COMPARING ASSESSMENT METHODS IN UNDERGRADUATE STATISTICS COURSES

by

Sarah E. Baxter

An executive position paper submitted to the Faculty of the University of Delaware in partial fulfillment of the requirements for the degree of Doctor of Education in Educational Leadership

Spring 2017

COMPARING ASSESSMENT METHODS IN UNDERGRADUATE

STATISTICS COURSES

by

Sarah E. Baxter

Approved: ____

Ralph P. Ferretti, Ph.D.
Director of the School of Education

Approved: ____

Carol Vukelich, Ph.D.
Dean of the College of Education and Human Development

Approved: ____

Ann L. Ardis, Ph.D.
Senior Vice Provost for Graduate and Professional Education

I certify that I have read this executive position paper and that in my opinion it meets the academic and professional standard required by the University as an education leadership portfolio for the degree of Doctor of Education.

Signed: ____

Amanda M. Jansen, Ph.D.
Professor in charge of executive position paper

I certify that I have read this executive position paper and that in my opinion it meets the academic and professional standard required by the University as an education leadership portfolio for the degree of Doctor of Education.

Signed: ____

Douglas A. Archbald, Ph.D.
Member of executive position paper committee

I certify that I have read this executive position paper and that in my opinion it meets the academic and professional standard required by the University as an education leadership portfolio for the degree of Doctor of Education.

Signed: ____

Dawn M. Berk, Ph.D.
Member of executive position paper committee

I certify that I have read this executive position paper and that in my opinion it meets the academic and professional standard required by the University as an education leadership portfolio for the degree of Doctor of Education.

Signed: ____

Joan L. Buttram, Ph.D.
Member of executive position paper committee

ACKNOWLEDGMENTS

Amanda Jansen, Ph.D., for her continuous advice, patience and support over the past many years.

This manuscript is dedicated to:

My husband, Jason, for his unconditional love and support through every semester along the way.

My children, Andrew, Daniel, and Ryan for providing needed diversions from all things academic. You remind me every day that the most rewarding job of all is being a mom.

My parents, Dr. Milton and Sandra Loyer, who encouraged my love of reading and school from day one and have supported me through every step of my journey.

TABLE OF CONTENTS

Appendices

LIST OF TABLES

ABSTRACT

The purpose of this study was to compare undergraduate students' academic performance and attitudes about statistics in the context of two different types of assessment structures for an introductory statistics course. One assessment structure used in-class quizzes that emphasized computation and procedural fluency as well as vocabulary recognition in three sections during the semester (quiz-based). The other assessment structure used projects that students conducted outside of class and which involved students generating and analyzing their own data for two sections during the semester (project-based). All five sections were assessed using the same midterm and final exam.

Findings show that while students performed similarly on the midterm exam, students in quiz-based sections performed higher, on average, than students in project-based sections on the final exam. There were few significant differences in attitudes, as measured through a quantitative survey, at the end of the semester, with students in project-based sections scoring higher in a few instances. Overall, very few differences or changes in attitudes were found in this study across all groupings.

Results differed by learning environment, according to a post-hoc analysis. Two sections (one each of a project-based and quiz-based assessment) were taught in a hybrid learning environment. Three sections (one project-based and two quiz-based) were taught in a face-to-face environment. Students in a face-to-face environment had higher final exam scores than students in hybrid sections. In addition, students in face-to-face sections tended to have more positive attitudes. Future research should

focus on assessment method structure in a common learning environment to remove the complicating factor of learning environment found in this study. In addition, learning environment can be studied separately in terms of how it affects both academic performance and attitudes in statistics.

Chapter 1

INTRODUCTION

In recent years a change in assessments has occurred in mathematics courses. This is evident at the elementary and secondary levels, as the introduction of the Common Core State Standards seeks to help students understand concepts and engage in reasoning along with learning to execute procedures. On its webpage, there is repeated emphasis of the importance of conceptual understanding behind mathematical rules (Core Standards, (n.d.)). This demonstrates the commitment of those who developed the Common Core standards to help students think deeply about the mathematical procedures they are performing and make sure they are understanding the meanings underlying the procedures. A shift in standards for students' learning necessarily leads to a shift in assessments designed to measure these different learning outcomes.

This shift in assessments can also be seen at the college level where steps are being taken to encourage literacy and understanding to complement the development of computation and procedural skills. One national published report on this topic is the Guidelines on Assessment and Instruction in Statistics Education (GAISE) (Garfield et al., 2005). Published in 2005, the GAISE College Report provides the following six recommendations for instructors looking to improve their undergraduate statistics course:

1. Emphasize statistical literacy and develop statistical thinking
2. Use real data
3. Stress conceptual understanding, rather than mere knowledge of procedures
4. Foster active learning in the classroom
5. Use technology for developing conceptual understanding and analyzing data
6. Use assessments to improve and evaluate student learning

The sixth recommendation of this report focuses on assessment. Instructors were encouraged to look for new ways of assessing students that went deeper than computation and procedure, toward understanding. In the process of considering new methods of assessment, however, instructors need to make sure that students *can* still perform the necessary computations to solve problems.

In this EPP, I am investigating undergraduate statistics courses that were designed with two different approaches to assessment. Some sections were assessed through the use of in-class quizzes and tests that emphasized vocabulary and computational and procedural fluency. I refer to these sections as “quiz-based” sections. In contrast, other sections were given a series of mini-projects, as well as one semester-long project, to assess their understanding of the material. I refer to these sections as “project-based” sections. All sections were given the same midterm and final exam, which focused primarily on procedural fluency. I will compare students in these sections on two factors:

- Academic performance as measured by midterm and final exam
- Attitudes toward statistics as measured through (1) an attitudes inventory scale at the beginning and end of the semester, and (2) a course evaluation feedback form designed by the instructor of the course (myself)

Research Questions

Students in college math courses are most often assessed by measures such as in-class quizzes and tests that are comprised of mathematical tasks that capture procedural fluency (Garfield et al., 2005). The college where I teach is no exception. Statistics, although a broader type of math course, is no exception, either. In fact, at my college, I am the only instructor to have attempted a project-based assessment structure for our statistics course. The majority of instructors use quizzes and tests that emphasize computation and procedural fluency to measure student knowledge. Some instructors do include group work and activities as part of their assessment structure, but these learning experiences count for minimal credit toward the final grade in the course. The learning outcomes for introductory statistics which have been designed by the New Jersey Council of Community Colleges (NJCCC) and are mandatory for inclusion on our syllabi are written in ways that align with traditional measures. These learning outcomes are as follows:

- Compute measures of descriptive statistics.
- Apply basic rules of probability (binomial, conditional, addition, etc).
- Solve problems involving probability distributions.

- Formulate conclusions through inference.
- Analyze bivariate data through linear correlation and regression.
- Apply basic statistical concepts.

Although these learning outcomes could also be assessed through alternative ways such as projects, group work, in-class experiments, etc., most instructors choose more standard measuring assessments like in-class quizzes and tests which emphasize computation and procedural fluency.

Currently at our college, instructors design their own assessment items for final exams; however, many community colleges in New Jersey have moved toward using a common final exam across all statistics sections. The benefits of common final exams are two-fold. Colleges can provide a measure of consistency across sections by requiring all sections to be tested using the same final exam. In addition, colleges can more easily gather and analyze data on the learning outcomes designed by the NJCCC. Our college is hoping to move toward a common final exam for all statistics sections. This would allow us to provide more consistency across sections as well as put ourselves in a better position to more easily gather and analyze data on student success in achieving learning outcomes. This common final exam would be largely comprised of multiple choice questions and focus heavily on computation, procedural fluency, and vocabulary recognition. While GAISE suggests that assessments that address conceptual understanding are more valuable, most colleges find that using a multiple-choice format with calculation and procedural emphases is easier to use, grade, and correlate data from across sections. Therefore, it is essential that any instructor using alternative means like projects and experiments to assess student knowledge throughout the semester be certain that these students could also perform well on a procedural-based final exam that students take in other sections, so that students have similar opportunities to learn across sections.

At first glance, it would be easy to be skeptical about students in project-based sections being able to perform at least as well as students in quiz-based sections on this common midterm and final exam, which emphasize computation and procedural fluency. Wouldn't having had multiple opportunities for students in quiz-based sections to practice these skills through in-class quizzes give them an advantage on the midterm and final exam? Wouldn't the *lack* of these opportunities through in-class quizzes for students in project-based sections hurt them when they took the midterm and final exam? My conjecture is that students in project-based sections *will* be able to perform at least as well as students in quiz-based sections on this common midterm and final exam. I have several reasons to support this conjecture. The notes I design for students to complete in class are set up with examples nearly identical in format to computational and procedural problems they will see on the midterm and final exam. Therefore, students in project-based sections, though not completing in-class quizzes, will still have completed problems in the same format during our class sessions together. By having access to and studying these notes, students in project-based sections should have the material and practice they need to perform at the same level on the midterm and final exam as students in the quiz-based sections. I also believe that students who complete projects throughout the course are able to reinforce skills in computation and procedure in a way that is similar to students in quiz-based sections who routinely practice these skills on in-class quizzes. Using their own data, these students in project-based sections practice the same computations and procedures that students in quiz-based sections do in a quiz-format. Therefore, regardless of assessment method, students in all sections are practicing the computations and procedures they will be evaluated on with the midterm and final exam. Finally, I believe that because of the extra interest created by students completing projects based on a topic of their choice, that this higher interest level will lead students to make connections between content in the classroom and the project content again, in ways that are similar to students performing these calculations and procedures on an in-class quiz. Being exposed to the calculations and procedures in new ways through the collection of their own data, could lead them to make even stronger intellectual connections between course content and application in a given problem. In summary, it is important that students who experience a project-based course, at the least, do not have impoverished opportunities to learn; based on the reasons above, I believe that students in project-based sections are indeed *not* at a disadvantage. I conjecture that students in project-based sections are able to learn the material just as well as students in quiz-based sections, and that this will lead them to be able to perform at a similar level on the common midterm and final exam which emphasize computation and procedural fluency as students in quiz-based sections. It is possible, and even hopeful, that students in project-based sections actually have even richer opportunities to learn the course material.

Investigating whether students in project-based sections can perform as well as students in quiz-based sections on standardized exams is worth doing. However, if there are additional advantages to being in project-based sections, I would like to explore this possibility and identify them. An example of one such advantage is the potential for student attitudes in project-based sections to be better than in quiz-based sections. When participating in projects throughout the semester, students have the opportunity to plan and carry out small research projects multiple times. For the majority of these research projects, students are free to choose the type of data they wish to collect. I hope the experience students receive in actively being involved in statistics in a more hands-on way gives them more confidence in their ability to perform statistical tasks. I also hope students in project-based sections have better attitudes about statistics in terms of the value of statistics in their lives, their interest in the subject matter, and, in general, liking the subject of statistics. Therefore, in addition to putting forth a conjecture about students in project-based sections being able to perform academically at a level similar to students in quiz-based sections, I also conjecture that students in project-based sections will exhibit more positive attitudes toward statistics at the end of the course than students in quiz-based sections.

These conjectures lead directly to my formal research questions. I am seeking to investigate both academic performance and attitudes in this study. Therefore, I have created research questions for each of these components to help investigate the conjectures I have put forward. My research questions for this study are as follows:

1. How does students' performance compare between those who enrolled in project-based sections of statistics and those enrolled in quiz-based sections?

    1a. How does the performance of students in the project-based sections on the midterm and final exam compare to students in the quiz-based sections?

    1b. What kind of correlation exists between grades on quizzes during the semester and grades on the midterm and final exam for students in the quiz-based sections? What kind of correlation exists between grades on projects during the semester and grades on the midterm and final exam for students in the project-based sections?

2. How do students' attitudes toward statistics compare between those enrolled in project-based sections of statistics and those enrolled in quiz-based sections?

    2a. How do student attitudes in project-based sections, as reported on a quantitative rating scale, compare to students' attitudes in quiz-based sections at the end of the semester?

    2b. How have student attitudes, as reported on a quantitative rating scale, changed over the course of the semester?

2c. What kind of differences in students' attitudes exist in the feedback provided on a course evaluation form for students in the quiz-based sections vs. students in the project-based sections?

Chapter 2

LITERATURE REVIEW

Statistics is an emerging discipline. While initially reserved only for students of mathematics, in the 1960s it began to seep into a broader spectrum of college courses. However, at that time, statistics courses emphasized probability, which required students to be mathematically advanced enough to understand complex calculations. In the 1970s, exploratory data analysis evolved as a statistics area separate from probability models. This had the effect of allowing students to access statistical ideas and methods without tedious calculations (Garfield et al., 2005).

From 1995 to 2010, enrollment figures for introductory statistics courses at four-year colleges and universities increased from 164,000 to 312,000 students, a 90% increase (Doehler et al., 2013). A main reason for this increase is the number of undergraduate programs that began to require introductory statistics as a part of their major. The development of the Advanced Placement Statistics course in 1997 provides further evidence of the growing interest in and demand for statistical knowledge. In this first year offering, 7500 students took the AP statistics exam. By 2004, more than 65,000 high school students took the AP statistics exam each year. In 2013, this number reached over 151,000 students world-wide. (Total Registration, (n.d.)).

Today statistics is a requirement for most undergraduate majors. Its increased presence in education is seen beginning at the primary grade level. My7-year old son is reading and interpreting graphs, performing experiments and making predictions about events like the likelihood of heads vs. tails in a series of coin tosses. Statistics and probability are required material for the mathematics curriculum from K-12 (Groth, 2008) and in fact, the GAISE project produced a second report for the instruction of students at the K-12 level (Franklin et al., 2007).

It is an exciting time to be a statistics instructor. With the emergence of social media and technology available 24 hours a day, students are exposed more and more to information that they need to process, interpret, and understand. Statistics provides a lens for them to critically evaluate information they see and not just passively accept all information as true or reliable.

