

THE IMPACT OF INTEGRATED TELESCOPE USE ON ATTITUDES AND
CONCEPTUAL UNDERSTANDING OF INTRODUCTORY
ASTRONOMY STUDENTS

by

Kisha Marie Delain

A professional paper submitted in partial fulfillment
of the requirements for the degree

of

Master of Science

in

Science Education

MONTANA STATE UNIVERSITY
Bozeman, Montana

July 2015

©COPYRIGHT

by

Kisha Marie Delain

2015

All Rights Reserved

DEDICATION

To my husband Charles, for all your support. To my daughter Amaya, who despite being young tried to understand that mommy was doing her “homework.” To my parents Judy and Dave for all your encouragement through the years, especially my mother for discussions on the topic of education. Special thanks to Dr. Eric Brunsell in the MSSE program at Montana State University for advising and critique throughout this process, to Dr. Elizabeth Wehner for letting me co-opt her classroom and Dr. Gerry Ruch for putting up with my “do this now” demands. Lastly, many thanks to Sarah Sargent, without whom I would not be the writer I am today.

There are too many other colleagues, friends and mentors who have helped me in innumerable ways on this road to list individually, but I could not have done it without you. Thank you.

TABLE OF CONTENTS

1. INTRODUCTION AND BACKGROUND	1
2. CONCEPTUAL FRAMEWORK	3
3. METHODOLOGY	9
4. DATA AND ANALYSIS	14
5. INTERPRETATION AND CONCLUSION	33
6. VALUE	35
REFERENCES CITED	36
APPENDICES	43
APPENDIX A: IRB Form	44
APPENDIX B: ATSI Additional Questions	46
APPENDIX C: TOAST Additional Questions	49
APPENDIX D: Survey 1: H-R Diagram	51
APPENDIX E: Survey 2: Measuring Jupiter's Mass	53
APPENDIX F: Interview Questions	56
APPENDIX G: Histograms of Attitude Survey	59

LIST OF TABLES

1. Triangulation Matrix	14
2. Chi-squared Goodness-of-fit for ATSI Data, Comparison to Treatment	18
3. Chi-squared Goodness-of-fit for ATSI Data, Pre- to Post-Instruction	19
4. Chi-squared Goodness-of-fit for Student Confidence	19
5. Chi-squared Goodness of fit for Real Data Importance.....	27
6. Wilcoxon Rank-sum Test for TOAST, Comparison to Treatment.....	28
7. Wilcoxon Rank-sum Test for TOAST, Pre- to Post-Instruction.....	28
8. TOAST Pre-test Means, Medians, and Standard Deviations.....	31
9. TOAST Post-test Means, Medians, and Standard Deviations	32
10. TOAST Gains	33

LIST OF FIGURES

1. Rubric to Evaluate Inquiry from Bruck et al. 2008	5
2. How Do Students Feel About Science?.....	16
3. Do Students Think They Can Do Science?.....	17
4. Survey Responses for Real Data Importance, Jupiter Lab	21
5. Survey Responses for Real Data Importance, H-R Diagram.....	22
6. Survey Responses for Simulation vs. Real Data.....	23
7. Survey Responses for Own Telescope Data Importance	24
8. Survey Responses for Authentic Data Importance	26
9. TOAST Content Test Results, Related Questions	30
10. TOAST Content Test Results, All Questions	31

ABSTRACT

This paper discusses a pilot program using real data from our university telescope in laboratory exercises with our students at the University of St. Thomas. In this paper, I evaluate the effectiveness of these labs and their impact on student attitudes using standardized learning assessments combined with student interviews and surveys. One hundred and sixteen undergraduate students enrolled our fall introductory astronomy course completed our newly developed exercises as a normal part of the laboratory portion of the introductory astronomy course, with comparison groups utilizing “canned” or simulated data. Student attitudes improved in the treatment group more than in the comparison group in both questions of interest: how students feel about science, and whether or not students feel they can do science. In addition, student content knowledge gains are about 10% higher in treatment groups than in comparison groups.

Keywords

College

Astronomy

College non-majors

Research in teaching/learning

Teaching approaches

Course curriculum

Laboratory exercises

Assessment

INTRODUCTION AND BACKGROUND

Introductory astronomy is a course that often also includes a laboratory component in order to allow students to fulfill the laboratory science distribution requirements. There are many available laboratory exercises, including several simulations of the telescope experience for introductory majors. However many institutions do not have access to required equipment to run observational exercises using real data in their classrooms (Edelson, Gordin & Pea 1999; Lee and Songer 2003). At the community college level and high school level, this lack of access can be even more challenging. One of our faculty members received an NSF grant to expand access to our telescope by developing flexible curriculum and data processing software for use at a variety of educational levels. We have developed a set of three laboratory exercises appropriate for use in an introductory astronomy course using real data from our telescope, with the goal of making these exercises and access to the telescope available to local high schools and community colleges as well as our own students. This paper discusses the pilot program using these laboratory exercises with our students at the University of St. Thomas.

The University of St. Thomas is a private Catholic liberal arts university with 6,234 undergraduate students and 3,995 graduate students in the fall of 2014. The student population is primarily white suburban students, with 14% of undergraduates as students of color; only 205 undergraduate students are international students. I have been teaching introductory astronomy and its associated laboratory in some form or another since 1996 and at the University of St. Thomas as an adjunct for five years, during which

the observatory has come online largely through the efforts of another faculty member. In addition, we have collaboratively developed laboratory exercises to utilize the new capabilities of the telescope. Notably, the telescope can be run manually, with students at the computer choosing objects and taking pictures as well as robotically with observing plans submitted to a scheduler. The laboratory exercises developed to provide access involve both types of observing, with the manual observing easily replaced by robotic observations if attending an observing session is not possible. One hundred and sixteen undergraduate students in our fall introductory astronomy course completed our newly developed exercises as a normal part of the laboratory portion of the introductory astronomy course, with comparison groups utilizing “canned” or simulated data. In this paper, I evaluate the effectiveness of these labs and their impact on student attitudes using standardized learning assessments combined with student interviews and surveys.

Specifically, I attempt to address the question of how student ownership of telescope data in laboratory exercises impacts student attitudes toward science. I investigate the difference in student attitudes and content retention compared to several different laboratory elements: using real vs. simulated data; using real data that is their own vs. real data which is given to them; and the authentic process of developing and executing an observing plan and analyzing the results. These questions will drive future development of inquiry-style labs targeted toward our audience of non-science students in the introductory courses.

CONCEPTUAL FRAMEWORK


Our interest in student data ownership comes as we transition our labs from a more cookbook style to guided and open inquiry. In the past several decades, research has shown that student engagement through inquiry methods improves student attitudes and engagement with the material (Cohen 1994; Basey & Francis 2011; Enghag & Niedderer, 2008; Martin-Hansen 2002; Prins et al 2009) as well as increasing student learning at least in the short term (Basey & Francis 2011; Chin and Brown 2000; Bunterm, Lee, Kong, Srikoon, Vangpoomyai, Rattanaovongsa, & Rachahoon 2014). Inquiry, which is grounded in constructivist learning theory (Piaget 1928; Vygotsky 1934), is often divided into levels depending on the amount of control the student has. Constructivist pedagogy consists of teaching methods in which students “actively develop, test, and revise their ideas... through collaborative firsthand inquiry with their peers” (Smith, Maclin, Houghton, & Hennessey 2000). Several different methods of categorizing level of inquiry have been developed (Nadelson, Walters, & Waterman 2010; Buck, Bretz & Towns 2008; Chinn & Malhotra 2002; NRC 2000; Martin-Hansen 2002), with the ultimate goal of giving students both the skills and the knowledge to generate their own, independent inquiry. Indeed this is precisely what science entails: asking questions, determining methods of answering those questions, evaluating data, and eventually drawing conclusions (Power, B. 2012; Chi, Feltovich, & Glaser 1981). The many levels of inquiry can be designed to provide students with scaffolding to eventually come to this expert level (Bell, Smetana & Binns 2005; Kirschner, Sweller & Clark 2006;

Buck Bretz & Towns 2008; Chatterjee, Williamson, McCann & Peck 2009; Sadeh & Zion 2009, 2012; Bunterma et al. 2014).

Unfortunately, the literature is inconsistent in the application of the term “inquiry,” which has different meanings to different researchers and settings. The definitions used in the secondary education settings have little overlap with those used in collegiate settings (Bruck et al 2008). Even within subsets of inquiry (e.g. guided inquiry) there is little agreement on the meaning of terms. Several researchers note that while hands-on or active learning tasks are often part of inquiry, it is not the case that all hands-on or active learning tasks are inquiry (Nadelson, Walters & Waterman 2010; Tang, Coffey, Elby & Levin (2009)). Rather, it is the processes of scientific research that are important, including “mechanistic reasoning, model-based reasoning, scientific argumentation, and sense making” (Tang, Coffey, Elby & Levin 2009; Hammer, Russ, Mikeska, & Scherr, 2005; Lehrer & Schauble, 2004; Warren, Ballenger, Ogonowski, & Roseberry, 2001). However, the term inquiry is often applied to the former set of active learning tasks in addition to the more open, unguided processes inherent to authentic science. Because of this lack of consistency, several rubrics have been discussed as a means to connect the terminology in use to levels of student independence (Bruck 2008; NRC 2000; Schwab 1962).

A rubric to characterize inquiry in the undergraduate laboratory.

Characteristic	Level 0: Confirmation	Level ½: Structured Inquiry	Level 1: Guided Inquiry	Level 2: Open Inquiry	Level 3: Authentic Inquiry
Problem/Question	Provided	Provided	Provided	Provided	Not provided
Theory/Background	Provided	Provided	Provided	Provided	Not provided
Procedures/Design	Provided	Provided	Provided	Not provided	Not provided
Results analysis	Provided	Provided	Not provided	Not provided	Not provided
Results communication	Provided	Not provided	Not provided	Not provided	Not provided
Conclusions	Provided	Not provided	Not provided	Not provided	Not provided



More structure **Less structure**

Figure 1: Rubric to evaluate inquiry from Bruck et al. 2008.

In this paper I will be using the levels of inquiry as described in Figure 1, though it is also useful to understand inquiry based on the following seven principles given by Jonassen (1994) on page 36 of their discussion of constructivism:

1. Provide multiple representations of reality – avoid oversimplification of instruction by representing the natural complexity of the world
2. Focus on knowledge construction not reproduction
3. Present authentic tasks (contextualizing rather than abstracting instruction)
4. Provide real world, case based learning environments, rather than predetermined instructional sequences
5. Foster reflective practice
6. Enable context, and content, dependent knowledge construction
7. Support collaborative construction of knowledge through social negotiation, not competition among learners for recognition.

