.	0	1	2	3	4
9. I do not expect to use science much when I get out of school.	0	1	2	3	4
10. I value careers in science, such as a hydrologist or engineer.	0	1	2	3	4
11. Making decisions can be difficult when I don't understand the choices.	0	1	2	3	4
12. My intuition helps me make decisions in science.	0	1	2	3	4
13. Science will help me to understand the effect I have on the environment.	0	1	2	3	4
14. Science helps me to ask others for help with my work.	0	1	2	3	4
15. Using scientific methods helps me think things through.	0	1	2	3	4
16. Science can help me decide how to treat my cold or illness.	0	1	2	3	4
17. Science impacts my feelings about global issues I am considering.	0	1	2	3	4
18. Science should be required in college.	0	1	2	3	4
19. Science could help me figure out how to spin / shoot / throw / hit a ball.	0	1	2	3	4
20. Science class helps me to evaluate my own work.	0	1	2	3	4

21. I have support from others to excel at science.	0	1	2	3	4
22. Using scientific methods helps me decide what to buy in the store.	0	1	2	3	4
23. Science will help me understand the importance of recycling.	0	1	2	3	4
24. Learning science can help me understand about things that affect people's health.	0	1	2	3	4
25. Science can help me to make better choices about various things in my life.	0	1	2	3	4

Appendix 7: Wright Map Example Diagram

Wright Map Example



Q1: Item showing the least agreement among the sample.

Q5: Item representing the average attitude of the sample

Q4: Item representing the most agreement among the sample.

Appendix 8: CARS survey Wright Maps

Spring

CARS A

```
MEASURE   PERSON - MAP - ITEM
              <more>|<rare>
   3                +
                    |
                    |
                    |  Q13
              XX    |
   2           X  T+T
             XXX    |  Q.14
            XXXX    |
              XX  S|  Q12
     XXXXXXXXXXX    |
   1 XXXXXXXXXXXX   +S
        XXXXXXXX M|  Q17    Q18
       XXXXXXXXX    |  Q2
       XXXXXXXXX    |  Q.6    Q1     Q11    Q24    Q7
           XXXXX  S|  Q20
   0         XXX    +M Q3     Q4
          XXXXXX    |  Q10    Q21
                  T|  Q16
                    |  Q.15   Q19
               X    |
  -1                +S Q23    Q8
                    |  Q.9    Q25
                    |  Q.22
                    |
                    |  Q5
  -2                +T
              <less>|<freq>
```

CARS B

```
MEASURE PERSON - MAP - ITEM
          <more>|<rare>
   5            +
                |
                |
           X    |
                |
                |
   4            +
                |
                |
           X    |
                |
                |
   3       X  T+
         XXX    |
                |
                |
          XX    |
          XX    |
   2       X  S+T Q19
           X    |
         XXX    |  Q.23
       XXXXX    |
     XXXXXXX    |
    XXXXXXXX    |  Q13    Q22
   1 XXXXXXXX M+S Q.18
       XXXXX    |  Q10    Q4
    XXXXXXXX    |  Q14
          XX    |  Q6
          XX    |  Q12    Q24
        XXXX    |
   0      XX  S+M Q16    Q7
       XXXXX    |  Q2
           X    |  Q.17
         XXX    |  Q15    Q25
          XX    |  Q5
              T|  Q20    Q9
  -1            +S Q3
                |  Q.8    Q11    Q2
                |
                |
                |
                |
  -2            +T Q.1
                |
                |
                |
                |
                |
  -3            +
```

CARS C

```
MEASURE PERSON - MAP - ITEM
        <more>|<rare>
   5       X  +
              |
              |
              |
              |
              |
   4          +
              |
              |
           X  |
              |
           X  |
   3       X T+
              |
              |
              |
           X  |
           X  |
   2      XX S+
              |
     XXXXX    |
   XXXXXXX    |T Q22
        XX    |
       XXX    |  Q14
   1      XX M+  Q7
   XXXXXXX    |  Q20    Q21
     XXXXX    |S Q1     Q19
       XXX    |
       XXX    |  Q.9    Q12    Q16    Q3
       XXX    |  Q15
   0   XXX    +M
        XX  S|  Q10    Q17    Q18    Q23
         X    |  Q4
         X    |  Q25
              |S
              |  Q13    Q2     Q24    Q6
  -1          +  Q11    Q8
            T|  Q.5
              |
              |T
              |
              |
  -2          +
         X    |
              |
              |
              |
         X    |
  -3          +
        <less>|<freq>
```

Fall

CARS A

```
MEASURE     PERSON - MAP - ITEM
                <more>|<rare>
   3                  +
                      |
                 .    |  Q13
                      |
                .#    |
   2             # T+T
                 .    |  Q.14
                .#    |  Q12
                .# S|
                 #    |
   1      .#######   +S Q17
            .##### M|  Q1
            .#####    |  Q18
             #####    |  Q11      Q2
            .#####    |  Q.6      Q24      Q7
   0           .## S+M Q4
                      |  Q16      Q20      Q3
               .##    |  Q.15     Q19
                 . T|  Q10      Q21
                 .    |  Q23      Q25
  -1                 +S Q.9
                      |  Q.22
                      |
                      |  Q8
                      |  Q5
  -2                 +T
                <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

CARS B