It can also be a frustrating time to be a statistics instructor. Statistics is a course with a bad reputation – often feared by college students and one that creates high levels of anxiety (Ben-Zvi & Garfield, 2008; Curran et al., 2013; Lesser et al., 2013). The level of math required can be daunting for students and the rules and ideas in statistics are complex. The required emphasis on reasoning and interpretation – often different than in most traditional math courses – can be met with resistance from students (Ben-Zvi & Garfield, 2008). In addition, navigating the ever-growing options of available resources through online technology is time-consuming. Keeping an undergraduate statistics course fresh, current, and relevant is a semester-by-semester challenge.

The GAISE College Report, published in 2005, provides six recommendations for instructors looking to improve their undergraduate statistics course and help meet the challenges presented above. The sixth of these recommendations, listed here, is the focus of this EPP: Use assessments to improve and evaluate student learning.

My conjecture is that project-based assessments provide a formative experience that could allow students to learn at least as much as students who do not have these assessments, as well as develop more positive dispositions toward statistics. I conjecture this because the experience of engaging in statistics outside the classroom by choosing topics and research questions of personal interest, collecting data, and analyzing that data allows students to see a relevance and feel an excitement for the material that may be absent by just reading about statistics and solving textbook problems in class. Two main topics for this EPP are using alternative assessments in statistics and addressing student attitudes in statistics. While there are many examples of alternative learning opportunities and alternative learning environments in statistics, I found no literature directly investigating the effects of alternative assessment approaches in a statistics course. The literature does address student attitudes in statistics, and in some cases, these attitudes are linked with alternative learning opportunities and/or academic performance. In this section I will address some of the prior research that has been done on these topics.

Several studies show a connection between alternative learning opportunities in statistics courses and both academic performance and attitudes. I define alternative learning opportunities as tasks students are asked to complete or participate in that are different than individually taken quizzes or tests taken during the class period. They have often been linked to a more active learning approach in the classroom. One such approach was a workbook approach in which students read a chapter and answered a set of questions prior to coming to class. Answers were submitted online prior to class and feedback was provided to students prior to class. While there was a short lecture time at the beginning and end of class, the majority of class was devoted to students working, often in small groups, in their workbooks (Carlson and Winquist, 2011). Results at the end of the semester showed that not only did students in this workbook setting have more positive attitudes toward statistics than students who were not in the workbook setting, but that these positive attitudes were correlated with high performance. Autin et al. (2013) compared students taught in a traditional, lecture-based method with students taught in a student-centered environment that focused on problem-solving in the classroom. Test scores at the end of the semester were similar, but qualitative student reflections about what was learned were deeper from students in the non-traditional environment, whose comments included more ideas about understanding and connecting the material in the course to real-life problems, feeling an excitement for the material, and seeing the relevance of statistics in the world around them. Similarly, Prins (2009) found a student-centered approach to learning to be more challenging than the traditional teacher-centered approach, but also found it more enjoyable and more rewarding. Having students work on problems together during class promoted conversation between peers, and not just teacher to student. Wilson (2013) applied the flipped classroom idea to her statistics class, shifting the majority of knowledge acquisition outside of the classroom to allow time in the classroom for more interactive learning opportunities. In comparing classes taught with the flipped design to traditional courses she also taught, she found that not only were students' attitudes toward the class improved in the flipped class, their exam scores were also significantly higher. While these studies highlight alternative learning opportunities, I found no literature investigating the effects of altering the assessment approach in an undergraduate statistics course.

Using student-generated data in class, whether gathered about the students themselves or by students in regards to a research question they design, has also been shown to have positive impact on student attitudes (Neumann et al., 2010; Woodward & McGowan, 2012; Neumann et al., 2013). In one example, students were presented with a series of questions to answer about themselves – some open-ended and some forced response – which included both qualitative and quantitative responses (Neumann et al., 2010). This survey had the effect of increasing student interaction as well as allowing the instructor and students to get to know each other. At the end of the course, interviews were conducted with students and the most common categories of response (seen at least 45% of the time in all interviews) were that the survey created interest, increased relevance, and helped understanding (Neumann et al., 2010; Neumann et al., 2013). Additional categories coded were fun activity, increased participation, reduced negative mood, and allowed for a different approach to a course. In all cases, the use of real data added value to the course by proving intrinsically interesting (Neumann et al., 2010; Woodward & McGowan, 2012; Neumann et al., 2013). Libman (2010) and Warton (2007) both found that using students' own data, as gathered by the students themselves through individually-designed research questions, resulted in students asking countless questions in class. The assignments came alive for the students and they were motivated beyond normal classroom levels to interpret and understand their data. Libman (2010) stated that, “on the whole, there were quite a few students who did not stop at doing the standard statistical procedures, but went further and related them to background material and made an effort to make sense of their outcomes” (p. 10). Libman refers to this as Learner Empowerment. He describes this as the student directing the class time with various questions about their data sets. The dialogue that occurred between teacher and student were often based on questions students brought to class as they were dealing with their own data sets.

In the majority of examples above, the idea of attitude change was measured qualitatively through student feedback at the end of the course or else addressed in related categories like increasing learner engagement and interest, and learner empowerment. The changes seemed due to an emphasis in the classroom on active participation by students in the process of asking and answering research questions and a student-centered environment that allowed students to gather, discuss, and analyze data of their choice. For my EPP, in addition to using student feedback provided on course evaluation forms, I would like to use a quantitative survey to measure attitude change because this provides a way of capturing the degree of change students experience during the course of the semester.

Why is studying attitudes in statistics important? With the current shift in revised learning goals and assessment methods that focus on literacy and understanding discussed earlier, an increased interest in student attitudes toward statistics has emerged. Schau & Emmioglu (2012) suggest that student attitudes have been found to be at least as important a learning outcome as knowledge of the subject matter itself. In statistics, they claim that students who leave their statistics courses with negative attitudes are unlikely to ever apply or use the material learned in that course. The majority of non-STEM major undergraduates students will likely take only one introductory course in statistics; this makes the experience of that introductory statistics course critical for their future interactions with and interest in statistics.

The *Survey of Attitudes Toward Statistics-36* [SATS-36] (Schau, 2003) is one of the most widely used quantitative assessment on attitudes in introductory statistics courses used in the United States. Several studies have used the SATS-36 to evaluate student attitudes in introductory statistics. Unfortunately, many of those studies investigating attitudinal change over the course of the semester of introductory statistics have typically shown that, on average, student attitudes either did not change or became more negative by the end of the semester (Evans, 2007; Sizemore & Lewandowski, 2009; Schau & Emmioglu, 2012). Some of these studies examined only the issue of attitude change. Others linked attitude assessment with certain methods. Schau and Emmioglu (2012) looked only at attitude change over the course of the semester. Their research used data from many institutions and included over 100 sections and 2200 students in total. No information is provided about what type of assessment methods were used in these sections. They state that when attitude changes have been found, those changes have often occurred in contexts where there have been changes in instructional design. Schau and Emmioglu (2012) recognize the need to identify what type of instructional and course characteristics are related to desired attitude changes and then shift our courses to mirror those.

Student-centered learning, where the student holds an active role in the learning process, has often been associated with positive attitudes. By using discovery projects in a course where students plan and conduct their own research studies, some positive change in attitudes has been identified (Spence et al., 2011; Bailey et al., 2013). In a large experimental study currently in process, preliminary results show that students in treatment sections (with the discovery project component) score higher than students in control sections (without the discovery project component) on measures of self-efficacy for data collection, and in content knowledge for identifying which statistical analysis is appropriate for a given scenario (Bailey et al., 2013). Carnell (2008) completed a similar study comparing two sections of introductory statistics. The difference between the sections was that students in one section completed a student-designed data collection project and students in the other section did not. While statistically similar in attitudes at the beginning of the course, students in the project section did *not* have more positive responses than students in the section without the project. In fact, each group showed a significant decrease on the interest subscale over the course of the semester. Yet in other cases, researchers found that improvement in students' attitudes *did* occur, particularly in courses where the learning environment consisted of a number of constructivist elements (Mvududu 2003).

In summary, the literature on student attitudes in statistics courses shows mixed results. The topic of attitudes is confounded by complex interactions between achievement and affect (Spence et al., 2011) and results differ greatly by instructor (Bailey et al., 2013). In addition, there are limitations mentioned in several of the studies. Carnell (2008) used only a small sample of students in her study and also admits that having just one project during the semester may not have been enough to facilitate changes in student attitudes. In other cases, no investigation of instructional methods or assessment methods across sections was done in conjunction with the study on attitudes (Bailey et al., 2013).

I hope to address some of these limitations in my EPP. For example, I will control for the factor of instructor effect because I will be teaching all the sections in my study. In addition, I am using a larger sample of 150 students across five sections. Also, instead of studying just attitude change (Bailey et al., 2013), I will be attempting to link attitude changes with assessment method. I will explicitly define and distinguish differing assessment methods across sections. Instead of having just one project during the semester (Carnell, 2008), I will have a total of 5 projects in the project-based sections.

Finally, attitude change in the literature has sometimes been linked more often with alternative teaching approach and instructional methods than with alternative assessments (Neumann et al., 2010; Carlson & Winquest, 2011; Autin et al., 2013; Neumann et al., 2013; Wilson, 2013). A teaching approach involves the ways a teacher presents material in the classroom, whereas assessments have to do with the way student learning is quantitatively measured, often for the purpose of providing a grade for that student. A teacher may use active learning in the classroom as a teaching approach, but this might not be related to what she records for the purposes of student grades (designed assessments she uses to measure student learning). In other words, a teacher could fall into four groups: (1) use a lecture-based approach in the classroom and a assess students through in-class quizzes, (2) use an active learning approach in the classroom and assess students through in-class quizzes, (3) use a lecture-based approach in the classroom and assess students in ways other than in-class quizzes (this could include projects), or (4) use an active learning approach in the classroom and assess students in ways other than in-class quizzes. It is unclear in most of the studies above just how student knowledge was measured and graded during the semester and if teachers in these courses, despite using active learning in the classroom, still assessed student learning using an in-class quiz format. What I hope to do in my EPP is more directly connect alternative methods of *assessment* (not teaching approach) with both achievement scores as well as attitude data. In fact, my approach of teaching in the classroom is a more traditional, lecture-based approach. Therefore, my EPP will differ from much of the previous literature because I hope to see attitude change not as a result of a different *teaching approach* (the ways the instructor chooses to introduce and present the content material to the students in class) or active learning method in the classroom, but as a result of an alternative *assessment approach* (the ways in which student learning is measured and graded) outside of procedure-based in-class testing.

Chapter 3

METHODS

Site

Data were collected at Rowan College of Gloucester County (RCGC). RCGC is one of 19 public community colleges in the state of New Jersey. More than 7000 students are enrolled annually in courses at RCGC. About 56% of the students are full-time. The student body is predominantly white (70%) and African American (13%). Females comprise about 53% of the student body. About 60% of students are between the ages of 18-24, so there is a significant population of students outside of the traditional college age (coming directly from high school), including about 14% age 45 and older.

Sample

I collected data from five sections of statistics in the spring 2016 semester. Each section began with an enrollment of 30 students, for a total of 150 students. At the midpoint of the semester, 141 students took the midterm exam. During the last week of class, 133 students took the in-class final exam. At the beginning of the semester, 140 students completed the pre-test attitude survey. At the end of the semester, 113 students took the post-test attitude survey. A total of 108 students completed the course evaluation at the end of the semester.

Intervention Description

The five sections were similar in several ways. I taught each section in this study. The same midterm, in-class final exam and take-home final exam was used in all sections. The sections varied from each other in two important ways: assessment type (project-based and quiz-based) and learning environment (face-to-face and hybrid). I provide definitions of each assessment type and each learning environment below:

Assessment Type

Project-based. Students in these sections were assessed throughout the semester with four mini-projects and one larger semester project. For each project, students were able to research topics that interested them, gather their own data, analyze their data using techniques learned in class, create appropriate visual displays using computer packages, and write up their results in a report form.

Quiz-based. Students in these sections were assessed throughout the semester with seven short in-class quizzes which emphasized primarily computation and procedural fluency, as well as some emphasis on vocabulary recognition.

Learning Environment

Face-to-face. Students in these sections met twice a week for 75 minutes each class session. All notes were provided in class through the use of handouts. E-learning was used only as a means to post notes for students who missed class.

Hybrid. Students in these sections met once a week in the classroom for a 75-minute session. Prior to coming to class each week, students were required to study notes posted online for that chapter. In class, students completed handouts containing additional notes for each chapter or practiced examples which required the use of the online posted notes.

Students *did* know ahead of time what type of learning environment they were signing up for: hybrid or face-to-face. However, students did *not* know ahead of time what type of assessment structure would be used in their section. The first day of class students were given a syllabus that described their assessment structure. The table below summarizes the type of learning environment and assessment differences by section for this intervention. Note that the table gives each section a number. These numbers are simply the number assigned by the universities to these sections, based on the order of the times the classes are offered in the master schedule. The official labels for the sections taught were: Math 103-04, Math 103-06, Math 103-07, Math 103-09, and Math 103-10. I have abbreviated these on the chart below by identifying them as Sections 4, 6, 7, 9, and 10. Other instructors taught sections 5 and 8. Each section began with an enrollment of 30 students.

Assessment Method	Learning Environment	
	Face-to-Face (FTF)	Hybrid (H)
Project-Based (P)	Section 4	Section 10
Quiz-Based (Q)	Section 6 Section 7	Section 9

There were very little instructional differences between the project-based sections and the quiz-based sections. The time allotted for giving quizzes in the quiz-based sections was used to explain, and then provide feedback on, the projects in the project-based sections. Therefore, the same overall time was spent in each section, regardless of assessment method, on teaching the notes through class handouts, which was the primary method of teaching approach used by the instructor. In addition, there were very few instructional differences between the hybrid and face-to-face sections. Students in all sections, regardless of learning environment, received the same set of notes, just delivered differently. That is, the notes available to students online in the hybrid sections were the same notes that were presented in-class to the face-to-face sections. All five sections were so similar, that a student could attend any section and experience the same teaching style of going through class handouts and using examples to practice the material.