Because many students arrive in college without significant open inquiry experiences (Hume 2009; Hume & Coll 2010; Bruck et al. 2008), and because of the wide range of scientific backgrounds of students in the introductory astronomy course, it is challenging to introduce inquiry directly to the students; at the very least enough theory and background must be provided before students can move forward (Leonard and Penick 2009; Edelson 1998). Researchers seeking to develop inquiry thus typically focus on one aspect to further. Development of scientific thinking skills and content knowledge can occur separately or in conjunction; many researchers focus on one aspect or another in a specific study (White and Frederiksen 1998; Linn and Songer 1991; Lee and Songer 2003). However, it is important to capture the complexity of scientific thinking by incorporating content knowledge, inquiry skills and resources together (Lee and Songer 2003). Ideally students who continue on in other science courses would fully develop inquiry methods (Sunal, Wright & Day, 2004), but this is not the general population of our astronomy course. Even for majors in upper level courses, several researchers have found that students are rarely given inquiry opportunities, despite the documented benefits to providing students with these opportunities (Nadelson, Walkers & Waterman; Bruck et al 2008; Chinn & Malhotra, 2002; Martin-Hansen, 2002).

One of the common themes of inquiry is the concept of authenticity. As with inquiry itself, authenticity has a wide range of definitions in the literature. As proposed by Bruck et al. (2008), authentic inquiry would only include investigations where none of the information is given, including the concept under investigation. While this is fully authentic science, many others believe that there can be components of authentic science

within other types of inquiry, effectively creating authentic situations which are nonetheless suited to the students' background (Leonard and Penick 2009; Lee and Songer 2003). For example, authenticity may be added to a lower level inquiry task by tying the material to the students' lives, addressing real-world problems, involving students in processes similar to those faced by scientists, or otherwise integrating the scientific community, data, or processes in to the inquiry (Lee and Songer 2003).

Toward this end, a number of groups have attempted to create educational virtual environments (EVEs) using either real data or simulated data to invoke the processes that scientists use (Lin, Hsu & Yeh 2012; Trundle & Bell 2010). While some of these virtual environments utilize immersive virtual reality with multi-sensory interactions, most use real-world, authentic tasks involving visual representations of content and context to construct knowledge (Mikropoulos and Natsis 2011; Smetana & Bell 2012). In addition, these EVEs tend to help represent authentic science through complexity present in the world (Mikropoulos and Natsis 2011; Rutten, van Joolingen, van der Veen 2012; Chen 2010). Smetana & Bell's literature review of 61 papers (2012) finds that simulations with supported technology use is most effective when used in inquiry-based situations where students are encouraged to reflect on activities which promote cognitive dissonance. Used in this way, learning of science content is improved compared to traditional lecture-based formats, and is similar to hands-on inquiry teaching methods (Smetana & Bell 2012). The use of simulations can improve over hands-on inquiry if the number of repetitions is higher with simulations and students are allowed to explore concepts more fully with targeted activities. However, the impact on long-term

conceptual understanding is minimal, as prior conceptual models are extremely difficult to replace (Smetana & Bell 2012).

Astronomy is in a unique position in that there is a huge amount of archival data available, and this data can be used to incorporate real, authentic research opportunities into the classroom (Taasoobshirazi, Zuiker, Anderson & Hickey 2006). However, it can be challenging for students at the introductory level to access and to process this data in a feasible way. By providing some initial processing, teachers can scaffold student experiences with real data (Taasoobshirazi, Zuiker, Anderson & Hickey 2006; Ucar and Trundle 2011; Raeside, Busschots, Waddington, & Keating 2008). In particular, web-based EVEs can provide interactions with the images which can be accessed anywhere without additional installation of software, thus greatly increasing availability of real data (Raeside et al. 2008). This use of real data is one way to increase authenticity in our introductory laboratory class.

To extend the use beyond just archival data, our laboratory design is for the students to take ownership of the data by creating observing plans and observing at the telescope. Ownership of data and the process has been investigated to a lesser extent. Prins et al (2009) found that ownership influenced student interest in the topic by providing a clear link between theory and practice along with the broader value in student appreciation for ‘understanding models and learning to construct models’. In our case, our goal is not to arrive at complete authentic inquiry, nor even open inquiry, with our introductory students, as we will need to scaffold the students’ knowledge and skill sets. Instead, we have developed labs using some guided inquiry and some structured inquiry,

with more of a focus on authenticity and ownership of data with a guided analysis process close to the process that scientists use.

METHODOLOGY

The goal of this research is to determine whether integrating telescope use into the astronomy laboratory impacts both student attitudes and conceptual understanding in our introductory astronomy class. This course is a traditional lecture format introductory course with separate lab component. All students enrolled in the course are in both lecture and lab, with up to 24 students in each of five lab sections. Two of the sections were taught using simulations and three sections were taught using data from the telescope, with two instructors teaching both a comparison and intervention group. The last section is not included in this analysis, as the instructor declined to participate. Our goal is to introduce students to the real process of science, by using real data from our own telescope, to raise both their confidence and their awareness of the scientific process.

Participants

There were 116 students enrolled in the fall course, of which one declined to answer any demographic information and two of which declined to participate in the study. The remainder of students in this class were predominantly white (81%), with 1% Hispanic-American, 3% Asian-American, 4% African-American, 6% mixed race students and 1% international students, with 5% declining to answer this question. 69% of students were male and 31% were female. Students come from a variety of backgrounds, predominantly suburban at 71% (3% rural, 16% small town, and 9% urban). This course is targeted toward non-majors fulfilling the “lab science” requirement (83% of this class),

although a few students (5%) take the course to fulfill the physics minor or as interested science students (11%). As such, the course consists of a mixed group of underclassmen (16% freshman, 44% sophomores) and upperclassmen (18% juniors, 21% seniors). Students indicated a fairly bell-shaped curve in confidence on the difficulty of the course, with 21% expecting the course to be difficult for them, 54% unsure, and 25% expecting the course to be easy. Despite this, students skewed toward believing their math skills to be average or higher, with only 11% rating themselves as “poor or very poor” in math skills (31% average, 57% as good or very good). Lab sections were divided into comparison group ($n = 44$) and treatment group ($n = 48$) based on pre-test attitude surveys. For each instructor, the lab section with less confidence received the treatment. Demographics for each lab reflected the overall demographics of the class. The research methodology for this project received an exemption by Montana State University's Institutional Review Board and compliance for working with human subjects was maintained (Appendix A).

Intervention

Prior to creating our intervention activities, student laboratories involved pen-and-paper exercises, some computer simulations of the telescope experience, and a few laboratory experiments. With the ability to use the telescope we have developed several labs using our own data. The three labs discussed here are the first to be implemented within the introductory astronomy course. One of the labs had a simulation equivalent, the second with paper-and-pen exercise and real data, while the third was run for all groups with the telescope data. All students went to the telescope one night for

observing, and were able to “drive” the telescope using the computer-controlled system. Some of the data for the intervention labs was taken during this time.

The first new lab, Measuring Jupiter’s Mass, consists of a two parts separated by several weeks. In the first class, students work through an exercise designed to make them think about how to measure Jupiter’s mass using telescopic images. Students are expected to devise an observing campaign based on certain constraints (number of allotted observations, weather, and amount of time Jupiter is above the horizon). The class then decides on a consensus set of observations to “submit” to the telescope, which is submitted by the instructor to the telescope for robotic observing. In the second part, students analyze the data received from the telescope (treatment groups) or run a CLEA simulation to find the same “data” (comparison groups). From the data set given, students measured Jupiter’s mass. Due to timing constraints with Jupiter’s actual position in the sky and weather-related issues, this second lab was run nearly two months later than the initial lab.

The second lab targets student understanding of stellar lives using the H-R Diagram. As this topic is very important to understand stars and their lives, we used both our usual exercise using pen-and-paper in addition to a newly written lab using real data from the telescope. All students performed the pen-and-paper first lab. In the second experimental lab, students use a web interface to gather photometric data of brightness and color index (B-V) of stars in two clusters, one old and one young. They then enter this data into a spreadsheet, which allows them to graph the brightness (magnitude) vs. color. Unlike the Jupiter lab, students do not design the observing plan but instead use

data given to them to investigate stellar properties and ages. In the treatment groups, students are aware that the data is taken from the telescope during observation nights. In the comparison groups, students were not told anything about the origin of the data.

The last laboratory exercise targets galaxies and galaxy evolution, tracing Hubble's logic for galaxy classification and adding new information such as color and our understanding of stellar lives (as developed in the previous H-R Diagram lab). Because all students observe at the telescope, all students complete the same laboratory exercise, which begins with a pen-and-paper discussion of the morphology of galaxies. Students are given images from the internet from various telescopes and asked to come up with their own classification system based on appearance (noting that they have no spatial size information). Students then are guided through questions about the ages of galaxies based on their knowledge of stellar lives and asked to come up with a sequence of galaxy evolution. Lastly, students use their own telescope data from the observing nights to make an appealing image using photo editing software, and are asked to classify it using their classification scheme, Hubble's scheme, and comment on its age.

Data Collection

In order to measure student attitudes, the Attitudes Toward Science Instrument (ATSI) (Bailey, Slater & Slater 2010) was given at the beginning and end of the course. To address the challenge identified by the authors of another attitude survey (CLASS, Adams, W., private communication), namely that students tend to equate "science" with many different things, two questions were added to ATSI survey asking students about their definition of science and also on what experiences they based their answers to the

questions (See Appendix B). A range of attitudes were chosen from the initial ATSI survey results with an aggregated “positivity” score created as an average of the answers for each survey question with “positive” answers all coded as positive numbers despite wording of the question, and “negative” answers coded with negative numbers. Thus an average of zero gives a student who answered equally positive and negative on various questions.

We used the Test of Astronomy Standards (TOAST) (Bailey, Slater & Slater 2010) as a pre- and post-test measure of student understanding; however TOAST does not include information on Kepler’s Laws, which are the basis of the Jupiter’s Mass lab, or questions on galaxy stellar content. Three additional questions were added to TOAST to target these content areas, for a total of 30 questions (see Appendix C). Analysis is focused on these additional content questions along with the TOAST questions involving stellar properties. Photocopies of student lab papers for each section were kept for comparison for each of the four labs involved. As students turn in one paper per group, this was 6 papers for each lab, to be used for comparing each group. In addition, teacher notes on the classes and interactions with students were kept.