```
MEASURE  PERSON - MAP - ITEM
             <more>|<rare>
    4                 +
              X    |
                   |
              X    |
              X    |
                   |
    3              T+
             XX    |
             XX    |
                   |
              X    |  Q.23
             XX    |T
    2       XXX S+  Q19
            XXX    |
         XXXXXX    |
          XXXXX    |  Q.18    Q13
          XXXXX    |
          XXXXX M|S Q22
    1 XXXXXXXXXXX  +
          XXXXX    |  Q10     Q4
         XXXXXX    |
           XXXX    |  Q14     Q24     Q6
           XXXX    |
         XXXXXX S|  Q16
    0         X    +M Q2
           XXXX    |  Q12
              X    |  Q.17    Q7
              X    |  Q15
                  T|  Q9
                   |  Q11     Q25
   -1              +  Q20
              X    |S Q5
                   |  Q3
                   |  Q21
                   |  Q.1     Q.8
                   |
   -2              +
             <less>|<freq>
```

CARS C

```
MEASURE PERSON - MAP - ITEM
           <more>|<rare>
    5             +
                  |
            X     |
            X     |
                  |
           XX     |
    4       X     +
           XX T|
                  |
                  |
                  |
            X     |
    3       X     +
                  |
            X     |
            X S|
           XX     |
          XXX     |
    2      XX     +
           XX     |T
          XXX     |
        XXXXX     |  Q22
           XX     |
           XX M|  Q.9
    1    XXXX     +  Q12     Q14     Q16     Q21
         XXXX     |S Q19     Q7
         XXXX     |  Q1      Q20
         XXXX     |  Q10
         XXXX     |
           XX     |  Q15
    0 XXXXXXXXXX  +M Q18     Q3
            X S|
         XXXX     |
            X     |  Q17     Q25     Q4
                  |  Q23
                  |S Q.5
   -1       X     +  Q2      Q24
                  |  Q13     Q6      Q8
                  |
                 T|  Q11
                  |
```

Appendix 9: SAI II survey Wright Maps

Spring

Pre-class SAI II

```
MEASURE    PERSON - MAP - ITEM
              <more>|<rare>
   70               +
                    |
                    |
                    |  Q.31
                    |
                    |
                 #  |
                   T|T Q.24
                    |  Q.9
   60          .##  +
              #### S|  Q.15
           .######  |  Q36
              ####  |S Q.22         Q27
           ####### M|  Q.2
             #####  |  Q.39
             .#### S|  Q.26         Q30          Q40
              ####  |  Q.35         Q21          Q34
   50            .  +  Q.8          Q1           Q12
                 # T|M Q.14         Q20
                    |  Q.13         Q17          Q29          Q33
                    |  Q23
                    |  Q.32         Q.37         Q.7          Q19
                       Q4
                    |  Q.11         Q10          Q25
                 .  |  Q16
                    |S Q28
                    |  Q.38         Q.6          Q18
   40               +  Q.3
                    |
                    |
                    |T
                    |
                    |  Q.5
                    |
                    |
                    |
   30               +
              <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Post-class SAII II

```
MEASURE  PERSON - MAP - ITEM
            <more>|<rare>
                  |  Q.9
                  |  Q.31
                  |T Q.24
               X  |
   60         XX T+
            XXXX  |
           XXXXX S|  Q.22         Q36
     XXXXXXXXXXX  |  Q.15
        XXXXXXXX  |S Q27
     XXXXXXXXXXX M|  Q.2          Q.39         Q40
     XXXXXXXXXXX  |  Q.26
            XXXX S|  Q.14         Q21          Q30          Q34
             XXX  |  Q.35         Q12
   50         XX T+M Q.13         Q.8          Q1           Q20
               X  |
               X  |  Q.37         Q.7          Q17          Q33
                  |  Q23          Q29
                  |  Q19
                  |  Q.11         Q.32         Q10          Q25
                     Q28          Q4
                  |S Q.38         Q16
                  |  Q.6
   40             +  Q.3
                  |  Q18
                  |
                  |
                  |T
                  |  Q.5
                  |
                  |
                  |
   30             +
            <less>|<freq>
```

Fall

Pre-class SAI II

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
                     |  Q.31
                  .  |  Q.9
                  .  |
                  . T|T
   60                +  Q.24
                ##   |
               .###  |
                 .# S|  Q.2    Q.22   Q36
               .###  |  Q.15
                ###  |S
             ###### M|  Q.35
              #####  |  Q.26   Q.39   Q27
           .######## |  Q40
                .##  |  Q30    Q34
   50          .### S+  Q.14   Q12
                 ##  |M Q.32   Q1     Q20    Q29
                  .  |  Q.13   Q.8    Q21
                     |  Q23    Q33
                  . T|  Q.7    Q19
                     |  Q17    Q4
                  .  |  Q.37   Q10    Q25
                     |S Q.11   Q.38   Q16    Q28
                  .  |
                     |  Q18
   40                +  Q.6
                     |  Q.3
                     |
                     |T Q.5
                     |
                     |
                     |
                     |
                     |
                     |
   30                +
               <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```

Post-class SAII II

```
MEASURE    PERSON - MAP - ITEM
               <more>|<rare>
                     |T Q.31
   60               #  +  Q.9
                  .  |  Q.22   Q.24
                 ## T|  Q36
                 .#  |  Q.15
                .## S|S Q27
                .##  |  Q40
            #######  |  Q.2    Q.39   Q1     Q30
                ### M|  Q.14   Q.35
          .######## |  Q12    Q20
   50          .####  +  Q.13   Q.26   Q34
             .##### S|M Q.8    Q21
                  .  |  Q29    Q33
                 .# T|  Q23
                     |  Q.32   Q.37   Q17
                  .  |  Q.7    Q25    Q4
                     |  Q.38   Q.6    Q10    Q19    Q28
                     |S
   40                +  Q.11   Q.3    Q16
                     |  Q18
                     |
                     |
                     |T
                     |  Q.5
                     |
                     |
                     |
   30                +
               <less>|<freq>
EACH "#" IS 2: EACH "." IS 1
```