Data Collection Plan

The data for this study were collected at several points throughout the semester. The timeline on the following page illustrates the data collection plan. The paragraphs that follow the timeline more fully describe this data collection plan.

Time Frame (15 week course)	Project-Based Sections	Quiz-Based Sections
First week of class	• Baseline Math Skills Survey (N=140) • Survey of Attitudes Toward Statistics (SATS) Pre-Test (N=140)	
First half of semester	Mini Projects 1, 2, 3 Semester Project Proposal	Quizzes 1-3
Halfway Point	Common Midterm Exam (N=141)	
Second half of semester	Mini Project 4 Semester Project	Quizzes 4-7
Last Week of class	• Survey of Attitudes Toward Statistics (SATS) Post-test (N=113) • Common Final Exam (N=133) • Course Evaluation Form (N=108)	

During the first week of class, two types of data were collected for this proposal. The first was a baseline measure of mathematic and statistical skills entering the course (Appendix A). This measure was designed by myself and consisted of 16 multiple choice questions focusing on basic algebra skills. It was administered to students the first day of class, during the last 15 minutes of the class period. The purpose of this baseline measure was to hopefully establish that students across sections were similar entering the course. Also during the first week of the course, the pre-test for the Student Attitudes Toward Statistics (SATS-36) was completed (Appendix B). This survey was completed online by students. This 36-item questionnaire assesses student attitudes in 6 components: affect (6 items), cognitive competence (6 items), value (9 items), difficulty (7 items), interest (4 items), and effort (4 items). Students responded to each of the 36 items on the survey using a 7-point Likert scale. A student's score for each of the six attitude components was calculated by using the mean for all questions that correspond to that component. This inventory was repeated during the final weeks of the course to determine what changes occurred, for each student, during the semester.

Throughout the semester data were collected in the form of projects for the project-based sections and in-class quizzes for the quiz-based sections. To more fully see the differences in these assessments, I have included some samples of quizzes and projects in the Appendix. **Quiz #1** (Appendix C) is an example of the vocabulary-based matching assessment given. This is also similar to questions on the midterm and final exam that address key vocabulary terms and statistical methods used throughout the course. **Quiz #2** (Appendix D) is an example of the computation-based quiz questions that address procedural fluency. **Project #1** (Appendix E) illustrates the student-centered, data-driven approach to assessment that allows students to form their own conjectures, create a data collection plan, gather their data, and then assess their initial conjectures through a report form. **Semester Project** (Appendix F) is the semester-long project that students work on over the course of many weeks that draws from material throughout the entire course.

Other academic performance data collected during the semester were the midterm and final exam. The midterm exam and final exam were used for comparisons across sections. Despite learning environment (hybrid or face-to-face) or assessment structure (quiz-based or project-based), all sections received the same midterm exam and the same final exam. The midterm exam took place, in all sections, during one 75-minute class period. No notes were permitted for this midterm exam. About one-third of the midterm exam was comprised of matching and multiple choice questions that addressed vocabulary and statistical method identification. The remaining two-thirds were short answer problems based on computation and procedure. One sentence of interpretation may have been required for some computation questions. The final exam took place in two parts. One part was an in-class portion during one 75-minute class period for all sections. The second part was a take-home portion students had one week to complete and return. No notes were permitted for the in-class portion. For the take-home portion, notes could be used. Students signed a pledge for the take- home portion indicating that they did not ask for or receive help for any questions and did not consult any other person while completing the exam. All work was to be done by the individual alone and all work needed to be shown for full credit. A small portion, about 10%, of the final exam was comprised of matching and multiple choice questions that addressed key vocabulary terms from the semester. About 75% of the final exam was problems based on several 5-step procedures learned in class for inference. These 5-step procedures included both computation portions as well as interpretation and written conclusion portions. The remaining portion of the final exam was computation problems involving probability.

During the last week of class, two additional types of data were collected. First, I re-administered the SATS-36 post-test (Appendix G) to all students, to assess their attitudes at the end of the semester. Students completed this online. Second, each student completed a course evaluation form (Appendix H). This course evaluation contained three open-ended questions for students to answer regarding their understanding of the purpose of and goals of statistics. It also contained 6 statistical tasks (like 'write a research question') and asked students to rate themselves on how confident they felt about their ability to complete these tasks if they were asked to do so.

Data Analysis

I provide tables on the following pages that outline the types of analyses that were conducted for each set of data described in the previous section. A separate table is provided for each sub-question. The first two tables focus on the two sub-questions related to the first research question comparing the academic performance of students enrolled in project-based sections with students enrolled in quiz-based sections. The next three tables outline the data analysis plan for the three sub-questions relating to the second research question comparing the attitudes of students in project-based sections with those in quiz-based sections.

Research Question	**Data**	**Analysis**
1a. How does the performance of students in the project-based sections on the midterm and final exam compare to students in the quiz-based sections?	• Midterm Exam • Final Exam	• ANOVA to look for differences in mean scores across five sections; multiple comparisons when appropriate • T-tests to look for differences when sections are grouped by assessment method • T-tests to look for differences when grouped by learning environment

Research Question	**Data**	**Analysis**
1b. What kind of correlation exists between grades on quizzes during the semester and grades on the midterm and final exam for students in the quiz-based sections? What kind of correlation exists between grades on projects during the semester and grades on the midterm and final exam for students in the project-based sections?	• Midterm Exam • Final Exam • Quiz Average (for students in quiz-based sections) • Mini-Project Average (for students in project-based sections) • Semester Project (for students in project-based sections)	• Pearson's Correlation Coefficient Computation for Quiz Average, Midterm, and Final Exam (for students in quiz-based sections) • Pearson's Correlation Coefficient Computation for Mini Project Average, Semester Project, Midterm, and Final exam (for students in project-based sections) • Hypothesis Test for Significant Correlation

Research Question	Data	Analysis
2a. How do students' attitudes in project-based sections, as reported on a quantitative rating scale, compare to students' attitudes in quiz-based sections at the end of the semester?	Survey of Attitudes Toward Statistics (SATS-36) pre- and post-tests	• ANOVA to look for differences in mean scores across five sections; multiple comparisons when appropriate • T-tests to look for differences when sections are grouped by assessment method • T-tests to look for differences when grouped by learning environment

Research Question	Data	Analysis
2b. How have students' attitudes, as reported on a quantitative rating scale, changed over the course of the semester?	Differences in pre- and post-test scores for Survey of Attitudes Toward Statistics (SATS-36)	• ANOVA to look for differences in mean differences (post-pre) across five sections; multiple comparisons when appropriate • T-tests to look for differences when mean differences are grouped by assessment method • T-tests to look for differences when mean differences are grouped by learning environment

Research Question	Data	Analysis
2c. What kind of differences in students' attitudes exist in the feedback provided on a course evaluation form for students in the quiz-based sections vs. students in the project-based sections?*	Course Evaluation Form	• ANOVA to look for differences in mean scores across five sections; multiple comparisons when appropriate • T-tests to look for differences when sections are grouped by assessment method • T-tests to look for differences when grouped by learning environment

*_Note_: Data from the open-ended questions on the course evaluation form were thrown out of the study due to error in student interpretation of the questions.

Chapter 4

RESULTS PART I

In this chapter, I revisit the first research question in which academic performance was used as a means for comparison between students enrolled in project-based sections and students enrolled in quiz-based sections. This first research question contained two sub-questions. I will address each sub-question separately.

First Academic Performance Research Sub-Question

1. How does students' performance compare between those who enrolled in project-based sections of statistics and those enrolled in quiz-based sections?

   1a. How does the performance of students in the project-based sections on the midterm and final exam compare to students in the quiz-based sections?

Before I address the first research question, I will discuss the results of the Baseline Math Skills Survey. This analysis sets the context for the study because it allows me to assess the mathematical skills for students in different sections at the start of the course. If a section of students did have stronger (or weaker) incoming skills, this incoming difference could explain findings if that section of students performed better (or worse) by the end of the semester. In addition to section, I also tested for differences between the groupings focused on in this study: assessment type and learning environment.

No significant difference in the mean score on the Baseline Math Skills Survey was found for students when the data was grouped in any of the three ways focused on in this project. These three groupings are: by section (4, 6, 7, 9, and 10), by assessment type (Project and Quiz), and by learning environment (Face-to-Face and Hybrid). This is helpful to know because it indicates that at the beginning of the semester, the various groupings had, on average, similar basic math skill sets. The descriptive statistics, along with p-value results when testing for differences, for each grouping are listed in Table 1, Table 2, and Table 3 below.

Table 1 Baseline Math Skills Survey Results by Assessment Type

Assessment Type	Count	Mean	St. Dev.
Project	56	70.45	15.72
Quiz	84	69.73	19.72
p = 0.819			

Table 2 Baseline Math Skills Survey Results by Learning Environment

LearningEnvironment	Count	Mean	St. Dev.
Face-to-Face	87	70.24	18.33
Hybrid	53	69.64	18.08
p = 0.851			

Table 3 Baseline Math Skills Survey Results by Section

Section	Count	Mean	St. Dev.
4 (P, FTF)	29	71.41	17.71
6 (Q, FTF)	29	68.21	20.18
7 (Q, FTF)	29	71.10	17.45
9 (P, H)	27	69.41	13.53
10 (Q, H)	26	69.88	22.12
$p = 0.965$			

Using ANOVA, we now have evidence that the five sections were all similar at the start of the course in terms of mathematical skills. When grouped either by assessment type or by learning environment, we also have evidence, using t-tests, that mathematical skills were similar at the start of this study. The results of this analysis are encouraging because now, if differences in midterm or final exam are found in any of these groupings at the end of the semester, then we have ruled out that those differences could be attributed to differences in mathematical skills present between the sections or groups at the start of the semester.

Having controlled for incoming capabilities with the baseline math skills test, I now focus on the two assessment measure used in this study to measure academic performance: midterm exam and final exam. The final exam was given in two pieces: the in-class portion and the take-home portion. Altogether, therefore, there are three assessment measures I am comparing across the groups: midterm, in-class final exam and take-home final exam.

First, I looked at the midterm exam. Using t-tests, there was no significant difference in the average midterm exam scores when students were divided by assessment type (p = 0.228). In the analysis when grouped by section, however, an ANOVA showed that the differences in means across the five sections approached significance (p = .090). In the follow-up pairwise comparisons using the Tukey-Kramer method, I found that the average midterm exam score for section 7 (Q, FTF) was found to be significantly higher than the average midterm exam scores for both section 4 (P, FTF) and section 10 (Q, H). No other significant differences were found. Table 4 and Table 5 below provide descriptive statistics for midterms exam scores by assessment type and by section, respectively. The p-values, when testing for differences, are found in the final row of each table.

Table 4 Midterm Exam Scores by Assessment Type

Assessment Type	Count	Mean	St. Dev.
Project	54	70.02	15.92
Quiz	87	73.56	17.47
p = 0.228			

Table 5 Midterm Exam Scores by Section

Section	Count	Mean	St. Dev.
4 (P, FTF)	28	67.96*	17.25
6 (Q, FTF)	29	73.24	17.76
7 (Q, FTF)	29	79.00*	15.98
9 (P, H)	26	72.23	14.35
10 (Q, H)	29	68.45*	17.55
p = 0.090			

Note: $*\mu_7 > \mu_4, \mu_{10}$

The second measure of assessment used was the in-class final exam. When comparing by assessment type, a t-test showed a significant difference found between project-based and quiz-based sections (p=.003). Students in the quiz-based sections scored higher, on average, than students in the project-based sections. When comparing the data by section, an ANOVA showed at least one difference among the section means. Using follow-up pairwise comparisons with the Tukey-Kramer method, several significant pairwise differences were found. Students in section 6 (Q, FTF) scored, on average, higher than students in sections 4 (P, FTF), 9 (P, H), and 10 (Q, H). In addition, students in section 7 (Q, FTF) scored higher, on average, than students in section 4 (P, FTF). Descriptive statistics for these two groupings can be found in Table 6 and Table 7 below. Standard deviations are provided in parentheses. P-values are listed in the final row of each table.

Table 6 In-Class Final Exam Scores by Assessment Type

Assessment Type	Count	Mean	St. Dev.
Project	50	61.80*	24.25
Quiz	83	74.14*	19.07
p = 0.003			

Note: $^*\mu_Q > \mu_P$

Table 7 In-Class Final Exam Scores by Section

Section	Count	Mean	St. Dev.
4 (P, FTF)	28	59.36*	25.80
6 (Q, FTF)	27	82.59*	11.30
7 (Q, FTF)	28	75.32*	17.35
9 (P, H)	22	64.91*	22.33
10 (Q, H)	28	64.82*	22.74
p = 0.0003			

Note: $*\mu_6 > \mu_4, \mu_9, \mu_{10}; *\mu_7 > \mu_4$

The final measure of assessment used in these comparisons involved the take-home final exam. When grouping by assessment type, a t-test showed no significant difference (p = 0.684) between the average score in the project-based sections as compared to the quiz-based sections. For the comparison among sections, an ANOVA showed that significant differences existed among the means. A follow-up analysis of pairwise comparisons using the Tukey-Kramer method showed that students in sections 4 (P, FTF), 6 (Q, FTF), and 7 (Q, FTF) scored significantly higher, on average, than students in section 10 (Q, H). In addition, students in section 7 (Q, FTF) scored significantly higher, on average, than students in section 9 (P, H). Descriptive statistics for the analyses for the take-home final exam can be found in Table 8 and Table 9 below. Standard deviations are provided in parentheses. P-values are listed in the final row of each table.