Five students agreed to interview, with three students from the treatment and two from comparison groups. Students were selected to get a mix of performance (high, medium and low) with a focus on non-science majors, though low response rate led to skewed gender balance (4 female; 1 male). Because of the low response rate, interpretation of the interviews is limited by the students who chose to participate and

there is likely a self-selection effect biasing the interview data. Data collection strategies are summarized in Table 1.

Table 1
Triangulation Matrix

Focus Question: How does student ownership of telescope data in laboratory exercises impact student attitudes toward science?				
Subquestions	Data Source			
Subquestion 1: How does the impact on student attitudes differ between data ownership vs. using a simulation?	Attitude Pre- and Post-Surveys	Student Interviews	Lab Surveys	Teacher journals
Subquestion 2: How does the impact on student attitudes differ between development and execution of an observing plan vs. only using data from the school's telescope?	Attitude Pre- and Post-Surveys	Student Interviews	Lab Surveys	Teacher journals
Subquestion 3: How does student understanding and retention of key concepts change based on ownership of the data vs. simulations?	Content Pre- and Post-Surveys	Laboratory Papers	Summative Assessments	

DATA AND ANALYSIS

Our goal in this research is understanding how telescope use and ownership of data affect students in our introductory class. We analyze our pre-and post-test attitude and content test to determine whether or not students feel happier about science, and whether they feel more confident about doing science. We also analyze the pre- and

post-test for content knowledge. We have divided our class into two comparison groups and two treatment groups, with one comparison and one treatment taught by each instructor. For the first instructor, the comparison group and treatment groups consisted of 23 and 24 students, respectively; for the second instructor the comparison group and the treatment had 21 and 23 students, respectively. Of these students, not all took both pre- and post- tests for either attitudes or content test; analysis includes only matched data for each test.

Impact of student ownership of data on attitudes toward science.

Our primary question is “How does student ownership of telescope data in laboratory exercises impact student attitudes toward science?” To address this question, we can look at the pre- and post- ATSI attitude survey. This attitude survey, designed for introductory astronomy, has questions that can be grouped into three categories: how students *feel* about science; do students think they can *do* science; and how important is science for society? It is these first two categories we are interested in in this paper. Each question in the Likert-style survey was scored from -2 to +2, such that positive attitudes were recorded positively regardless of question wording. The values for each student were then summed to give a total “positivity” score for each question category. A total of 68 students took both the pre and post-test scores, with 33 in the comparison sections and 35 in the treatment sections. Results are shown in Figures 2 and 3.

We can see that while the groups are fairly similar, the treatment group has a lower median attitude to begin with in both categories, with the skew for the comparison groups substantially shifted positive relative to the treatment groups. Post instruction, the

upper end does not change significantly for either group, but the median and skew shift slightly more positive for the treatment group, whereas only the skew shifts to the positive for the comparison groups. Appendix D contains histograms showing the skew for each group pre- and post-instruction for each question.

A two-way chi-squared test over all six questions shows that both prior to and post-instruction the comparison groups are not statistically significantly different from each other on either of the two questions; the two treatment groups are also not different

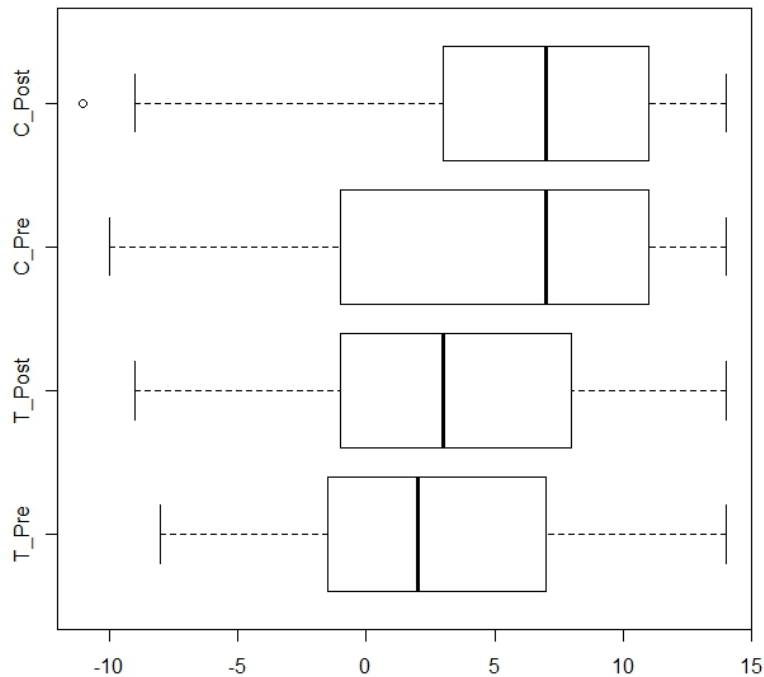


Figure 2: Box and whiskers plot of attitudes for question category “How do students feel about science?” Treatment groups are labeled “T_Pre” and “T_Post” while Comparison groups are labeled “C_Pre” and “C_Post.”

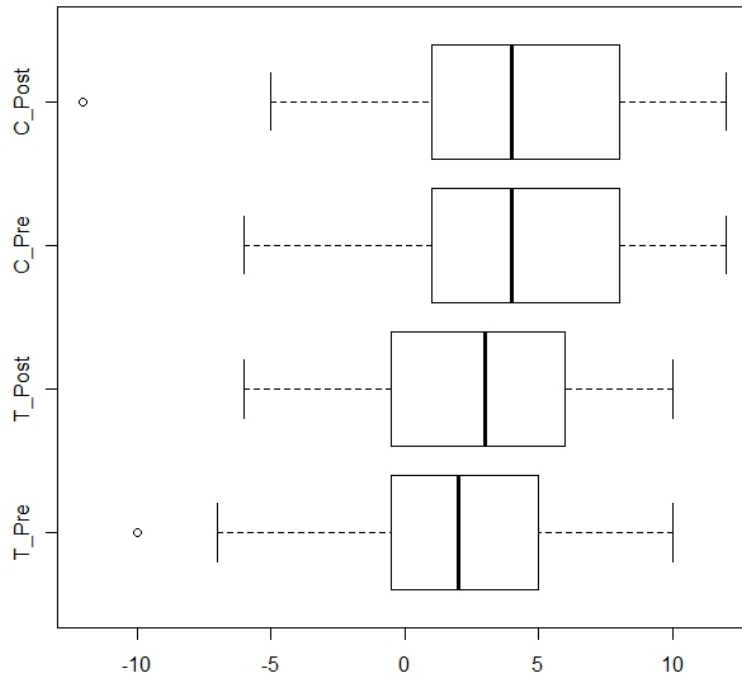


Figure 3: Box and whiskers plot of attitudes for question category “Do students think they can do science?” Treatment groups are labeled “T_Pre” and “T_Post” while Comparison groups are labeled “C_Pre” and “C_Post.”

from each other prior to instruction. We can therefore combine the two comparison and the two treatment groups for analysis. We do find there is a statistical significance that the treatment and comparison groups have different incoming attitudes for both of the questions “How do you feel about science?” and “Do you feel you can do science?” as shown in Table 2.

Table 2
Chi-squared Goodness-of-Fit p-values Comparing the Treatment (n = 35) and Comparison (n = 33) Groups For Both “How Do Students Feel About Science?” and “Do Students Feel They Can Do Science?” Using the ATSI Attitude Test.

Feel About?		Can Do?	
Pre	Post	Pre	Post
0.01	0.006	0.006	0.128

For the questions asking how the students feel about science, we see that the comparison and treatment sections are statistically different from each other both before and after instruction. However, when we compare the questions asking whether students feel they can do science, we find that prior to instruction, we have confidence that the treatment and comparison sections are statistically different, but after instruction the two are *not* statistically different (see Table 2). To see if we have a statistical change in attitudes on these questions, we must compare the groups before and after instruction. For the comparison groups, we find that there is a statistically significant difference in how the comparison groups feel about science, but that they do not feel any more capable at doing science than before instruction. However, for the treatment groups the pre- and post- instruction attitudes are different for both questions.

As we can see from Figure 2, the treatment sections shift toward more positive feelings toward science, and we have strong evidence ($p = 0.0004$) that the treatment sections feel they are more capable of doing science after instruction. The shift in attitude of the comparison group on how they feel about science suggests that the treatment group, which also shifted to the positive, was not enough of a shift to “catch up” with the comparison groups. For the latter question, since the comparison and

treatment sections post-instruction are not statistically different but were different prior to instruction (with the treatment feeling less confident), we can conclude that instruction has raised the confidence of the treatment sections more than the comparison sections.

Table 3

Chi-squared Goodness-of-Fit p-values Between Pre- and Post- comparison Groups and Pre- and Post- treatment Groups For Both “How Do Students Feel About Science?” and “Do Students Feel They Can Do Science?” Using the Pre-test As Null Probabilities For the Post-test

	Feel About?	Can Do?
Comparison (<i>n</i> = 33)	0.006	0.128
Treatment (<i>n</i> = 35)	0.005	0.0004

Table 4

Chi-squared Goodness-of-Fit For Survey Responses For Student Confidence Against the Null Hypothesis That There Is No Difference Between “Agree” “Neutral” and Disagree” Using Lab Surveys

Survey Question:	Do Students Feel Capable?	
	Understand Telescope	Design Confidence
Comparison	<i>p</i> = 0.012 (<i>n</i> = 34)	<i>p</i> = 0.60 (<i>n</i> = 31)
Treatment	<i>p</i> = 0.014 (<i>n</i> = 46)	<i>p</i> = 0.77 (<i>n</i> = 45)

In addition to the ATSI questions addressing student confidence, two questions in the final lab survey indicate that the two groups not statistically different ($p = 0.22$) for design confidence or for understanding the telescope ($p = 0.20$). Neither group is confident that they could design an observing plan if they wanted to, even though they do feel that they understand the telescope’s capabilities, as shown in Table 4. However, even though they are not confident that they could design a project themselves, we find

that our data from surveys shows that students prefer using our own telescope and to a lesser extent data from their own observing plans in lab sections, which we discuss in the following sub-section.