Table 8 Take-Home Final Exam Scores by Assessment Type

Assessment Type	Count	Mean	St. Dev.
Project	48	83.44	8.45
Quiz	83	82.70	12.21
$p = 0.684$			

Table 9 Take-Home Final Exam Scores by Section

Section	Count	Mean	St. Dev.
4 (P, FTF)	26	85.65*	8.35
6 (Q, FTF)	27	83.31*	11.64
7 (Q, FTF)	28	89.50*	8.54
9 (P, H)	22	80.82*	7.97
10 (Q, H)	28	74.82*	11.66
$p < 0.0001$			

Note: $*\mu_4, \mu_6, \mu_7 > \mu_{10}$; $*\mu_7 > \mu_9$

The first research question used three assessment measures common to all sections to compare the sections using the quiz-based assessments to the sections using the project-based assessments. At the midway point of the semester, the sections, when grouped by assessment-type, performed similarly on the common midterm exam. This is encouraging because it indicates that the students in the project-based sections, while not exposed to the quizzes that were similar in format to the common midterm exam, were still able to perform at a similar level on that midterm exam. That is, using projects as an assessment method for the first part of the course did not hinder the students in those sections from being able to perform at a similar level to their counterparts in quiz-based sections.

Unfortunately, by the end of the semester, however, a difference did emerge between students in project-based sections and students in quiz-based sections. Because the performance on the in-class final exam was significantly higher, on average, for students in the quiz-based sections it seems that students exposed to quizzes similar in format to the common in-class final exam may have been more prepared than students in the project-based sections. Students in quiz-based sections took three different quizzes that related to the material emphasized on the in-class final exam. In each case, the problems were presented in identical ways to what the students were asked to do on the in-class final exam. Students in quiz-based sections, therefore, received feedback three different times, through the final three quizzes in the course, on the material that was emphasized on the in-class final exam. Students in project-based sections had just one project, the Semester Project, which provided an opportunity for them to receive feedback on the material that was emphasized on the in-class final exam. The increased opportunity for feedback for students in the quiz-based sections may have contributed to their higher average score on the in-class final exam when compared with students in the project-based sections.

The difference in the in-class final exam performance did not appear in the take-home final exam portion, however. When grouped by assessment type, students in either group performed similarly, on average, on the take-home final exam portion. This is interesting because the material covered on take-home exam, probability, was not a topic that students in quiz-based sections had been quizzed on prior to that point. Students in the project-based sections, however, did have a project devoted to the topic of probability prior to the take-home final exam. This illustrates that students in the quiz-based sections were still able to perform at a similar level on the take-home final exam portion despite having not been assessed on the material prior to that point. I am cautious to attach too much weight to this, however, because the topic of probability tends to be one of the easier topics in this course. It is less difficult computationally, as well as more a more familiar topic for students prior to entering the course. Therefore, the fact that students performed similarly on this take-home exam covering probability despite the project-based sections having been assessed on the material prior to this point and the quiz-based sections *not* having been assessed on the material prior to this point, may not be surprising.

The examination of results by section opens up a further discussion topic. In *every* instance where a difference was found, that difference favored a section that was in a face-to-face learning environment. To explore this learning environment factor further, I performed t-tests for each of the three common assessments when the sections were grouped by learning environment. Using t-tests, there was no significant difference in the average midterm exam scores when students were divided by learning environment ($p = 0.271$). The difference in average in-class final exam score between these two groups approached significance ($p=.058$). For the take-home final exam, there was strong evidence indicating a significant difference between the two groups ($p < .001$). Students in the face-to-face sections scored significantly higher, on average, than students in the hybrid sections on both the in-class and take-home final exam. The descriptive statistics for these results can be found in Tables 10, 11, and 12. P-values are provided in the final row of each table.

Table 10 Midterm Exam Scores by Learning Environment

Learning Environment	Count	Mean	St. Dev.
Face-to-Face	86	73.47	17.41
Hybrid	55	70.24	16.08
$p = 0.271$			

Table 11 In-Class Final Exam Scores by Learning Environment

Learning Environment	Count	Mean	St. Dev.
Face-to-Face	83	72.30*	21.31
Hybrid	50	64.86*	22.33
$p = 0.058$			

Note: $^*\mu_{FTF} > \mu_H$

Table 12 Take-Home Final Exam Scores by Learning Environment

Learning Environment	Count	Mean	St. Dev.
Face-to-Face	81	86.37*	9.81
Hybrid	50	77.46*	10.54
$p < 0.001$			

Note: $^*\mu_{FTF} > \mu_H$

Looking over the results to this point, an interesting pattern emerged. When differences existed, they typically favored either quiz-based sections, face-to-face sections, or a combination of those two things. This led me to look at the data in four subgroups, as described in the chart below.

Assessment Method	Learning Environment	
	Face-to-Face	Hybrid
Project-Based	Section 4	Section 10
Quiz-Based	Section 6 Section 7	Section 9

To study this further, a multiple linear regression analysis was completed using three predictors: Baseline Math Skills Survey score, Learning Environment, and Assessment Type. All three of these predictors were statistically significant in the model. Students with higher baseline math skills scores, in face-to-face classes, and assessed with quizzes do better. The results for this analysis using In-Class Final Exam score as the dependent variable produced the following equation:

In-Class Final Exam Score = 43.2 +.21(Baseline Math Skills Score) + 7.7 (Learning Environment) + 10.5(Assessment Type).

This model indicates that the learning environment/assessment type combination has a large effect on student's performance on the in-class final exam. As further explanation, if a student is in hybrid and project-based assessment section (both coded “0”), his/her predicted in-class final exam score was 43.2 plus his/her baseline math skills score. For any given student, regardless of baseline math skills test score, if s/he was in a face-to-face learning environment and assessed with quizzes (both coded as “1”), his/her predicted in-class final exam is about 18 points higher (7.7 + 10.5 = 18.2 points). In other words, controlling for baseline math skills score, a student in a face-to-face learning environment and who is assessed with quizzes will, on average, outscore a student in a hybrid learning environment who is assessed with projects by about 18 points. If using quiz assessment (coded as “1”), but in a hybrid learning environment (coded as “0”), the difference is 10.5 points higher, relative to a hybrid, project-based assessment. If using project assessment (coded as “0”), but in a face-to-face learning environment (coded as “1”), then the difference is 7.7 points higher, on average, relative to a hybrid, project-based assessment.

Because of the way that learning environment and assessment method interact in this study, a two-way ANOVA would have been a natural analysis to complete at this point. However, performing a two-way ANOVA requires that certain assumptions be met. In particular, the cells would need to have similar variances. Because the variances of the four subgroups were *not* sufficiently equal, it was determined that a two-way ANOVA was not appropriate for this data. For this reason, one-way, rather than two-way, ANOVAs were used throughout this study. However, it is still useful to look at the summary statistics for each learning environment/assessment method combination. An apparent benefit for being in a face-to-face, quiz-based section can be seen by looking at Table 13 below. In this table, the means and standard deviations are presented for each learning environment/assessment method combination. As is evident when examining these combinations, in addition to having the highest mean, the face-to-face, quiz combination also has the smallest standard deviation due to having fewer lower test scores.

Table 13 Mean and Standard Deviation for In-Class Final Exam for Two Factors

	Assessment Type		
Learning Environment	Project	Quiz	Total
Face-to-Face	59.4 (25.8)	78.9 (15.0)	72.3 (21.3)
Hybrid	64.9 (22.3)	64.8 (22.7)	64.9 (22.3)
Total	61.8 (24.3)	74.1 (19.1)	69.5 (21.9)

Clearly, both assessment type and learning environment, therefore, were factors in determining how well students did on the in-class final exam. While the only statistically significant differences that were found between sections favored quiz-based, face-to-face sections (Table 7), the standard deviations in Table 13 above provide additional information when comparing assessments. The standard deviation for the quiz-based, face-to-face combined sections on the in-class final exam was over 10 points lower than the standard deviation for the project-based, face-to-face section. When comparing the hybrid sections, though, the standard deviation in the quiz-based section was almost the same as (less than 0.5 points different) the standard deviation for the project-based section. The quiz-based, face-to-face grouping, although it contained the highest number of students, had, by far, the smallest standard deviation.

This indicates, again, the less frequency of low scores on the in-class final exam for this assessment method/learning environment combination. The combination of more in-class time (due to the face-to-face environment) and more opportunity for feedback (due to the quiz-based assessment method) led to an overall increase in the average and decrease in the standard deviation for this grouping.

In summary, when examining academic performance using the midterm and final exam, there was no significant difference found between students in project-based section and students in quiz-based sections for either the midterm exam or the take-home portion of the final exam. For the in-class final exam, however, students in quiz-based sections scored significantly higher, on average, than students in project-based sections. These results are somewhat confounded by learning environment differences. When separating by section, every significant difference favored a section that was a face-to-face section. A multiple linear regression analysis showed that the optimum performance on the in-class final exam (where the most differences existed) was found in a quiz-based, face-to-face format.

Second Academic Performance Research Sub-Question

I now move on to the second sub-question for this research question involving academic performance:

1. How does students' performance compare between those who enrolled in project-based sections of statistics and those enrolled in quiz-based sections?

> 1b. What kind of correlation exists between grades on quizzes during the semester and grades on the midterm and final exam for students in the quiz-based sections? What kind of correlation exists between grades on projects during the semester and grades on the midterm and final exam for students in the project-based sections?

This sub-question focuses on all the assessments given throughout the semester. For quiz-based sections, I calculated the quiz average (based on 6 quizzes) for each student. For student in the project-based sections, I used both their mini-project average (based on 4 mini-projects) and their semester project score. For students in all sections I also used their midterm, in-class final exam, and take-home final exam score.

For the quiz-based sections, the correlation was strong and positive between the quiz average and both the midterm (.708) and in-class final exam (.707). The correlation between the quiz average and the take-home final exam was also positive, but not quite as strong (.599). For the project-based sections, the mini-project average was most strongly positively correlated with the midterm (.689). The correlation between the mini-project average and the in-class final exam (.525) and the take-home final exam (.310) was also positive, but not as strong. When examining the semester project score, the strongest positive correlation was found with the in-class final exam (.559). The correlation was positive, but not as strong, between the semester project score and the midterm (.238) and the take-home final exam (.299). Table 14 shows the Pearson's Pairwise Correlation Coefficients for students in the project-based sections. Table 15 shows the Pearson's Pairwise Correlation Coefficients for students in the quiz-based sections. Every correlation on each table was statistically significant.

Table 14 Pearson's Pairwise Correlations in Project-based Sections

	Mini Project Average	Semester Project	Midterm	In-Class Final Exam	Take-Home Final Exam
Mini Project Average	1.000				
Semester Project	.420*	1.000			
Midterm	.689*	.238*	1.000		
In-Class Final Exam	.525*	.559*	.491*	1.000	
Take-Home Final Exam	.310*	.299*	.259*	.284*	1.000
Final Grade	.697*	.681*	.642*	.826*	.347*

**Note:* Statistically significant correlation ($p < .05$)

Table 15 Pearson's Pairwise Correlations in Quiz-based Sections

	Quiz Average	Midterm	In-Class Final Exam	Take-Home Final Exam
Quiz Average	1.000			
Midterm	.708*	1.000		
In-Class Final Exam	.707*	.453*	1.000	
Take-Home Final Exam	.599*	.592*	.501*	1.000
Final Grade	.948*	.818*	.780*	.675*

**Note:* Statistically significant correlation ($p < .05$)

The quiz average for the quiz-based sections showed a stronger, on average, correlation with the midterm and final exam components than the project scores throughout the semester for the project-based sections. This difference could have been due to the format of the quizzes. Questions used on the quizzes were very similar in format to the questions used on the midterm and in-class final exam. Questions on both quizzes, midterm, and final exam components were also very similar in format to the examples used on the class handouts. Therefore, students in quiz-based sections were more familiar with the format of the questions on the midterm and in-class final exam than students in the project-based sections. Because the format was similar, this could have led to stronger correlation in performance between these items.

In addition, this stronger correlation could have been due to the differences in grading structure between quizzes and projects. With quizzes, students are graded on their ability to complete a statistical procedure, recognize a vocabulary word in a multiple choice situation, or calculate a statistical value. With projects, in addition to containing calculation and procedural elements, students are also graded on their writing and presentation of ideas in a paper format. Because of the more complex grading structure, project grades tend to vary more. Therefore, students could have a lower project grade, but that doesn't mean their statistical content was poor. Instead, they could have performed well on the statistical content but not performed as well on the written component. Similarly, students could have performed poorly in the statistical content, but their overall grade on a project might not reflect that. If their writing clearly explained their procedures (despite obtaining incorrect answers statistically), then their overall grade on a project might be higher than what their performance on the computational and procedural parts of the project alone reflected. Overall, it makes sense that the similarities in both format and grading components found in quizzes as compared to the midterm and final exam could lead to the results found for this research question: stronger correlations with quiz average and midterm and final exam, than the correlations found with project scores and midterm and final exam.

In summary, when examining pair-wise correlations within project-based and quiz-based sections, correlations with the midterm and final exam were stronger in the quiz-based sections. Tying together the two parts of this first research question, any differences that existed in the academic performance between project-based and quiz-based sections tended to favor the quiz-based sections. These results should be viewed with caution, however, because learning environment, as demonstrated in the multiple linear regression analysis, had a significant impact on academic performance as well.

Chapter 5

RESULTS PART II

In this chapter, I present results for the second research question regarding students' attitudes about statistics. This research question has three sub-questions and I will address each sub-question separately.

First Attitudes Research Sub-Question

2. How do students' attitudes toward and ideas about statistics compare between those in project-based sections of statistics and those in quiz-based sections?