Differences in interest between simulations and authentic observations

Student surveys for both the Jupiter and the H-R diagram labs indicated that students enjoyed having real data (see Figures 4 and 5). For the H-R diagram, the first pen-and paper exercise was less interesting to the students, with the majority of students indicating that the second lab was more interesting. For instance, one student wrote that “the hand graphing made it boring and excel made [the second lab] cool knowing it was real.” Another from the comparison group, who knew that the data was real but not our own noted that “I enjoyed seeing real star clusters instead of a table with plain numbers.” Those students who were told that the second lab involved real data from our telescope wrote “I think it made the lab more interesting knowing that we took this data from our telescope and it wasn't just a random picture. I liked it a lot.” Of the students who directly addressed data in their survey free written responses, 11 indicated that real data from our own telescope was more interesting; two indicated that they thought it was “not important but cooler”; and one indicated that they did not think real data was important at all. Both comparison and treatment groups felt that simulations would not be as interesting as using real data (Figure 6). While real data appears to be important to most students, ownership of data (from our own telescope or from their own observing plan) is important to a slightly lesser degree, as shown in Figure 7.

	Not Important	Neutral	Important
Treatment	9	4	33
Comparison	6	8	20

Figure 4: Student responses to the question “Is using real data important to you?” with Strongly Agree/Agree responses and the Strongly Disagree/Disagree responses aggregated for responses to the survey after the Jupiter Lab.

	Not Important	Neutral	Important
Treatment	8	6	28
Comparison	7	9	25

Figure 5: Student responses to the question “Is using real data important to you?” with Strongly Agree/Agree responses and the Strongly Disagree/Disagree responses aggregated for the H-R Diagram Lab.

	Not as Interesting	Neutral	Just as Interesting
Treatment	26	10	10
Comparison	12	15	6

Figure 6: Student aggregated responses to the question “How would a simulation compare to using real data?”

	Not Important	Neutral	Important
Treatment	11	7	24
Comparison	14	11	16

Figure 7: Student aggregated responses to the question “Is using data from our own telescope important to you?”

Chi-squared goodness-of-fit with null hypothesis that real data does not matter (equal probability for each “agree” “neutral” and “disagree”) was run on each of the individual survey questions with an $\alpha = 0.5$. We find that for both treatment and comparison groups the importance of real data is statistically significant ($p < 0.05$), but for the two comparison groups they feel that having data from our own telescope is not

important. It is possible that the reasoning may be indicated by the interview statements “I just kind of meant by that if you put in data that was just your made up data, it wouldn’t have made me upset or look at it differently. I would have just felt the exact same no matter what.” However, the treatment group felt that having their own data was important (p-value 0.004) “I definitely would prefer like, using the data that we collected. That is pretty cool. But it wasn’t like a huge deal, coming from someone who is kind of not super passionate about astronomy... just trying to get the credits for it.”

Interestingly, the treatment group did not feel that the authentic experience of the Jupiter lab was important (Figure 8), whereas the comparison group was more skewed toward feeling it could be important, though it also did not have any statistical significance. In addition, the long gap between project design for the Jupiter Lab and the data analysis may have impacted the treatment group’s attitudes toward the authentic experience; it is possible that they felt that it was not worth waiting so long for the data. As one of the interviewees from the treatment group commented, “The gap kind of made me forget about it. It was probably one of the reasons why I was like, this isn’t like a huge difference to me, was because it seemed like there was such a disconnect, because it had been such a long time. You know what I mean? Whereas, if it would have been like, right after, then I could have been like ‘oh that just happened!’” In contrast, for the comparison groups, it may be challenging to imagine how engaging having your own telescope data would be. Student survey responses about the authentic experience are shown in Figure 8, and chi-squared goodness-of-fit results for data ownership are summarized in Table 5.

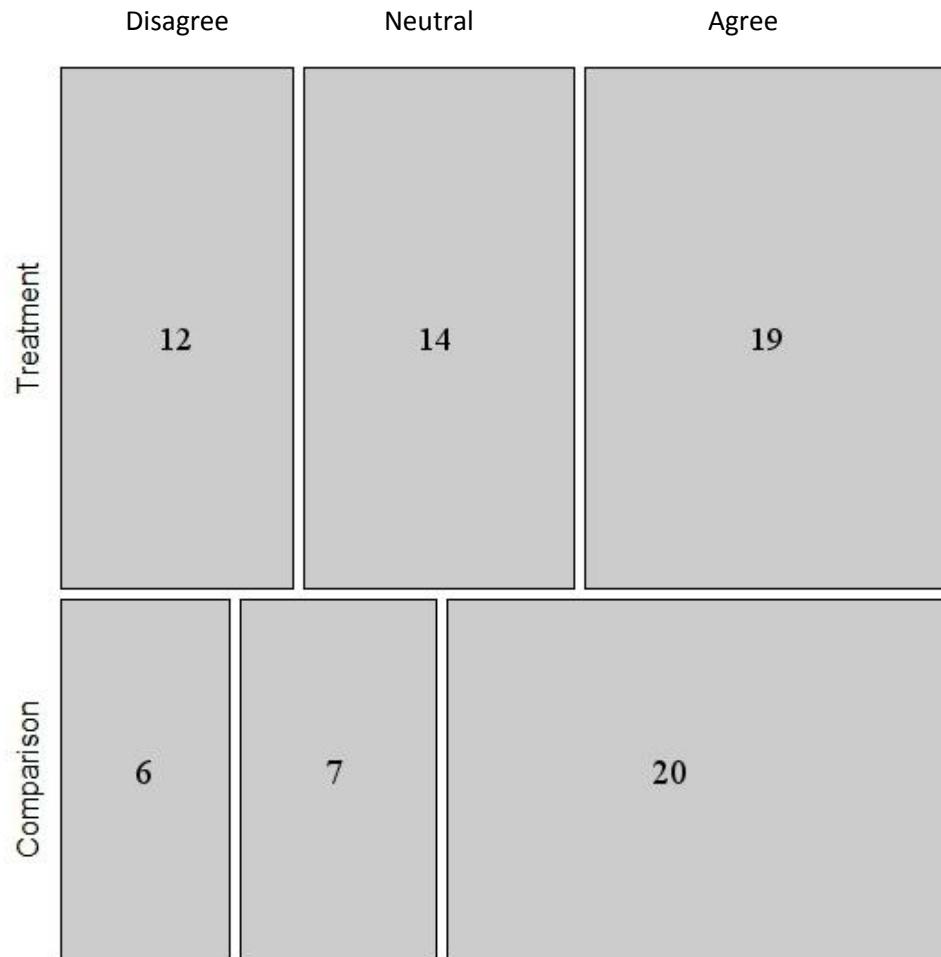


Figure 8: Student aggregated responses to the question “Is this type of authentic experience important to you?”

Table 5.
Chi-squared Goodness of Fit for Survey Responses For Whether Or Not Real Data Is Important To Students Against the Null Hypothesis That There Is No Difference Between “Agree” “Neutral” and Disagree.”

Survey Question:	Is Real Data Important?				
	Real Data 1 <i>p</i> -value	Real Data 2 <i>p</i> -value	Own Data <i>p</i> -value	Simulation <i>p</i> -value	Authentic Experience <i>p</i> -value
Comparison	0.006 (<i>n</i> = 34)	0.0008 (<i>n</i> = 41)	0.629 (<i>n</i> = 41)	0.148 (<i>n</i> = 33)	0.014 (<i>n</i> = 31)
Treatment	1.6e-7 (<i>n</i> = 26)	2.6e-5 (<i>n</i> = 42)	0.004 (<i>n</i> = 42)	0.004 (<i>n</i> = 46)	0.420 (<i>n</i> = 45)

While instructors made a concerted effort to inject the same level of energy and enthusiasm into the labs, we found that it was challenging to be as enthusiastic over not using real data from our own telescope as compared to simulation for the Jupiter Lab. For the second H-R diagram lab, it was also challenging to maintain the same level of enthusiasm when not informing students that the data was real and from our own telescope. It is possible that this may have impacted student attitudes toward the labs.

Differences in student understanding of key concepts between data ownership and simulations

Our last sub-question is to address the content retention by students. For this, students took a pre and post content test (TOAST). Eighty-three percent of the students (76 out of 84) took both the pre and post survey, with 17-19 students in each of the four sections. A Wilcoxon rank-sum test shows that there is no statistical difference between the treatment and comparison groups either before or after the course ($p = 0.72$ pre-test; $p = 0.18$ post-test), as shown in Table 6. This is due to the large standard

deviations in the samples, not any appreciable skew in the data, as we can see in the figures and from the fact that the means and medians are roughly equivalent for each group separately. Though the two groups are not statistically different from each other, overall, the students all improved on the entire survey. Using a Wilcoxon signed rank test for the both groups, we find that we have a very small probability ($p < 0.001$) that the distributions be by chance, with 95% confidence that the true location shift is not zero.

Table 7 shows pre- and post-instruction statistical significance for comparison and treatment groups.

Table 6
Pre- and Post-test TOAST Statistical Significance Using Wilcoxon Rank-sum Test Between Comparison and Treatment Groups For All Questions, and Only the Question Relating To Treatment Lab Concepts

	Pre-test			Post-test		
	p-value	Min 95% confidence interval	Max 95% confidence interval	p-value	Min 95% confidence interval	Max 95% confidence interval
All Questions	0.72	-6.0	7.0	0.18	-17.0	3.0
Relevant Questions	0.56	-7.4	12.5	0.23	-12.5	3.7e-5

Table 7
Wilcoxon Signed-rank Test between Pre- and Post-instruction Statistical Significance For both comparison and Treatment Groups Showing Statistically Significant Learning For Both Sets Of Students On All Questions, and Only the Question Relating To Treatment Lab Concepts

	Comparison			Treatment		
	p-value	Min 95% confidence interval	Max 95% confidence interval	p-value	Min 95% confidence interval	Max 95% confidence interval
All Questions	2.14e-6	-23.3	-13.3	1.88e-8	-21.7	-11.7
Relevant Questions	0.0003	-28.5	-14.0	1.77e-6	-35.5	-25.5

We are interested in particular in the questions that are directly related to the content in the treatment labs, and use these questions for the remainder of this analysis. This includes questions 13, 14, 16, 17 and 27 from TOAST plus our three additional questions (see Appendix C). The first of our additional questions had problems, with a significant number of students selecting the correct answer prior to taking the course, but with answers more evenly distributed among the answers after instruction. However, it seems commonly held knowledge among my students that if you do not know the answer to a multiple choice question to select “C” as it is more statistically likely; the correct answer in this case was indeed “C”. I postulate that this may possibly have led to the high incidence of selecting that answer in the pre-test even though the answers in the test do in fact vary with a slightly *lower* incidence of the choice “C”. We also find that excluding this question does not change the conclusions drawn in this section; we therefore leave it in this analysis. We find that for these questions only, at the 95% confidence level the comparison and treatment groups are again not statistically different ($p = 0.56$ pre-test; 0.23 post-test). These data are seen in Figures 9 and 10. Means, medians and standard deviations given in Tables 8 and 9.

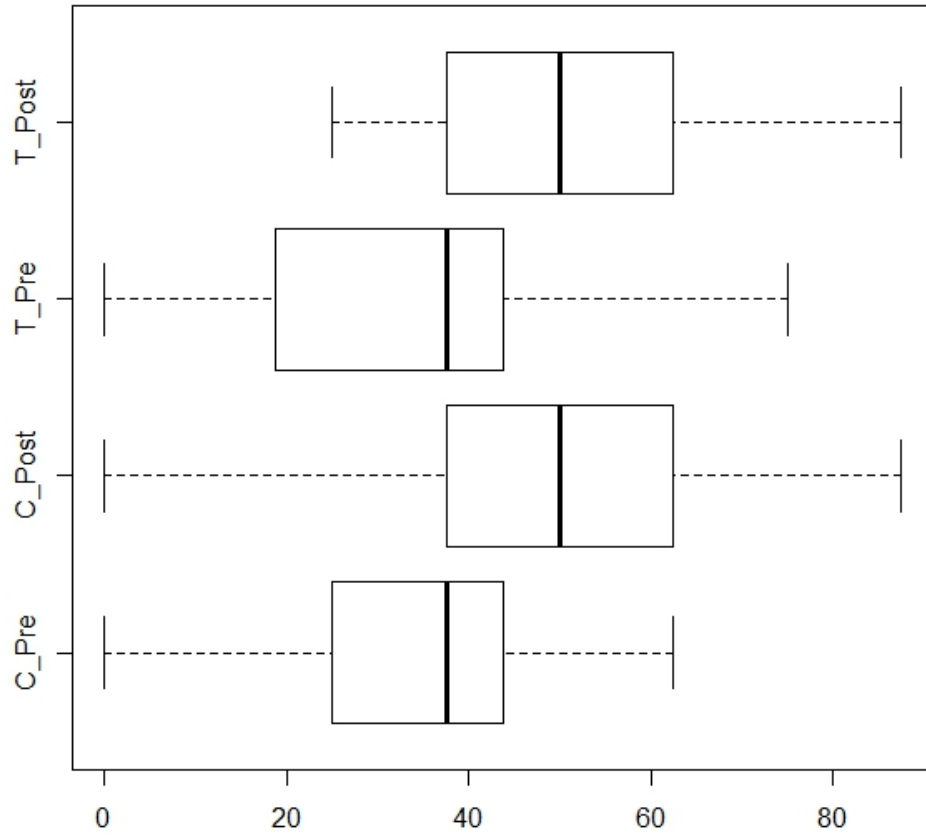


Figure 9: Improvements on TOAST content test for only the questions related to treatment labs.

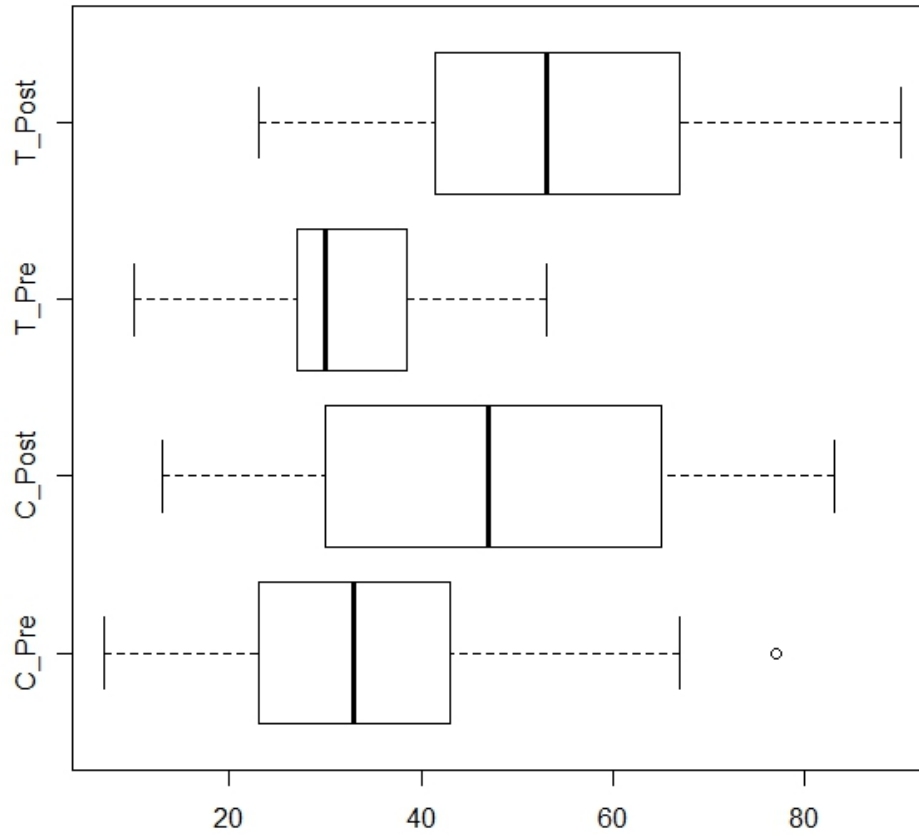


Figure 10: TOAST content test results for all questions.

Table 8

Pre-test Means, Medians, and Standard Deviations For Comparison and Treatment Groups For the TOAST Content Test.

	Pre-test					
	All Questions			Relevant Questions		
	Mean	Median	σ	Mean	Median	σ
Comparison	34.2	33.0	16.2	35.1	37.5	16.6
Treatment	31.7	30.0	10.3	32.5	37.5	16.9

Table 9
Post-test Means, Medians, and Standard Deviations For Comparison and Treatment Groups For the TOAST Content Test.

	All Questions			Relevant Questions		
	Mean	Median	σ	Mean	Median	σ
Comparison	46.8	47.0	20.1	46.2	50.0	20.2
Treatment	53.2	53.0	18.0	52.5	50.0	17.7

As there is little difference between the entire content test and the questions specifically relevant to the lab, we conclude that any difference in engagement among the students applied to the class as a whole, and not just the specific content areas of the lab. Though we find no statistical difference between the comparison and treatment groups due to the large standard deviation, calculating the gains, leads to the surprising discovery that students in the treatment groups have a gain about 10% greater than the gain in the comparison groups. While it is normal that students who perform higher tend to have smaller gains (Pollok 2005; McDermott 1990; Redish & Steinberg 1999), there was little difference between comparison and treatment groups at the outset, making the increase in gain of treatment groups is particularly robust. In particular, looking at comparison and treatment group for a single instructor (e.g. Comparison 1 and Treatment 1), we see that this trend is even more robust, as the matching is even more clear within a single instructor's groups, and similar difference in gains for each (13% higher for instructor 1 and 12% for instructor 2).

Table 10

Gains for Each Lab Section, As well As For Comparison and Treatment Groups For All Questions, As well As For Those Questions Relevant To the Treatment Labs

	All Questions			Relevant Questions		
	% Pre	%Post	%Gain	% Pre	%Post	%Gain
Comp 1 ($n = 17$)	29.2	42.7	19.1	30.9	41.9	16.0
Comp 2 ($n = 19$)	38.8	50.5	19.2	38.8	50.0	18.3
Comp all ($n = 36$)	34.3	46.9	19.2	35.1	46.2	17.1
Treat 1 ($n = 21$)	28.6	50.0	30.0	26.2	47.6	29.0
Treat 2 ($n = 19$)	35.4	56.9	33.2	39.5	57.9	30.4
Treat all ($n = 40$)	31.7	53.2	31.4	32.5	52.5	29.6

Beyond the simple gains in treatment or comparison sections, we find that the students that initially had happier feelings toward science (our comparison sections) were more likely to choose what I considered the main distracter question on the content pretest if they did not choose the correct answer. In contrast, those who were in our treatment sections tended to choose the same minor distracter question, rather than being more randomly distributed among the answers. This suggests that for the two populations there is a difference between preconceptions held by the students, and which may be worth exploring further in future work.

INTERPRETATION AND CONCLUSION

In our study, we find that all students feel a little better about science after instruction. In particular, those students in the treatment sections feel they can do science more than prior to class. As this course is directed primarily to non-science students who tend to not be confident in their ability to do science, this is a very positive result. We find that students say that they prefer using real data, and to a lesser extent from our own

telescope. However we find an encouraging result when actually using real data from our own telescope compared to real data that is just given to them without context or a simulation, in that the treatment sections improve in attitude and also in the gains on the post-test. For the post-test, gains in treatment sections were approximately 10% higher than in comparison groups. The increase in understanding of students involved in inquiry-based situations, where learning of science content is improved compared to traditional lecture-based formats (Smetana 2010; Bell, Smetana & Binns 2005; Buck, Bretz & Towns 2008; Sadeh & Zion 2009, 2012). This use of real data is a new addition, showing a greater understanding beyond labs using identical inquiry-based labs with simulations or data that does not seem real. Since data ownership has a weaker effect, it is possible that developing labs utilizing available real archival astronomical data, paying attention to speak specifically to the realness of the data may have the same effect.

I would like to continue this research by developing a few labs to take advantage of archival data. We also plan to continue this work by developing more laboratory exercises using our own telescope, and are working with local K-12 institutions to get them access to our telescope and run similar or modified age-appropriate exercises to introduce students of younger ages to real telescope work, with the goal of instilling interest and enthusiasm for science at younger ages. As a separate project, I would like to continue exploring the differences in incoming understanding that our college students have of the material by probing further the content question answers and which distracter question students choose based on their confidence in science-related activities.

VALUE

This research has been particularly useful for me in a number of ways. This is the first time we have done any survey of our students either pre- or post- instruction on a standardized survey. While the TOAST survey is largely geared toward astronomy concepts which are taught at the K-12 level, it provides a baseline of comparison for our future teaching. I would like to include more complicated and higher level concepts that are being developed by other groups in the future. In addition to the content survey, I am very excited to see that our impact on the attitudes of students was a positive one; one of my main goals in teaching introductory astronomy is that students leave the course feeling more confident about science in general. While these students are not planning to be science majors, an understanding of science is important to society at large – something the students agree with – and having students feel more capable and confident about science is very exciting. The use of data to guide our instruction gives me confidence that those things I thought were working are indeed actually working. This will impact my teaching by making me more enthusiastic toward developing new labs and teaching them to students, knowing that they will feel better about science after instruction.