   2a. How do student attitudes in project-based sections, as reported on a quantitative rating scale, compare to students attitudes in quiz-based sections at the end of the semester?

Results from the pre-test scores on the Survey of Attitudes Toward Statistics (SATS) showed no significant difference when students were grouped either by assessment method or by learning environment. In addition, among sections, there was no significant differences found for any of the component averages on the pre-test scores. This is helpful because it indicates that students across sections, assessment methods, and learning environments had similar attitudes about statistics at the start of the course.

In the post-test score analysis, no significant differences were found for any of the six component averages when the students were grouped by assessment method. However, when grouped by section and by learning environment, analyses did show significant differences. When grouped by section, four of the six component averages showed a significant difference among component averages. In each of the four cases, section 10 (Q, H) scored significantly lower than another section. In three of those cases, section 7 (Q, FTF) scored significantly higher than section 10 (Q, H), and in the fourth case, section 6 (Q, FTF) scored significantly higher than section 10 (Q, H). When grouped by learning environment, again, four of the six component averages showed a significant difference. In all four cases, students in the face-to-face sections scored higher, on average, than students in the hybrid sections.

The SATS (Survey of Attitudes Toward Statistics) was used to address this research question. The six components identified with SATS are Affect (students' feelings concerning statistics), Cognitive Competence (students' attitudes about their intellectual knowledge and skills when applied to statistics), Value (students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life), Difficulty (students' attitudes about the difficulty of statistics as a subject), Interest (students level of individual interest in statistics), and Effort (amount of work the student expects to learn statistics). Information on which of the 36 items were grouped within each component are presented in Appendix B, along with the SATS survey itself.

Grouping by section

First, I consider the students grouped by section. Table 16 and Table 17 below show the pre- and post-test component averages for this grouping. The standard deviation for each average is provided in parentheses. Using ANOVA, no significant differences were found among sections for any of the six component averages on the pre-test. For the post-test results, an ANOVA indicated at least one significant difference was found in four of the six component averages. The p-value results for the ANOVA are listed in the final column of Table 16 and Table 17. Table 18 identifies the pair-wise differences found to be significantly different.

Table 16 Pre-Test Component Averages by Section

	04 (P, FTF) N=27	06 (Q,FTF) N=30	07 (Q,FTF) N=28	09 (P,H) N=27	10 (Q,H) N=28	Overall N=140	P-Value
Affect	4.75 (1.20)	4.42 (1.20)	4.56 (1.32)	4.74 (1.19)	4.34 (1.44)	4.56 (1.27)	0.681
Cognitive Comp.	5.12 (0.97)	4.86 (1.05)	5.26 (1.14)	5.19 (1.27)	5.02 (1.30)	5.09 (1.14)	0.710
Value	4.90 (1.31)	4.77 (0.95)	5.00 (1.16)	4.92 (1.01)	4.81 (1.34)	4.88 (1.15)	0.947
Difficulty	3.83 (0.70)	3.75 (0.71)	3.66 (0.78)	3.84 (0.83)	3.66 (0.95)	3.75 (0.79)	0.852
Interest	4.53 (1.64)	4.66 (1.26)	4.84 (1.24)	4.94 (1.08)	4.45 (1.65)	4.68 (1.38)	0.672
Effort	6.41 (0.75)	6.40 (0.67)	6.34 (0.68)	6.57 (0.57)	6.50 (0.61)	6.45 (0.65)	0.789

Table 17 Post-Test Component Averages by Section

	04 (P,FTF) N=22	06 (Q,FTF) N=26	07 (Q,FTF) N=22	09 (P,H) N=21	10 (Q,H) N=22	Overall N=113	P-Value
Affect	4.74 (1.55)	5.29 (1.07)	5.27 (1.63)	4.91 (1.65)	4.17 (1.73)	4.89 (1.56)	0.090**
Cognitive Comp.	5.14 (1.36)	5.55 (0.95)	5.76 (1.42)	5.25 (1.27)	4.56 (1.68)	5.26 (1.38)	0.042*
Value	4.86 (1.07)	5.10 (1.05)	5.34 (1.32)	4.81 (1.25)	4.23 (1.55)	4.88 (1.29)	0.050*
Difficulty	3.93 (1.02)	3.95 (0.86)	4.08 (0.80)	3.77 (1.01)	3.62 (1.13)	3.87 (0.96)	0.578
Interest	4.07 (1.82)	4.78 (0.94)	5.03 (1.59)	4.55 (1.45)	3.73 (1.78)	4.44 (1.58)	0.035*
Effort	6.08 (0.77)	6.07 (0.92)	6.38 (0.57)	6.30 (0.61)	6.08 (0.94)	6.18 (0.78)	0.538

Note: *ANOVA found significant difference among means ($p \leq .05$)
Note: **ANOVA found significant differences among means ($.05 < p < .10$)

Table 18 Pair-wise Differences for Items with at least one Significant Difference Among Section Means

Component	Significant Pairwise Difference
Affect	$\mu_6(Q, FTF) > \mu_{10}\,(Q, H)$
Cognitive Competence	$\mu_7(Q, FTF) > \mu_{10}\,(Q, H)$
Value	$\mu_7(Q, FTF) > \mu_{10}\,(Q, H)$
Interest	$\mu_7(Q, FTF) > \mu_{10}\,(Q, H)$

This analysis by section highlights some interesting conclusions. Table 18 above identifies the four components in which a significant difference was found. For each of these four components (affect, cognitive competence, value, and interest), a higher average score indicates a more positive attitude. In every instance where a difference was found, that difference favored a section that was a quiz-based, face-to-face section. Also, in every instance a difference was found, the section that scored significantly lower was section 10, which was a quiz-based, hybrid section. This indicates that learning environment seemed to play a factor in student attitudes for students in quiz-based sections. It is also interesting that none of the differences involved a project-based section. It would have been nice to see a project-based section score significantly higher than a quiz-based section for one of these four components, as this would have supported my original conjecture, that students in project-based sections showed a difference in attitudes at the end of the course as compared to students in quiz-based sections. While the absence of such difference is not encouraging, it is encouraging that at least the students in project-based sections never scored significantly lower, on average, than students in quiz-based sections on any of these four component averages. This indicates that their averages were not worse at the end of the semester.

Grouping by assessment method

Second, I consider students grouped by assessment method. As stated earlier, no significant differences were found in either pre- or post-test results for any of the six component averages. Table 19 and Table 20 provide the pre- and post-test component averages by assessment method. The standard deviation for each average is provided in parentheses. The p-value for the t-test is found in the final column on each table.

Table 19 Pre-Test Component Averages by Assessment Method

	Project-Based (N=54)	Quiz-Based (N=86)	P-Value
Affect	4.74 (1.18)	4.44 (1.31)	0.166
Cognitive Competence	5.15 (1.12)	5.04 (1.16)	0.575
Value	4.91 (1.16)	4.86 (1.15)	0.803
Difficulty	3.84 (0.76)	3.69 (0.81)	0.291
Interest	4.74 (1.39)	4.65 (1.39)	0.719
Effort	6.49 (0.66)	6.42 (0.65)	0.572

Table 20 Post-Test Component Averages by Assessment Method

	Project-Based (N=43)	Quiz-Based (N=70)	P-Value
Affect	4.83 (1.58)	4.93 (1.55)	0.729
Cognitive Competence	5.20 (1.31)	5.30 (1.44)	0.698
Value	4.83 (1.16)	4.90 (1.38)	0.785
Difficulty	3.85	3.89	0.850

	(1.01)	(0.94)	
Interest	4.30 (1.65)	4.53 (1.54)	0.462
Effort	6.19 (0.69)	6.17 (0.83)	0.895

While I would have liked to see positive differences for the components of affect, cognitive competence, value, or interest in project-based sections, the *lack* of a difference for the effort and difficulty component is worth interpreting here. Students in project-based sections put in more time outside of class than students in quiz-based sections. The amount of work to design projects, gather data, and write up results for the projects is significantly more time-consuming than studying for a series of in-class quizzes. It is surprising and encouraging, therefore, that students in project-based sections ranked their attitudes in both the effort and difficulty components as similar to students in quiz-based sections at the end of the semester. This is encouraging because while students in project-based sections certainly would have put in more time and effort outside of class, this difference was not present in their formal attitude rankings at the end of the semester. I interpret this to mean that the 'work' students in project-based sections were doing did not feel like more work than students in quiz-based sections were doing. The amount of work invested in the projects did not stand out to students in project-based sections as being harder (more difficult) or requiring more time (more effort) than students whose outside of class work involved only studying for quizzes and tests. I believe this could at least partly be due to the fact that students choosing their own topic and researching things that interest them is an enjoyable experience and doesn't feel as much like 'work' as studying from a textbook or set of notes.

Grouping by learning environment

As noted in the discussion when students were grouped by section, the results indicated that learning environment did play a factor in differences found between sections on the post-scores for student attitudes. The t-tests confirmed these findings when students were grouped by learning environment. No significant differences were found among component averages on the pre-test. For the post-test, however, four component averages were found to be significantly different. For the components of value and cognitive competence, the component average for students in the face-to-face sections was significantly higher ($p < .05$) than for students in the hybrid sections. For the components of affect and interest, the difference in component averages between the learning environments approached significance ($.05 < p < .10$). In each of these cases, again, the component average for students in the face-to-face sections was significantly higher than for students in the hybrid sections. As discussed earlier, a higher average for the four components here that showed significance (affect, cognitive competence, value, and interest), indicate a positive shift in attitude. Table 21 and Table 22 provide the pre- and post-test scores for students grouped by learning environment. P-values from the t-tests are provided in the final column of each table. The standard deviation for each component average is provided in parentheses.

Table 21 Pre-Test Component Averages by Learning Environment

	Face-to-Face (N=85)	Hybrid (N=55)	P-Value
Affect	4.57 (1.23)	4.54 (1.33)	0.897
Cognitive Competence	5.07 (1.06)	5.10 (1.27)	0.885
Value	4.89 (1.13)	4.86 (1.18)	0.902
Difficulty	3.75 (0.73)	3.75 (0.89)	0.980
Interest	4.68 (1.38)	4.69 (1.41)	0.980
Effort	6.39 (0.69)	6.53 (0.59)	0.216

Table 22 Post-Test Component Averages by Learning Environment

	Face-to-Face (N=70)	Hybrid (N=43)	P-Value
Affect	5.11 (1.42)	4.53 (1.71)	0.054**
Cognitive Competence	5.49 (1.25)	4.90 (1.52)	0.028*
Value	5.10 (1.15)	4.51 (1.42)	0.018*
Difficulty	3.98 (0.88)	3.69 (1.07)	0.125
Interest	4.64 (1.50)	4.13 (1.66)	0.097**
Effort	6.18 (0.78)	6.19 (0.79)	0.895

Note: *T-test shows significant difference between means ($p < .05$)
Note: **T-test shows significant difference between means ($.05 < p < .10$)

It is again interesting to note where significant differences did *not* emerge in this analysis. Although students in hybrid sections *should* be putting in more work (higher effort) outside of the course due to the learning environment design, and many students comment to me anecdotally that participating in a hybrid section *is* harder (more difficult), these differences did not emerge when students ranked themselves on the effort and difficulty component at the end of the course. It is problematic to think, actually, that students in a hybrid course (meeting just once a week) would *not* rank their effort higher at the end of the semester than students in a face-to-face course (meeting twice a week). Perhaps this is one reason why students in the hybrid sections seem to be at a deficit on academic performance; they may not be putting in the effort required to succeed.

Therefore, to address the first part of this research question on students' attitudes, there were no significant differences found in students' attitudes at the end of the semester, when comparing students enrolled in project-based sections with students enrolled in quiz-based sections. Significant differences did exist when students were grouped by learning environment, however. In summary, while assessment method did not have an impact on student attitudes at the end of the semester, as measured by the SATS (36) post-test, learning environment did have an effect, with students in face-to-face sections scoring significantly higher, indicating a better attitude, on four of the six component averages than students in hybrid sections.

Second Attitudes Research Sub-Question

Because all tests with the pre-test scores showed no significant difference among means, any significant differences found in post-test scores can be attributed to the students' experiences during the semester. While testing for differences using the post-test scores was valuable and highlighted several differences in various groupings, all instances where learning environment, not assessment method, was a factor, I also felt it was important to compute the change between pre-test and post-test score within all combinations of groupings. This leads directly into the second sub-question involving students' attitudes:

> 2. How do students' attitudes toward statistics compare between those enrolled in project-based sections of statistics and those enrolled in quiz-based sections?
>
> 2b. How have student attitudes, as reported on a quantitative rating scale, changed over the course of the semester?

Even when no significant difference was found for a particular component in the pre-test scores, and then again in the post-test scores, this tells us nothing about what the actual value is for the average (on the scale of 1 to 7). If pre-test average scores for two sections are both around 3.0 and post-test average scores for the same sections are both around 5.0, then the analyses done to this point would simply report that no significant differences were found among component averages for either pre- or post-test scores. However, a significant piece of information would have been overlooked if I stopped there. Going from a 3.0 average to a 5.0 average would represent a significant change in attitude score from pre- to post-test score. Therefore, I now present information on the change from pre- to post-test scores for all groupings in this study.

Grouping by section

First I look at changes for each of the six components when the students were combined across all sections. In this analysis, two components showed a significant change in average reported score. Breaking the students down by section, several more significant changes emerged. Table 23 on the following page shows pre- test average, post-test average, and change (bold print) between averages when students were grouped by section. In addition, the overall change is recorded when students were combined across sections. The p-values for each analysis when comparing means from the pre- to the post-test average are also provided.