In addition to my own personal value, the university is interested in knowing the impact of the observatory on students, who will eventually become alumni. This project has been supported by NSF grant #1140385.

REFERENCES CITED

- Adams, W. K., Perkins, K. K., Podolefsky, N. S., Dubson, M., Finkelstein, N. D., & Wieman, C. E. (2006). A new instrument for measuring student beliefs about physics and learning physics: the Colorado learning attitudes about science survey. *Physics Review Special Topics: Physics Education Research*, 2(1), 010101-010115. doi:10.1103/PhysRevSTPER.2.010101
- Anderson, J. R., Greeno, J. G., Kline, P. J., & Neves, D. M. (1981). Acquisition of problem-solving skill. In J.R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 191–230). Hillsdale, NJ : Erlbaum.
- Bailey, J., Slater, S. & Slater, T. (2010). *Astronomy Education Research Primer*. (pp. 99-109). New York, NY: WH Freeman & CO.
- Basey, J. M. & Francis, C. D. (2011). Design of inquiry-oriented science labs: impacts on students' attitudes. *Research in Science & Technological Education*, 29(3), 241-255. doi:10.1080/02635143.2011.589379.
- Bell, R. L., Smetana, L., & Binns, I. (2005). Simplifying inquiry instruction. *The Science Teacher*, 72(7), 30–33.
- Bruck, L. B., Bretz, S. L., & Towns, M. H. (2008). Characterizing the level of inquiry in the undergraduate laboratory, *Journal of College Science Teaching*, 38(1), 52-58.
- Bunterm, T., Lee, K., Kong, J. N. L., Srikoon, S., Vangpoomyai, P., Rattanavongsa, J. & Rachahoon, G. (2014). Do different levels of inquiry lead to different learning outcomes? A comparison between guided and structured inquiry. *International Journal of Science Education*, 36(12), 1937-1959. doi:10.1080/09500693.2014.886347
- Chatterjee, S., Williamson, V. M., McCann, K., & Peck, M. L. (2009). Surveying students' attitudes and perceptions toward guided-inquiry and open-inquiry laboratories. *Journal of Chemical Education*, 86(12), 1427–1432. doi:10.1021/ed086p1427
- Chen, S. (2010). The view of scientific inquiry conveyed by simulation-based virtual laboratories. *Computers & Education*, 55(3), 1123-1130. doi:10.1016/j.compedu.2010.05.009

- Chi, M. T. H., Feltovich, P. J., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, 5(2), 121–152.
- Chin, C., & Brown, D. E. (2000). Learning in science: A comparison of deep and surface approaches. *Journal of Research in Science Teaching*, 37(2), 109–138.
doi:10.1002/(SICI)1098-2736(200002)37:2<109::AID-TEA3>3.0.CO;2-7
- Chinn, C.A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. *Science Education*, 86(2), 175–218. doi:10.1002/sce.10001
- Cohen, E. (1994). *Designing Groupwork Strategies for the Heterogeneous Classroom*. New York: Teachers College Press.
- Edelson, D. C, Gordin, D. N., and Pea, R. D. (1999). Addressing the challenges of inquiry-based learning through technology and curriculum design. *Journal of the Learning Sciences*, 8(3/4), 391-451.
- Elio, R. & Peternela B. Scharf, P. B. (1990). Modeling novice-to-expert shifts in problem-solving strategy and knowledge organization. *Cognitive Science*, 14(4), 579–639. doi:10.1207/s15516709cog1404_4
- Enghag, M. and Niedderer, H. (2008). Two dimensions of student ownership of learning during small-group work in physics. *International Journal of Science and Mathematics Education* 6(4), 629-653. doi:10.1007/s10763-007-9075-x
- Hammer, D., Russ, R., Mikeska, J., & Scherr, R. (2005). Identifying inquiry and conceptualizing students abilities. In Duschl, R. & Grandy, R. (Eds.), *Teaching scientific inquiry* (pp. 138-156). Rotterdam, The Netherlands: SensePublishers.
- Heller, J. I., & Reif, F. (1984). Prescribing effective human problem-solving processes: problem descriptions in physics. *Cognition and Instruction*, 1(2), 177–216.
doi:10.1207/s1532690xci0102_2
- Hume, A. (2009). Authentic scientific inquiry and school science. *Teaching Science*, 55(2), 35-41.

- Hume, A. & Coll, R. (2010). Authentic student inquiry: the mismatch between the Intended curriculum and the student-experienced curriculum. *Research in Science & Technological Education*, 28(1), 43–62.
- Jonassen, D. H. (1994). Thinking technology: toward a constructivist design model. *Educational Technology*, 34(4), 34–37.
- Kirschner, P. A., Sweller, J. & Clark, R. E. (2006). Why minimal guidance during instruction does not work: an analysis of the failure of constructivist, discovery, problem-based, experiential, and inquiry-based teaching. *Educational Psychologist*, 41(2), 75-86. doi:10.1207/s15326985ep4102_1
- Lee, H. S., & Songer, N. (2003). Making authentic science accessible to students. *International Journal of Science Education*, 25(8), 923-948. doi:10.1080/0950069022000017315
- Lehrer, R., & Schauble, L. (2004). Modeling natural variation through distribution. *American Educational Research Journal*, 41(3), 635 - 679.
- Leonard, W. H. and Penick, J. E. (2009). Is the inquiry real? Working definitions of inquiry in the science classroom. *The Science Teacher*, 76(5), 40-43.
- Lin, L., Hsu, Y., & Yeh, Y. (2012). The Role of Computer Simulation in an Inquiry-Based Learning Environment: Reconstructing Geological Events as Geologists. *Journal of Science Education Technology*, 21(3), 370–383. doi:10.1007/s10956-011-9330-3
- Linn, M. C., Songer, N. B., & Lewis, E. L. (1991). Overview: students' models and epistemologies of science. *Journal of Research in Science Teaching*, 28(9), 729–732. doi:10.1002/tea.3660280903.
- Marschall, L., Snyder, G. A., Cooper, P. R., Hayden, M. P., Good, R., Hofman, H. *Project CLEA*. Department of Physics, Gettysburg College, Gettysburg, PA, n.d. Web. 3 Dec. 2014. <<http://public.gettysburg.edu/~marschal/clea/CLEAhome.html>>
- Martin-Hansen, L. (2002). Defining inquiry: Exploring the many types of inquiry in the science classroom. *The Science Teacher*, 69(2), 34–37.

- McDermott, L. C. (1990) *A view from physics*. In *Toward a Scientific Practice of Science Education*, eds. Gardner, M., Greeno, J. G., Reif, F., Schoenfeld, A. H., diSessa, A., and Stage, E., pp. 3-30. Hillsdale, NJ:Lawrence Erlbaum Associates.
- Mikropoulos, T. A., & Natsis, A. (2011). Educational virtual environments: A ten-year review of empirical research (1999–2009). *Computers & Education* 56(3), 769–780. doi:10.1016/j.compedu.2010.10.020
- Nadelson, L. S., Walters, L., & Waterman, J. (2010). Course-integrated undergraduate research experiences structured at different levels of inquiry. *Journal of STEM Education*, 11(1&2), 27- 44.
- National Research Council (NRC). (2000). *Inquiry and the national science education standards*. Washington, DC: National Academies Press.
- Piaget, J. (1969). *The psychology of the child* (H. Weaver, Trans.). New York, NY: Basic Books.
- Pollock, S. (2005). No single cause: learning gains, student attitudes, and the impacts of multiple effective reforms. *AIP Conference Proceedings* 790, 137-140. doi:10.1063/1.2084720
- Power, B. (2012). Enriching students' intellectual diet through inquiry based learning. *Libri*, 62(4), 305–325. doi:10.1515/libri-2012-0024
- Prins, G. T., Bulte, A. M. W., Van Driel, J. H. & Pilot, A. (2009). Students' involvement in authentic modelling practices as contexts in chemistry education. *Res Sci Educ*, 39(5), 681–700. doi:10.1007/s11165-008-9099-4
- Raeside, L., Busschots, B., Waddington, S., Keating, J.G. (2008). An online image analysis tool for science education. *Computers & Education*, 50(2), 547–558. doi:10.1016/j.compedu.2007.09.021
- Redish, E. F., & Steinberg, R. F. (1999) Teaching physics: Figuring out what works. *Physics Today* 52(1), 24-30.

- Roll, I., Holmes, N. G., Day, J., & Bonn, D. (2012). Evaluating metacognitive scaffolding in guided invention activities. *Instructional Science*, 40(4), 691–710. doi:10.1007/s11251-012-9208-7
- Rutten, N., van Joolingen, W. R., van der Veen, J. T. (2012). The learning effects of computer simulations in science education. *Computers & Education*, 58(1), 136-153. doi:10.1016/j.compedu.2011.07.017
- Sadeh, I., & Zion, M. (2009). The development of dynamic inquiry performances within an open inquiry setting: a comparison to guided inquiry setting. *Journal of Research in Science Teaching*, 46(10), 1137–1160. doi: 10.1002/tea.20310
- Sadeh, I., & Zion, M. (2012). Which type of inquiry project do high school biology students prefer: Open or guided? *Research in Science Education*, 42(5), 831–848. doi:10.1007/s11165-011-9222-9
- Salehizadeha, M. R. & Behin-Aeinb, N. (2014). Effects of cooperative learning plus inquiry method on student learning and attitudes: a comparative study for engineering economic classrooms. *European Journal of Engineering Education*, 39(2), 188-200. doi:10.1080/03043797.2013.838542
- Schoenfeld, A. H., & Herrmann, D. J. (1982). Problem perception and knowledge structure in expert and novice mathematical problem solver. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 8(5), 484–494. doi:10.1037/0278-7393.8.5.484
- Schwab, J.J. (1962). *The teaching of science as enquiry*. In *The teaching of science*, eds. Schwab, J. J. and Brandwein, P. F., pp. 3–103. Cambridge, MA: Harvard University Press.
- Smith, C., Maclin, D., Houghton, C., & Hennessey, M. G. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences of epistemological development. *Cognition and Instruction*, 18(3), 349 - 422. doi:10.1207/S1532690XCI1803_3
- Sunal, D. W., Wright, E. L., & Day, J. B. (Eds.) (2004). *Reform in undergraduate science teaching for the 21st century*. Greenwich, CT: IAP Information Age.

- Taasoobshirazi, G., Zuiker, S. J., Anderson, K. T., & Hickey, D. T. (2006). Enhancing inquiry, understanding, and achievement in an astronomy multimedia learning environment. *Journal of Science Education and Technology*, 15(5/6), 383-395. doi:10.1007/S10956-006-9028-0
- Tang, X., Coffey, J. E., Elby, A. & Levin, D. M. (2010). The scientific method and scientific inquiry: tensions in teaching and learning. *Science Education* 94(1), 29-47. doi:10.1002/sce.20366
- Trundle, K. C. & Bell, R. L. (2010). The use of a computer simulation to promote conceptual change: a quasi-experimental study. *Computers & Education*, 54(4), 1078-1088. doi:10.1016/j.compedu.2009.10.012
- Ucar, S., Trundle, K. C. (2011). Conducting guided inquiry in science classes using authentic, archived, web-based data. *Computers & Education* 57(2), 1571–1582. doi:10.1016/j.compedu.2011.02.007
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. In *Interaction between learning and development*, pp. 79-91. Cambridge, MA: Harvard University Press.
- Warren, B., Ballenger, C., Ogonowski, M., & Roseberry, A. S. (2001). Rethinking diversity in learning science: The logic of everyday sense-making. *Journal of Research in Science Teaching*, 38(5), 529-552. doi:10.1002/tea.1017
- White, B. Y., & Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: making science accessible to all students. *Cognition and Instruction*, 16(1), 3-118. doi:10.1207/s1532690xci1601_2
- Wu, H. & Pedersen, S. (2011). Integrating computer- and teacher-based scaffolds in science inquiry. *Computers & Education* 57(4), 2352-2363. doi:10.1016/j.compedu.2011.05.011

APPENDICES

APPENDIX A
IRB APPROVAL

INSTITUTIONAL REVIEW BOARD

For the Protection of Human Subjects
FWA 00000165



960 Technology Blvd. Room 127
c/o Immunology & Infectious Diseases
Montana State University
Bozeman, MT 59718
Telephone: 406-994-6783
FAX: 406-994-4303
E-mail: cherylj@montana.edu

Chair: Mark Quinn
406-994-5721
mquinn@montana.edu
Administrator:
Cheryl Johnson
406-994-6783
cherylj@montana.edu

MEMORANDUM

TO: Kisha Delain and Eric Brunsell
FROM: Mark Quinn, Chair *Mark Quinn CJ*
DATE: September 26, 2014
RE: "The Effect of Ownership of Data on Student Attitudes in Introductory Astronomy" [KD092614-EX]

The above research, described in your submission of September 26, 2014, is exempt from the requirement of review by the Institutional Review Board in accordance with the Code of Federal regulations, Part 46, section 101. The specific paragraph which applies to your research is:

- (b) (1) Research conducted in established or commonly accepted educational settings, involving normal educational practices such as (i) research on regular and special education instructional strategies, or (ii) research on the effectiveness of or the comparison among instructional techniques, curricula, or classroom management methods.
- (b) (2) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures or observation of public behavior, unless: (i) information obtained is recorded in such a manner that human subjects can be identified, directly or through identifiers linked to the subjects; and (ii) any disclosure of the human subjects' responses outside the research could reasonably place the subjects at risk of criminal or civil liability, or be damaging to the subjects' financial standing, employability, or reputation.
- (b) (3) Research involving the use of educational tests (cognitive, diagnostic, aptitude, achievement), survey procedures, interview procedures, or observation of public behavior that is not exempt under paragraph (b)(2) of this section, if: (i) the human subjects are elected or appointed public officials or candidates for public office; or (ii) federal statute(s) without exception that the confidentiality of the personally identifiable information will be maintained throughout the research and thereafter.
- (b) (4) Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available, or if the information is recorded by the investigator in such a manner that the subjects cannot be identified, directly or through identifiers linked to the subjects.
- (b) (5) Research and demonstration projects, which are conducted by or subject to the approval of department or agency heads, and which are designed to study, evaluate, or otherwise examine: (i) public benefit or service programs; (ii) procedures for obtaining benefits or services under those programs; (iii) possible changes in or alternatives to those programs or procedures; or (iv) possible changes in methods or levels of payment for benefits or services under those programs.
- (b) (6) Taste and food quality evaluation and consumer acceptance studies, (i) if wholesome foods without additives are consumed, or (ii) if a food is consumed that contains a food ingredient at or below the level and for a use found to be safe, or agricultural chemical or environmental contaminant at or below the level found to be safe, by the FDA, or approved by the EPA, or the Food Safety and Inspection Service of the USDA.

Although review by the Institutional Review Board is not required for the above research, the Committee will be glad to review it. If you wish a review and committee approval, please submit 3 copies of the usual application form and it will be processed by expedited review.

APPENDIX B
ATSI ADDITIONAL QUESTIONS

21. Define "science" in one or two sentences.
22. On what personal experience is your definition of science based?
23. What is your college major (or current area of interest if undecided)?
24. What class are you?
Freshman
Sophomore
Junior
Senior
Graduate Student
Other
25. Which of the following describe your ethnic background? (Select all that apply)
African-American
Asian-American
Native American
Hispanic-American
African (not American)
Asian (not American)
White, non-Hispanic
Other
Decline to answer
26. What is your gender?
Female
Male
Other
Decline to answer
27. Which best describes your home community (where you attended high school)?
Rural
Small town
Suburban
Urban
Not in the USA

28. Which best describes the level of difficulty you expect from this course?

Extremely difficult for me

Difficult for me

Unsure

Easy for me

Very easy for me

29. How good at math are you?

Very poor

Poor

Average

Good

Very good

30. I expect that this course is going to be:

Drudgery

Tolerable

Unsure

Just ok

Lots of fun

APPENDIX C
TOAST ADDITIONAL QUESTIONS

1. Suppose that researchers have discovered a mathematical relationship between a star's mass, temperature, and size. If we can directly measure the star's temperature and size, which statement about its mass is the most correct?
 - a. Because we cannot directly measure the star's mass, we have no way of knowing its true mass.
 - b. Knowing the temperature and size will give us a definitive measure of its mass.
 - c. Knowing the temperature and size will let us calculate a mass, but it may not be its true mass.
 - d. Knowing the temperature and size will give information about its mass, but not an exact number.

2. An image of a galaxy shows that it has very blue color. This means that
 - a. It contains young blue stars and old red stars
 - b. It contains old blue stars and young red stars
 - c. It only contains young blue stars
 - d. It contains old and young blue stars, along with old red stars
 - e. It contains old and young red stars, along with young blue stars

3. If the sun were to somehow increase in mass, in order to stay in the same orbit, a planet would have to
 - a. Increase in speed
 - b. Decrease in speed
 - c. It could not stay in the same place; it would spiral in
 - d. It could not stay in the same place; it would be flung outward

APPENDIX D

SURVEY 1

H-R Diagram Survey

This survey is to help us improve the H-R diagram lab sequence. Please take a moment to tell us what you think! (You do not have to write your name but it may help with followup questions. The other lab instructor will read your comments, not me.)

Thinking just about the data and graph:

Is it important to you to have real data?

Not at all important 1	Not really important 2	Neutral 3	Somewhat important 4	Really important 5
------------------------------	------------------------------	--------------	----------------------------	--------------------------

Is it important for you to have data from our own telescope?

Not at all important 1	Not really important 2	Neutral 3	Somewhat important 4	Really important 5
------------------------------	------------------------------	--------------	----------------------------	--------------------------

How challenging is it for you to graph by hand?

Very challenging 1	Somewhat challenging 2	Neutral 3	Easy 4	Very Easy 5
--------------------------	------------------------------	--------------	-----------	----------------

How challenging is it to use Excel?

Very challenging 1	Somewhat challenging 2	Neutral 3	Easy 4	Very Easy 5
--------------------------	------------------------------	--------------	-----------	----------------

Which did you like better and why?

This question was verbally clarified to refer to the data itself.

APPENDIX E

SURVEY 2

Jupiter Lab Survey

Note: Comparison group wording in brackets

How challenging was the first lab, where you designed the observing plan?

Very Challenging	Somewhat Challenging	Neutral	Easy	Very Easy
1	2	3	4	5

How well do you think you understand about the capabilities of the telescope?

Not at all	A little	Neutral	A lot	Extremely Well
1	2	3	4	5

How challenging was the second lab, where you did the data analysis?

Very Challenging	Somewhat Challenging	Neutral	Easy	Very Easy
1	2	3	4	5

How important is [would] using real telescope data [be]?

Not <u>at all</u> Important	Not really Important	Neutral	Somewhat Important	Really Important
1	2	3	4	5

How would [did] using a simulation compare?

Just as Interesting	A little less Interesting	Neutral	A lot less Interesting	Not <u>at all</u> Interesting
1	2	3	4	5

Typically astronomers design a set of observations and then have to wait to implement them. Is this type of authentic experience important to you?

Not <u>at all</u> Important	Not really Important	Neutral	Somewhat Important	Really Important
1	2	3	4	5

Given the understanding of how the telescope and camera work you have developed in this course, how confident are you that you could design an observing project if you wanted to?

Really Confident	Somewhat Confident	Neutral	Not very Confident	Not at all Confident
1	2	3	4	5

If you could use the telescope to figure out some project, what kinds of projects would you be interested in?

Rank order the labs in terms of how interesting you found them:

- _____ Going to the observatory (leave blank if you went to public night)
- _____ Measuring Jupiter's Mass (both labs together)
- _____ Galaxies
- _____ H-R Diagram (with telescope data [web data])

APPENDIX F
INTERVIEW QUESTIONS

Science Content Question:

If you were to look at this picture, what would you be able to tell me about the stars in that galaxy?

What kind of stars does it contain?

How would you know if a star is young or old?

Are all blue stars young? Are all red stars old?

General Course Attitudes: [reference ATSI answers as needed]

Let's just talk about how you felt the class went. How did you feel about taking astronomy or any general science versus how you felt after the class.

Is there anything else you want to tell us that would help us with the labs? Things you liked, didn't like, hated?

Real Data: [reference surveys as needed]

The other thing we're interested in was the idea of using real data and using our telescope.

Can you just kind of talk a little bit about having real data versus not real data?

What about real data from "wherever" versus our own telescope?

There are sort of two pieces of that, one is that the data from the first one, it was real data but it was from somewhere else and it was just a table of numbers, versus actually using our image. Can you maybe talk about that a little bit?