Table 23 Change in Component Average (Post-Test – Pre-Test) by Section

		04 (P, FTF)	06 (Q,FTF)	07 (Q,FTF)	09 (P,H)	10 (Q,H)	Overall N=140
Affect	Pre	4.75	4.42	4.56	4.74	4.34	4.56
	Post	4.74	5.29	5.27	4.91	4.17	4.89
	Change	**-.01**	**0.87***	**0.71**	**0.17**	**-0.17**	**0.33****
	P-Value	0.991	0.006	0.101	0.689	0.701	0.067
Cog. Comp.	Pre	5.12	4.86	5.26	5.19	5.02	5.09
	Post	5.14	5.55	5.76	5.25	4.56	5.26
	Change	**0.02**	**0.69***	**0.50**	**0.06**	**-0.46**	**0.17**
	P-Value	0.953	0.013	0.189	0.856	0.293	0.278
Value	Pre	4.90	4.77	5.00	4.92	4.81	4.88
	Post	4.86	5.10	5.34	4.81	4.23	4.88
	Change	**-0.04**	**0.33**	**0.34**	**-0.11**	**-0.58**	**0**
	P-Value	0.917	0.227	0.345	0.739	0.172	0.997
Diff.	Pre	3.83	3.75	3.66	3.84	3.66	3.75
	Post	3.93	3.95	4.08	3.77	3.62	3.87
	Change	**0.10**	**0.20**	**0.42****	**-0.07**	**-0.04**	**0.12**
	P-Value	0.714	0.364	0.068	0.797	0.901	0.268
Interes t	Pre	4.53	4.66	4.84	4.94	4.45	4.68
	Post	4.07	4.78	5.03	4.55	3.73	4.44
	Change	**-0.46**	**0.12**	**0.19**	**-0.39**	**-0.72**	**-0.24**
	P-Value	0.353	0.691	0.639	0.313	0.151	0.205
Effort	Pre	6.41	6.40	6.34	6.57	6.50	6.45
	Post	6.08	6.07	6.38	6.30	6.08	6.18
	Change	**-0.33**	**-0.33**	**0.04**	**-0.27**	**-0.42****	**-0.27***
	P-Value	0.140	0.134	0.960	0.133	0.077	0.004

Note: *Significant Change: $p < .05$
Note: **Significant Change: $.05 < p < .10$

First, I examine differences in the change between pre- test and post-test averages overall. Using t-tests so look for significant difference between pre- and post-test scores for each of the six component averages, there were two component average differences that resulted in a significant change. Overall, students scored significantly higher on the Affect component, on average, at the end of the semester. This higher average indicates a more positive attitude toward how students feel about the subject of statistics. For the Effort component, a significant decrease was found from pre- to post-test scores, on average, for all students in the study. This decrease in average score indicates that students ranked the amount of effort they *actually* put into the course (post-test score) lower than the amount of work they *expected* to put into the course (pre-test score). This does not represent a difference in attitude (positive or negative) so much as a difference between expected effort and actual effort. So in this case, a decrease in average score, does not indicate a more negative attitude.

When the data is grouped by section, more differences emerged. Section 4 (P, FTF) and Section 9 (P, H), which were both project-based sections, each showed no significant change from pre-test to post-test score averages for any of the 6 components. For section 6 (Q, FTF), two of the component averages showed a significant change. For the variables of Affect and Cognitive Competence, students showed a significant increase, on average, from pre- to post-test score. This higher average indicates that students in section 6 (Q, FTF) showed a positive shift in how they felt about statistics (affect) and in their knowledge and skills when applied to statistics (cognitive competence) by the end of the semester. For section 7 (Q, FTF), students showed a significant increase from pre- to post-test score average for the component of Difficulty. A decrease in a difficulty score does not represent a more negative attitude, but rather a difference in how hard students feel the subject matter of statistics is. Scoring significantly lower on the post-test indicates that students section 7 (Q, FTF) *anticipated* the subject of matter of statistics (pre-test) was going to be harder than they *actually* felt like it was (post-test) by the end of the semester. Finally, for section 10 (Q, H), students' attitudes for the Effort component significantly decreased from pre- to post-test score, on average. As discussed earlier, a decrease in effort does not indicate a more negative attitude, but rather a difference in the *expectation* between the work they thought they would put into the course at the beginning of the semester and the amount of work they found they actually put into the course at the end of the semester.

In each case, when changes were found in the table above, they did not involve sections that were project-based; the only significant changes in attitudes occurred in quiz-based sections. This is an interesting result because it is similar to what I saw when I examined post-differences alone. For those results, I also saw that any differences that existed involved only quiz-based sections. No significant differences, therefore, either in the examination of post-test scores, or in the examination of change from pre- to post-test scores involved project-based sections. As indicated early, the absence of change for the project-based sections could be discouraging because student attitudes did not become more positive over the course of the semester, as I had hoped. However, it is also encouraging to note that while there was not a positive increase in attitude, there also was not a negative change observed either. There seemed to be very little change, overall, in project-based sections in terms of student attitudes over the course of the semester. It is possible that students in project-based sections entered the course with positive attitudes already, and that those attitudes remain unchanged throughout the semester. Therefore, no change in attitude does not indicate a negative result, just a lack of change from beginning to end of the semester.

Grouping by assessment method

When grouped by assessment method, there was just one significant change found in the component average differences for the project-based sections. Two significant changes were found for the quiz-based sections. Table 24 on the following page shows the change in attitude from pre- to post-test scores for students when grouped by assessment method. The change for each component is presented in bold print. P-values for the t-tests performed when analyzing differences between pre- and post-test scores for each component are also provided.

Table 24 Change in Component Average (Post-Test – Pre-Test) by Assessment Method

		Project-Based	Quiz-Based
Affect	Pre	4.74	4.44
	Post	4.83	4.93
	Change	**0.09**	**0.49***
	P-Value	0.779	0.036
Cognitive Competence	Pre	5.15	5.04
	Post	5.20	5.30
	Change	**0.05**	**0.26**
	P-Value	0.865	0.226
Value	Pre	4.91	4.86
	Post	4.83	4.90
	Change	**-0.08**	**0.04**
	P-Value	0.756	0.828
Difficulty	Pre	3.84	3.69
	Post	3.85	3.89
	Change	**0.01**	**0.20**
	P-Value	0.941	0.174
Interest	Pre	4.74	4.65
	Post	4.30	4.53
	Change	**-0.44**	**-0.12**
	P-Value	0.172	0.611
Effort	Pre	6.49	6.42
	Post	6.19	6.17
	Change	**-0.30***	**-0.25***
	P-Value	0.035	0.039

Note: *Significant Change: $p < .05$

For the Effort component, there was a significant decrease from pre- to post-test score, on average, for students in project-based sections. When separated by individual sections, both project-based sections were close to showing a significant change (Table 23). Both p-values for Effort were just over 0.010. When the sections were combined (Table 24), the increase in the sample size led to the change for this component as being found significant. Effort is one component, as discussed earlier, where a decrease does not indicate a more negative attitude, but a difference in expectation. That is, students in project-based sections found that the amount of effort they actually put into the course was significantly lower than what they expected they would put into the course. For the quiz-based sections, significant changes were found for two of the component averages. Scores for the Affect component showed a significant increase, on average, for students in quiz-based sections. This means that students in quiz-based sections did have more positive feelings toward the subject of statistics at the end of the semester than they had at the beginning of the semester. Also, just like for students in project-based sections, scores for students in quiz-based sections were significantly lower on the Effort component at the end of the semester.

Grouping by learning environment

In the final grouping for this research sub-question involving change in attitudes, students were grouped by learning environment. The highest number of significant differences in the change from pre- to post-test scores were observed when looking at this grouping as compared to any previous grouping. Table 25 on the following page provides information on scores for the pre- and post-test averages as well as the change between averages when grouped by learning environment. The changes are presented in bold print. P-values from the t-tests performed to look for significant differences in the change from pre- to post-test scores are also provided.

Table 25 Change in Component Average (Post-Test – Pre-Test) by Learning Environment

		Face-to-Face	Hybrid
Affect	Pre	4.57	4.54
	Post	5.11	4.53
	Change	**0.54***	**-0.01**
	P-Value	0.013	0.980
Cognitive Competence	Pre	5.07	5.10
	Post	5.49	4.90
	Change	**0.42***	**-0.20**
	P-Value	0.031	0.480
Value	Pre	4.89	4.86
	Post	5.10	4.51
	Change	**0.21**	**-0.35**
	P-Value	0.247	0.196
Difficulty	Pre	3.75	3.75
	Post	3.98	3.69
	Change	**0.23****	**-0.06**
	P-Value	0.076	0.789
Interest	Pre	4.68	4.69
	Post	4.64	4.13
	Change	**-0.04**	**-0.56****
	P-Value	0.849	0.082
Effort	Pre	6.39	6.53
	Post	6.18	6.19
	Change	**-0.21****	**-0.34***
	P-Value	0.064	0.020

Note: *Significant Change: $p < .05$
Note: **Significant Change: $.05 < p < .10$

First, I examine the face-to-face sections. Students in these sections scored significantly higher for three of the components: Affect, Cognitive Competence, and Difficulty. Students in these face-to-face sections scored significantly lower on the post-test for the Effort component. This represents a positive shift for students in face-to-face sections on their feelings about the subject of statistics (affect) and their skills and knowledge about statistics (cognitive competence). The increase in difficulty and decrease in effort do not represent a positive or negative shift in attitude so much as a difference between expectation at the beginning of the semester and what actually occurred by the end of the semester. That is, students in face-to-face sections viewed statistics as a more difficult (higher average score) by the end of the semester than what they expected it to be at the beginning of the semester. In addition, they viewed their effort in the course as less (lower average score) than what they expected it would be at the beginning of the semester.

For students in the hybrid sections, there was a significant decrease in the Interest component average from pre- to post-test scores. There was also a significant decrease in the Effort component average, similar to what was found for the face-to-face sections. This indicates a more negative attitude toward statistics for students in hybrid sections in terms of their individual interest in the subject of statistics. By the end of the semester, students in hybrid sections rated themselves as having a significantly lower interest in statistics than they did at the beginning of the semester. In terms of effort, the decrease indicates that the amount of work students actually put into the course (post-test) was significantly lower than what they expected to put into the course (pre-test) at the beginning of the semester.

Therefore, in addressing this second research sub-question involving students' attitudes, when evaluating the change that occurred over the course of the semester for students, the clearest findings can be found when the five sections were examined separately (Table 23). In this evaluation by section, the learning environment and assessment method factors are kept separate. There were 6 component difference averages across 5 sections. This yielded 30 different pre- to post-test score changes in attitudes. Of these 30 changes, only 4 of these were changes were significantly different ($p<.10$). All four of these changes involved quiz-based sections. There were no significant changes for project-based sections.

The first two sub-questions for this research question about student attitudes yielded similar results. In each case, whether examining differences between post-test scores, or examining change over the course of the semester from pre- to post-test scores, the highest number of significant differences involved a learning environment section or comparison. It seems that learning environment, more than assessment method, had a stronger impact on student attitudes. Overall, though, there were more *non*-significant differences than significant differences. So while changes did exist, and these changes tended to involve a learning environment difference or comparison, perhaps the biggest takeaway is how many components did *not* show a significant difference either between post-test scores, or in the change from pre- to post-test scores over the course of the semester.

Third Attitudes Research Sub-Question

I now move on to the third and final sub-question for this research question involving student attitudes in statistics.

> 2. How do students' attitudes toward and ideas of statistics compare between those in project-based sections of statistics and those in quiz-based sections?

2c. How do students in project-based sections differ in their self-reported ratings at the end of the semester of their confidence in their ability to perform various statistical tasks compared with students in quiz-based sections?

This research question has been revised to focus on just the quantitative results obtained from the course evaluation form. The results for the qualitative portion of this course evaluation were not helpful in addressing this research question. Many students interpreted the questions as relating directly to this particular course of statistics with this particular instructor. For example, when asked what the most important thing was about statistics, several students provided responses like, "Use the handouts to study for the tests. The tests are similar to the problems we do in class." Because so many students provided course-specific responses and not general responses to the subject of statistics, data did not reveal information about attitudes. Thus, I could not use the results from the open-ended questions on the course evaluation form to address my research question comparing students when grouped by assessment method.

This third part of the research question regarding student attitudes, therefore, focuses on the quantitative feedback that students provided on this course evaluation form. Students were asked to rate themselves on a scale of 1 to 5 for how confident they felt about their ability to complete six different tasks related to statistics: write a research question, design a good or scientific way to gather data, collect their own data, use the computer to make an appropriate chart of graph for a set of data, analyze their own data, and write a report about data they analyzed for a research project. The scale students used to evaluate their level of confidence was as follows:

1: Not confident at all

2: Below average confidence

3: Average confidence

4: Above average confidence

5: Very confident

I conducted four separate analyses for this data: ANOVA for the mean response between the six task-related items across all sections combined, ANOVA for differences in each task-related item when students were grouped by section, t-test for differences in each task-related item when students were grouped by assessment method, and t-test for differences in each task-related item when students were grouped by learning environment. Significant differences were found in each of the four analyses completed with this data set.

Overall results

For the first analysis I combined the data for all sections to study whether students overall reported a higher average level of confidence for any particular of the six task-related items on the evaluation. I conducted an ANOVA test ($p=.002$) which indicated that there was at least one significant difference between these average confidence ratings reported by students for these six items. Using the Tukey-Kramer follow-up procedure for multiple comparisons, three pairwise comparisons were identified as being significantly different. The average confidence rating for the item pertaining to students collecting their own data was significantly higher than the average rating for the following three items: write a research question, design a good or scientific way to gather data, and write a report about data I analyzed for a research project. Table 26 below shows the overall average for all sections combined for each of the six items, based on the 5-point confidence level scale. The standard deviation for each item is presented in parentheses.