Do you feel like in terms of learning it [would/wouldn't have] mattered? Do you think it would affect how much you learned?

Authentic Experience: [reference surveys as needed]

For the Jupiter lab, we had you design an observing plan; and then the telescope did it and then you analyzed the data. That really is a long, involved, "really this is how we do it in astronomy" lab.

We could do that lab using a simulation and sort of all as one big lab, without using our telescope. How would that compare?

Or, if you submitted the plan to the telescope and then obviously if the weather continued to bad, we just used a prior semester's data? How would that compare to using your own data?

Can you see any benefits to doing it at one time?

What about the big gap? Could you talk about how that impacted you?

Do you like that authentic process?

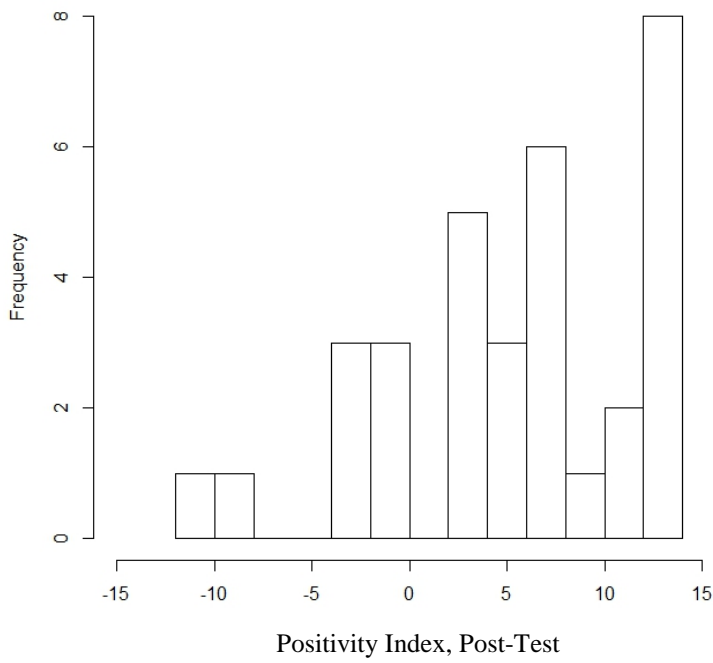
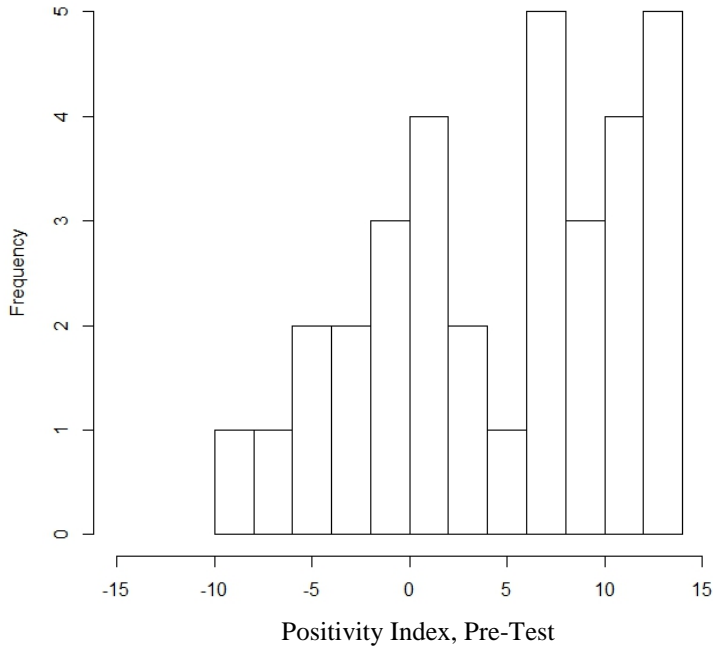
Can you see a long term project where this kind of thing might occur? Can you make any parallels with your major/future career?

Ranking of labs: [reference survey]

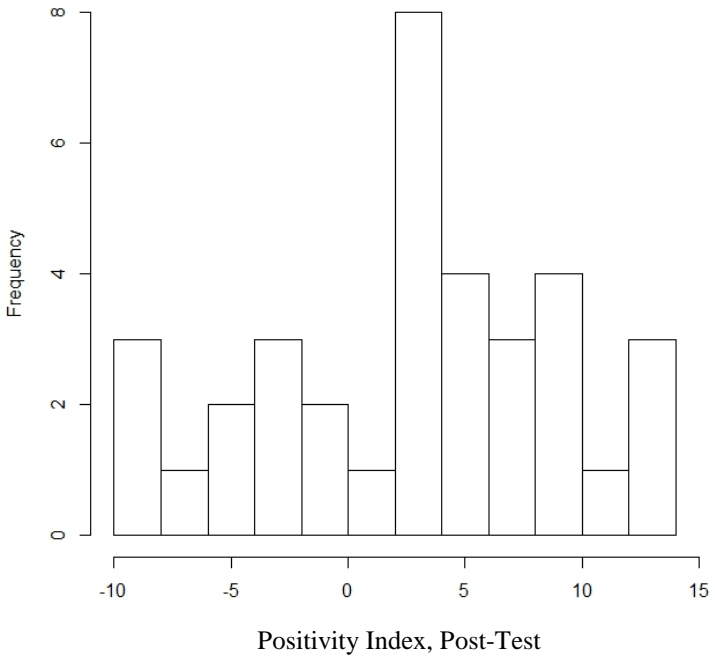
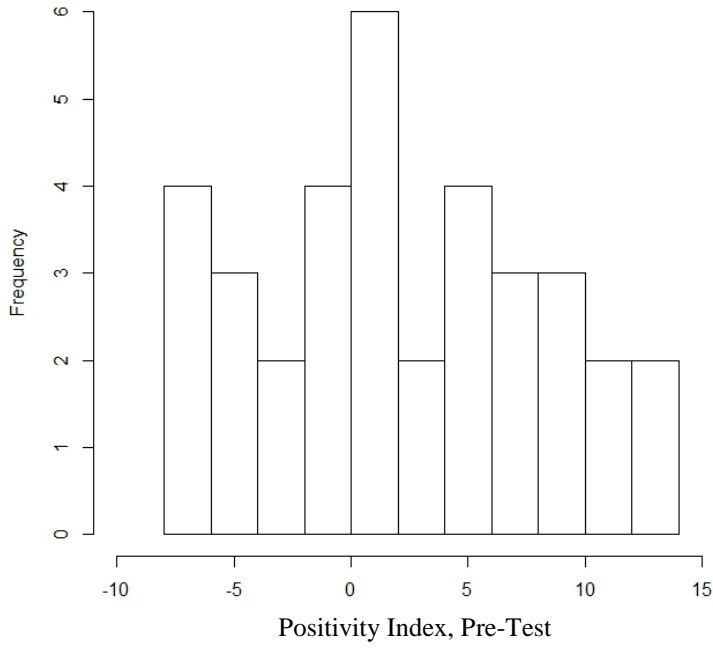
Can you maybe talk a little bit about the ranking and what you liked?

APPENDIX G
HISTOGRAMS OF ATTITUDE SURVEY

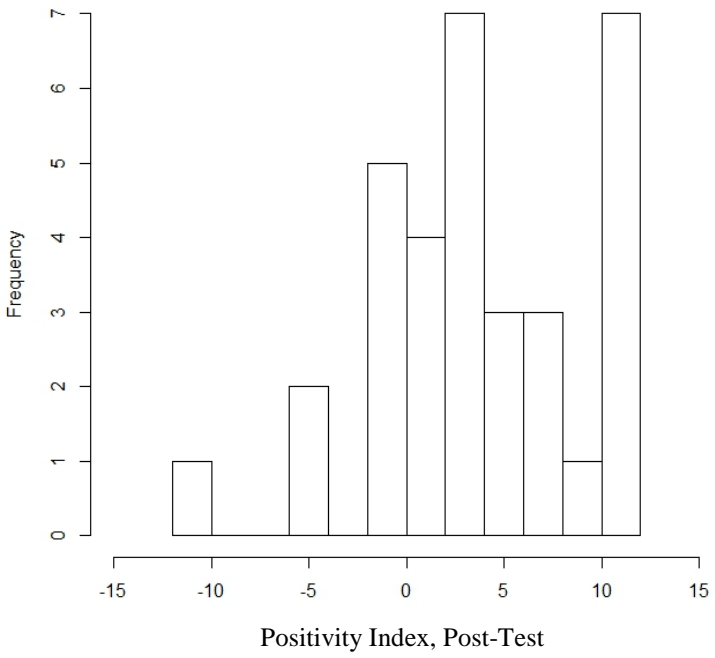
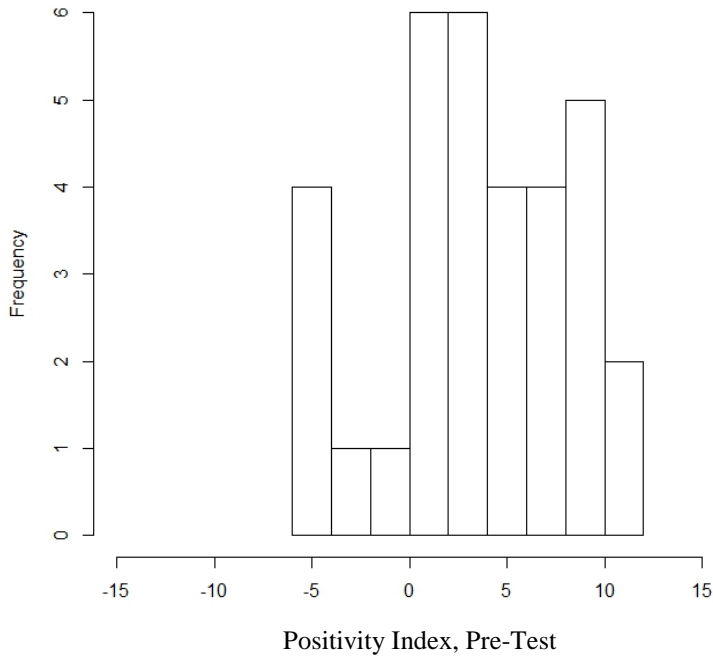
Comparison Group Positivity Index For “How Do Students Feel About Science?”



Treatment Group Positivity Index For “How Do Students Feel About Science?”



Comparison Group Positivity Index For “Do Students Think They Can Do Science?”



Treatment Group Positivity Index For “Do Students Think They Can Do Science?”