Table 26 Average Confidence Ratings for all Sections Combined

Item	Overall Mean
1. Write a research question	3.78* (0.97)
2. Design a good or scientific way to gather data	3.79* (1.00)
3. Collect my own data	4.25* (0.89)
4. Use the computer to make an appropriate chart of graph for a set of data	3.91 (1.12)
5. Analyze my own data	3.94 (0.94)
6. Write a report about data I analyzed for a research project	3.75* (1.03)

Note: $*\mu_3 > \mu_1, \mu_2, \mu_6$

Grouping by section

In the second analysis, students were grouped together by section. For each section, the mean confidence rating was calculated for each of the six task-related items. I performed an ANOVA for each of the six items, looking to see if any item showed at least one significant difference among any of the five section means for that item. The ANOVA for two of the task-related items showed that there was at least one significant difference ($p < .05$) among section means for each of those items. I used the Tukey-Kramer follow-up procedure for multiple comparisons to identify which means were significantly different. In each case, there was just one pairwise comparison among section means that showed a significant difference. For the task-related item regarding designing a good or scientific way to gather data, students in section 7 (Q, FTF) rated their confidence higher, on average, than students in section 10 (Q, H). For the task-related item regarding using the computer to make an appropriate display of data, students in section 9 (P, H) rated their confidence higher, on average, than students in section 10 (Q, H). There were two other task-related items where differences approached significance ($.05 < p < .10$) in the ANOVA. For the task-related item regarding confidence in writing a research question, students in section 9 (P, H) rated themselves higher, on average, than students in section 10 (Q, H). Also, for the item regarding writing a report about data analyzed for a research project, students in section 9 (P, H) again rated their confidence ability higher, on average, than students in section 10 (Q, H). The full results are provided in Table 27 below. The standard deviation for each section is provided in parentheses. In the final column on the table below, I report the p-value from the ANOVA for that particular item. Table 28 provides information on the pair-wise differences identified on items where a significant difference among section means was found.

Table 27 Average Confidence Ratings by Section

	04 (P, FTF)	06 (Q,FTF)	07 (Q,FTF)	09 (P,H)	10 (Q,H)	P-Value
Write a research question	3.88 (0.85)	3.91 (0.87)	3.81 (1.12)	4.05 (0.89)	3.20 (1.01)	.052**
Design a good or scientific way to gather data	3.96 (0.81)	3.91 (0.92)	4.09 (1.15)	3.70 (0.92)	3.20 (1.01)	.034*
Collect my own data	4.46 (0.66)	4.33 (0.86)	4.14 (1.04)	4.40 (0.68)	3.90 (1.12)	.241
Use the computer to make an appropriate chart of graph for a set of data	4.00 (1.14)	4.05 (1.20)	3.68 (1.25)	4.45 (0.69)	3.35 (0.99)	.023*
Analyze my own data	3.92 (0.93)	4.18 (0.91)	4.00 (1.07)	3.85 (0.88)	3.75 (0.91)	.637
Write a report about data I analyzed for a research project	3.83 (0.87)	3.77 (0.87)	3.72 (1.28)	4.15 (0.81)	3.25 (1.16)	.094**

Note: *ANOVA found significant difference among means ($p < .05$)
Note: **ANOVA found significant differences among means ($.05 < p < .10$)

Table 28 Pair-wise Differences for Items with at least one Significant Difference Among Section Means

Task-related Item	Significant Pairwise Difference
Write a research question	$\mu_9(P,H) > \mu_{10}\,(Q,H)$
Design a good or scientific way to gather data	$\mu_7(Q,FTF) > \mu_{10}\,(Q,H)$
Use the computer to make an appropriate chart of graph for a set of data	$\mu_9(P,H) > \mu_{10}\,(Q,H)$
Write a report about data I analyzed for a research project	$\mu_9(P,H) > \mu_{10}\,(Q,H)$

Grouping by assessment method

For the third analysis, students were grouped together by assessment method. A significant difference was found for one of the task-related items: students in project-based sections rated their confidence higher, on average, than students in quiz-based sections in their ability to use the computer to make an appropriate chart or graph for a set of data. Two other task-related items showed differences that approached significance ($.05 < p < .10$). Students in project-based sections seemed to score themselves higher, on average, for confidence in their ability to both collect their own data and write a report about data they analyzed than students in the quiz-based sections. Table 29 on the following page shows the average confidence ratings, with standard deviation in parentheses, for each assessment type on each of the six items. The p-value for each test is reported in the final column.

Table 29 Average Confidence Ratings by Assessment Method

Item	Project-Based (N=44)	Quiz-Based (N=64)	P-Value
Write a research question	3.95 (0.86)	3.65 (1.03)	.113
Design a good or scientific way to gather data	3.84 (0.86)	3.75 (1.08)	.643
Collect my own data	4.43 (0.66)	4.13 (1.01)	.082**
Use the computer to make an appropriate chart of graph for a set of data	4.20 (0.98)	3.70 (1.17)	.021*
Analyze my own data	3.89 (0.89)	3.98 (0.97)	.595
Write a report about data I analyzed for a research project	3.98 (0.85)	3.59 (1.12)	.058**

Note: *T-test shows significant difference between means ($p < .05$)
Note: **T-test shows significant difference between means ($.05 < p < .10$)

Grouping by learning environment

For the final analysis, students were grouped together by learning environment. A significant difference was found for just one of the task-related items: students in face-to-face sections rated their confidence higher, on average, than students in hybrid sections in their ability to design a good and scientific way to gather data. Table 30 on the following page shows the average confidence ratings, with standard deviation in parentheses, for each learning environment on each of the six items. The p-value for each test is reported in the final column.

Table 30 Average Confidence Ratings by Learning Environment

Item	Hybrid (N=40)	Face-To-Face (N=68)	P-Value
Write a research question	3.63 (1.03)	3.87 (0.94)	.218
Design a good or scientific way to gather data	3.45 (0.99)	3.99 (0.95)	.006*
Collect my own data	4.15 (0.95)	4.31 (0.86)	.361
Use the computer to make an appropriate chart of graph for a set of data	3.90 (1.01)	3.91 (1.19)	.963
Analyze my own data	3.80 (0.88)	4.03 (0.96)	.220
Write a report about data I analyzed for a research project	3.70 (1.09)	3.78 (1.01)	.702

Note: *T-test shows significant difference between means ($p < .05$)

As a whole, these self-reported confidence ratings showed several interesting results. First, the item that students, overall, reported their highest level of confidence in performing was the task of collecting their own data. This is surprising because the students in the quiz-based sections had no opportunity to collect their own data during the semester while students in the project-based had several opportunities. I was surprised that, when grouped by assessment method, this item showed a result that only approached significance ($p=.082$) in favor of the project-based sections. This indicates that students in quiz-based sections felt confident that based on their learning during the semester that they could collect their own data if they needed to.

In general, there was not as many differences as I expected when students were grouped by assessment method. Despite being given multiple opportunities during the semester to perform these six statistical tasks with their own data, students in project-based sections rated their confidence higher, on average, than students in quiz-based sections on only one item ($p < .05$), with just two of the remaining five items showing differences that approached significance ($.05 < p < .10$). I would have expected that the experience of actually performing these tasks in the project-based sections would have led students in those sections to rate themselves consistently higher in their ability to perform these tasks. The fact, therefore, that students in quiz-based sections feel as confident as students in project-based sections, on average, on several of these task-related items is both discouraging and encouraging. It is discouraging because one outcome I hoped to see by using the design of the assessments with projects was that by giving students the opportunity to experience statistics in the real world, that it would increase their confidence in the practice of statistics. While this was true in some cases, it was certainly not true for every task. These results are encouraging, too, however, because they indicate that students in quiz-based sections can exhibit confidence levels as high as students in project-based sections simply by practicing the material in class with data sets from the book, without being exposed to opportunities to collect and study their own data. Ideally, I want all students to feel confident in their ability to perform statistical tasks as a result of taking this course. I was pleased that every average confidence rating calculated in all four analyses was at least 3.0. This indicates that students, on average, across all sections, regardless of learning environment or assessment method, are exiting the course with at least an average confidence level in their ability to perform these statistical tasks.

The most interesting results came from looking at differences between sections. Four of the six items showed differences that were significant ($p < .05$) or that approached significance ($p < .10$) among means. In every instance, the section that rated their confidence levels significantly lower than another section was section 10 (Q, H). This does not surprise me, however. Students in the quiz-based, hybrid section had the least amount of exposure to these task-related items. They were not in a project-based section, and so therefore were not given opportunities during the semester to practice statistical tasks outside of class. In addition, because they were not in a face-to-face section, and so their time in the classroom was half of what their face-to-face counterparts experienced. It seems that in terms of measuring confidence in ability to perform statistical tasks, students in the quiz-based, hybrid section felt the least confident, on average, when compared to any other assessment method and learning environment combination. That being said, their average confidence rating on each of the six task-related items was still at least 3.0 in every case. Therefore, as I concluded earlier, despite there being several items on which section 10 (Q, H) rated themselves significantly lower, on average, than another section, their average confidence ratings never went below 3.0. So while they may not as felt as confident as students in other sections, they still exhibited an average confidence level (at least 3.0) for all tasks. Overall, while some differences did exist, students in every section exited the course with at least average (3.0) confidence levels on all six of these tasks, on average. This is encouraging because students, in general, are leaving this introductory course in statistics feeling at least an average level of confidence in their ability to perform these statistical tasks.

Therefore, to summarize the results for this third sub-question regarding students' attitudes when grouped by assessment method, three out of the six items on the quantitative feedback portion of the course evaluation form regarding students' confidence in their ability to perform certain statistical tasks yielded significant differences. Students in project-based sections rated their confidence ability higher than students in quiz-based sections for the following three items: collecting their own data, using the computer to make an appropriate chart or graph for a set of data, and writing a report about data they analyzed for a research project.

Summary for Attitudes Research Question

Overall, when looking at the second research question in its entirety, a few takeaway results emerge. First, there were not as many differences or changes in attitude as I would have expected. In the examination of post-test scores from the SATS, only four of the six components resulted in any significant differences among the five sections. In each of those four cases, just one pairwise comparison showed a significant difference. In the examination of change between pre- and post-test scores on the SATS, with 30 changes calculated (six for each component across five sections), just 4 of those 30 changes were significantly different. This illustrates that while discussing the changes that *did* occur is valuable, the fact that so many differences and changes were non-significant is also an important result. Secondly, when differences did exist, learning environment was generally a more significant factor than assessment method. For all four pairwise differences identified in follow-up analyses when examining post-test scores on the SATS, the difference involves a quiz-based, face-to-face section scoring higher than a quiz-based, hybrid section. In fact, there were no significant differences at all found among post-test scores on the SATS when students were grouped by assessment method, but when grouping by learning environment, four of the six component averages were significantly higher for face-to-face sections when compared with hybrid sections. When looking at change from pre- to post- test scores on the SATS, not one of the four changes involved a project-based section. Third, the course evaluation results yielded slightly different results than the SATS. For the course evaluation, more significant differences emerged when grouping by assessment method than when grouping by learning environment. Four of the six statistical tasks showed a significant difference for assessment method, with the project-based sections scoring higher in each instance. When grouped by learning environment, there was just one significant difference found among statistical tasks, with the face-to-face section scoring higher. Overall, just as with the results of the academic performance piece of this study, the learning environment/assessment method combination was a factor on the course evaluation. When grouped by section, four pairwise difference emerged following significant ANOVA results. In each case, section 10, a quiz-based, hybrid section, scored lower than another section.

Chapter 6

DISCUSSION

At the beginning of this EPP, I set forth two conjectures. First, I conjectured that students in project-based assessment sections could perform at least as well as students in quiz-based sections on both the midterm and final exam. Secondly, I conjectured that students in project-based sections would have better dispositions, on average, toward statistics at the end of the semester than students in quiz-based sections.

First, I examine my first conjecture. The results for the midterm supported this conjecture. No significant differences were found in the average midterm score between students when grouped by assessment method. That is, students in the project-based sections did perform at least as well as students in the quiz-based sections, on average. The results when comparing performance on the in-class final exam, however, did yield significant differences. Unfortunately, students in the project-based sections scored significantly lower, on average, than students in the quiz-based sections. One reason for this difference in the in-class final exam results could simply be that the in-class final exam did not measure the benefits the students in project-based sections received from their experience. If the in-class final exam had measured different skills, like, for example, the ability of a student to design a good survey question, or the ability of a student to write a paragraph summarizing the results of a study, it could be that students in project-based sections would score higher, on average, than students in quiz-based sections. In other words, in some ways, one might not be surprised by the results of this study, that students in quiz-based sections would score higher, on average, than students in project-based sections on a final exam that emphasized skills reinforced more directly with the quiz-based approach. I had hoped that due to having examples in the class notes similar to the quiz-based and exam format, and having opportunity to practice the computations and procedures emphasized on the final exam through the use of projects, that this would allow students in project-based sections to be sufficiently prepared to perform on the final exam at the same level as their peers in the quiz-based sections. In this study, however, that was not the case.

These results are somewhat complicated, though, due to the additional variable of learning environment. The multiple linear regression analysis provided deeper insight into the interaction between the assessment method and the learning environment. It turned out that the combination that resulted in the highest in-class final exam score, on average, was the face-to-face sections who were assessed with quizzes. The combination resulting in the lowest in-class final exam score, on average, was the hybrid section assessed with projects. One major limitation of this study was that assessment method could not be linked directly with academic performance without considering learning environment.

Further exploration about the interaction between assessment method and learning environment demonstrates how intertwined these factors are. Two of the measures I used for academic performance were in-class exams: midterm and the in-class portion of the final exam. In each case, section 4 (P, FTF) scored lower than section 7 (Q, FTF). Could that difference be attributed to assessment method? That is one possibility. It is also possible that students in section 7 (Q, FTF) were simply stronger test-takers that students in section 4 (P, FTF). Perhaps students in section 7 (Q, FTF) were more comfortable in an in-class testing environment and had less anxiety that students in section 4 (P, FTF). It is true that students in section 7 (Q, FTF) experienced in-class testing during the semester through multiple in-class quiz experiences, and that students in section 4 (P, FTF) did not have this opportunity throughout the semester to take in-class quizzes. Perhaps these experiences readied students in section 7 (Q, FTF) for the in-class midterm and final exam in ways that were not available for students in section 4 (P, FTF). Interestingly, for the take-home final exam portion, students in sections 4 (P, FTF), 6 (Q, FTF), and 7 (Q, FTF) (all the face-to-face sections) performed higher, on average, than students in section 10 (Q, H). Perhaps students in section 4 (P, FTF) *did* learn the material as well as students in section 7 (Q, FTF), but they were not able to demonstrate this on an in-class testing experience (in-class final exam) as well as they were in an out-of-class testing experience (take-home final exam portion).

In some ways, I feel that grouping by assessment method or by learning environment was somewhat misleading. Differences could be hidden by the fact that the assessment method group contained one hybrid section and one face-to-face section. An example of this complexity of interpretation can be found by taking a closer look at Midterm Exam performance, for example. When grouped by assessment type, there was no significant difference between the groups. The project-based average was based on two sections. Section 4 (FTF) had an average of 68.0 and section 9 (H) had an average of 72.2. Together that resulted in a project-based combined average of 70.0. The quiz-based average was based on three sections. Section 6 (FTF) and Section 7 (FTF) had averages of 73.2 and 79.0, respectively. Section 10 (H) had an average of 68.4. Together, the combined quiz-average for the three sections was 73.6. It is evident that students in the hybrid section brought down the overall quiz-based average. Looking at the midterm score by sections gives a more complete story. In the analysis of midterm exam scores when grouped by section, significant differences did exist between the sections. Section 7 (Q, FTF) scored significantly higher than section 4 (P, FTF) and section 10 (Q, H). This illustrates the difficulty of interpreting results with two interacting effects. We cannot conclude that the quiz-based format was the reason section 7 (Q, FTF) did better than section 4 (P, FTF) because section 10 (Q, H) performed just as poorly as section 4, on average, and section 10 was also a quiz-based section. We also cannot conclude that the face-to-face learning environment for students in section 7 (Q, FTF) was the reason they performed better, on average, than section 10 (Q, H), which was a hybrid section, because students in section 4 (P, FTF) performed just as poorly as students in section 10 (Q, H).

Ultimately, I learned that studying assessment methods with the additional effect of learning environment differences proved challenging. With only one section in three of the four assessment method/learning environment combinations, and two sections in the fourth assessment method/learning environment combination, I don't feel there were sufficient data to draw strong conclusions. Either this study could be conducted for multiple semesters so that the number of sections in each assessment method/learning environment would be increased, or else the study, and instructor's schedule, could be re-done and include sections with different assessment methods, but only one learning environment. If the study was repeated with only face-to-face sections, then it would be easier to compare assessment method directly, for example. Therefore, a limitation of this study was the limited number of sections per assessment method/learning environment combination.

My second conjecture involved students' attitudes toward statistics. When using the Survey of Attitudes Toward Statistics (SATS), the results were less complicated than the results for the previous research question involving academic performance. Looking by section, four of the six component averages showed significant differences among sections on the post-test. In *every* case, the section with the higher average was a quiz-based, face-to-face section. Also, in *every* case, the section with the lower average was section 10, which was a quiz-based, hybrid section. Therefore, something about the learning environment seemed to impact student attitudes on several components, and this was *not* confounded by the factor of assessment method. In this case, the results, when grouped by assessment method and by learning environment confirmed what the by section analysis indicated. No significant differences were found among any of the six component averages for the post-test scores when students were grouped by assessment method. This matched with, by section, no significant differences found involving either project-based section. However, when grouped by learning environment, four of the six component averages for the post-test were significantly higher in the face-to face sections as compared with the hybrid sections. This confirms that learning environment had a stronger impact on student attitudes as recorded on the post-test survey than assessment method.

There were not many significant differences in the change of attitudes, as recorded in the difference between pre- and post-test scores, over the course of the semester. I had hoped that students in the project-based sections would show a more positive increase change in attitude over the course of the semester than their peers in the quiz-based sections. This was not supported by the results. However, it is interesting to note that there was not much change in attitude, regardless of assessment approach. The most consistent change that occurred when looking at all six components was the negative increase in the Effort component. As discussed in the Results section, however, this does not reflect a more negative shift in attitude; a decrease in the Effort component merely indicates that the amount of work students expected to do was higher than the amount of work they felt they actually did at the end of the course. While this was discouraging that students, in general, would not have more positive attitudes at the end of the semester, these results do fall in line with the majority of research which suggests that student attitudes are hard to change, and sometimes even become more negative over the course of the semester (Evans, 2007; Sizemore & Lewandowski, 2009 Schau & Emmioglu, 2012).

The course evaluation piece of the attitude research question focused on self-efficacy. My findings about self-efficacy mirrored the results found in the findings from the SATS. In both by section analyses, significant differences emerged. In every instance, the section that scored consistently lower than another section was section 10, a quiz-based, hybrid section. The results of my investigation regarding student attitudes and self-efficacy indicate that being in a quiz-based, hybrid section left students at a disadvantage. Their attitudes and reports of self-efficacy for certain statistical tasks were both lower, on average, at the end of the semester, than students in any other assessment method/learning environment combination. As noted in chapter 5, learning environment, more than assessment method, had a stronger impact on both student attitudes and self-efficacy in this study.

I set out to compare assessment methods in regards to academic performance and attitude; in the end, it turned out that learning environment proved to be the more impactful factor in this study. In exploring other research, I found that much has been recorded on the use of online models of instruction in statistics. Mills and Raju (2011) provide an extensive overview of the development of online courses in statistics over a decade. One of the challenges of studying learning environment is that two studies that refer to a hybrid course, for example, in statistics could define their hybrid course in very different ways. My hybrid course did not rely on technology heavily and did not use an online course management system. It was, for all intents and purposes, a traditional course where approximately half of the notes were presented through an online Power Point presentation. Contrast this with Ward (2003), who designed a study similar to my own where she compared hybrid and traditional courses on several academic performance measures as well as an attitude component. Her traditional and hybrid courses both looked very different than the ones I taught, though. In both of her courses, students met at times throughout the semester in a computer lab where they worked on student-generated data projects, worked collaboratively with classmates, and engaged in activities like applet demonstrations. For the hybrid course, students participated in a chat room, and used course content modules to navigate through the course objectives. She found no significant difference in the academic performance between the two sections, but a difference found in attitudes favored the hybrid sections. It is difficult to compare her results with my study, however, since our hybrid and traditional courses were designed in very different ways. Gundlach (2015) used three types of learning environments to test for differences: a web-augmented traditional course, a fully online course, and a flipped classroom. Her findings showed that students in the traditional classroom scored significantly higher than students in either the flipped or the fully online classroom for two of the three exams. There was no significant difference for the third exam. In addition, for her attitudes portion, the only significant change in attitudes during the course of the semester occurred in the traditional course. Again, however, there is difficulty in interpreting and comparing results when the learning environments are so varied. For example, was Gundlach's web-augmented traditional course more similar to my face-to-face course, or to my hybrid course? Learning environment, therefore, while worthwhile to study and explore, is also challenging to explore because of the shared terminology that often has varied definitions, depending on the researcher.

One limitation of this study was that it did not connect academic performance and attitudes by student. Because the attitudes piece was anonymous (labeled by section, assessment method, and learning environment only) we don't know if students who scored higher on the midterm and final exam showed more positive dispositions to statistics or not. The literature suggests that there is a connection between these two things. Carlson and Winquist (2011) were able to connect positive attitudes with high performance. Students with higher performance on exams did have more positive attitudes toward statistics at the end of the semester. Wilson (2013) also found that both student attitudes and exam scores were higher for students in a flipped class, as compared to students in a traditional course. Finney and Shraw (2003) were able to connect self-efficacy with academic performance. Students with higher scores on exams, did, on average, report higher levels of confidence in their ability to perform certain statistical tasks. Future research could connect this academic performance and attitudes piece directly for each student, rather than just grouping students by section, assessment method, or learning environment.

In conclusion, this study revealed that using a project-based assessment method had neither a positive or negative effect on student attitudes when compared with using a quiz-based method; attitudes in statistics are hard to change. In terms of academic performance, the results were mixed. While this study indicated that students in quiz-based sections did perform better, on average, than students in project-based sections, on the in-class final exam, this result was confounded by learning environment. On other measures like the midterm exam and take-home final exam portion, there was either no significant difference in performance, or a difference that favored a project-based section. Due to the limited number of assessment method/learning environment combinations, however, it is uncertain if the results would be similar if this study were replicated with future sections of statistics. Overall, the results regarding academic performance were inconclusive with regard to assessment method comparisons.

Moving Forward as a Leader

As a result of this study, I feel prepared to present a project-based assessment structure to my colleagues as a viable alternative to a more traditional quiz-based approach. Even though the results of this EPP showed that students in project-based sections scored lower on the final exam, I believe that there is validity to offering a project-based approach. In my pilot study for this EPP, with data taken from an earlier semester, there was *no* significant difference in final exam performance between students in quiz-based sections and students in project-based sections. Therefore, I feel the data results I have at this point are mixed.

As I think about taking a leadership role in my department at RCGC, I think the best approach I can take in working to strengthen the learning experiences our students are having in statistics is to emphasize that we need to be making decisions based on evidence. To this point, instructors have been free to design their own courses in terms of both how they structure their in-class time and how they assess their students. There is a wide variety of approaches taken in regards to assessment: if/how homework is included, how many quizzes/tests are given, and other assessments (like projects) that may be included. Instructors have complete freedom to design their course with any assessments they choose; in fact, instructors are not even required to give any final exam at all, if they so choose. In addition to differences in assessment method structures, RCGC offers three different learning environment options for statistics: face-to-face, hybrid, and fully online. One goal of our department should be to identify which types of learning environments and what types of assessment method structures are yielding the best outcomes for our students. Without consistency from section to section, it is very difficult to compare grades and outcomes across sections. My research attempted to bring consistency between a face-to-face course and a hybrid course by using a common midterm and final exam. My data currently shows that students in hybrid sections seem to be at a disadvantage for learning the material. This makes me wonder if and how effective a fully online course in statistics is at RCGC.

One challenge our department faces is that both learning environment and assessment method designs can be difficult to compare. Even two courses that are both designated as hybrid courses could be taught in very different ways, for example. A first step I would suggest as a leader would be for each instructor in statistics to document the following information for each of their sections: learning environment (including a brief description of what this means in their course like how class periods are structured, and what, if any, components take place online) and assessment method structure. A second step would be to emphasize the fact that we will not be able to identify what is best for our students until we have data. We will not have data unless we work together to bring at least one common element to all sections. The most logical consistent element to bring to each section is the introduction of a common final exam. This would be a major step for our school and some instructors may be reluctant or even oppose such an idea. To begin with, we could identify just 2 or 3 common questions that address key learning outcomes that all instructors would be required to put on their final exams. This way instructors would have the freedom to design the remainder of their final exam with questions of their choice. The bottom line is that we want to be doing things that work for our students. Until we can identify how well our *current* different learning environments and assessment method structures are working for students, we will not be able to take steps to strengthen our courses for the *future* and bring them closer in line with the GAISE recommendations in ways that are meaningful and that lead to good outcomes (in terms of academic performance and attitudes) for our students.

As the leader for the Statistics course at our college, I am often contacted by adjunct instructors who are looking for advice on how to set up their course. I have informally shared my experiences regarding using a project as one option for them as they set up their courses. I would like to put together a more formal portfolio of mini projects, as well as my semester project, that I could share with both adjunct and full-time instructors at our school. I would also like to set up a time when I can formally present the results of this study to all my STEM colleagues at RCGC. I believe that the inclusion of a research project could be a valuable addition in other courses beyond statistics, as well. The best way for me share my results would be to lead a session during our Professional Development Day experiences which occur twice a year. I plan to request to lead a session during the Fall 2017 semester Professional Development Day that would allow me to both share my results from my EPP as well as invite conversation amongst colleagues in the STEM department about new ways to think about and design assessments in our classrooms.

Through this experience I have also begun to consider how a combination of the two assessment structures I used in this study could produce an even stronger course that would still allow students to use a more hands-on approach with projects, but also give them the practice they need with in-class quizzes as they learn new and more difficult statistical techniques. One thing this study brought out was that perhaps the projects alone did not prepare students as well for the final exam as the quiz-based structure that students in quiz-based sections experienced. Instead of offering only projects during the semester, I am considering having a combination of a few quizzes (particularly as the material gets more difficult in the second half of the course), as well as a few projects. I will not give up the idea of including at least one project in every course. I believe the experience of having students generate and work with their own data is invaluable to them as they learn statistics. To this end, I will be an advocate in my department to introduce faculty to this idea of a revised assessment structure that would include at least one student-generated research project. I believe this is clearly in line with the GAISE College Report (Garfield et al., 2005), and I believe that RCGC and the Dean of STEM would fully support an inclusion of at least one research project in each section of introductory statistics.

Using a project-based assessment method has been rewarding for me, personally. It has allowed me to connect with students by understanding what topics they find interesting and choose to investigate. It appears that students are more engaged and motivated in class when they are applying the course content to their individual research topics outside of class. Student comments in emails and on course evaluations often indicate how much students enjoyed collecting and working with their own data. Students mention that it was refreshing to be in a course where they were able to conduct their own research and learn content material in fun and creative ways. I do not know why these comments I've received over many years of teaching did not translate into a more positive disposition toward statistics on the formal SATS used in this study. But I'm reminded that the majority of research has indicated that student attitudes are hard to change. Therefore, while I have not received the quantitative evidence to support endorsing a project-based assessment method structure in favor of a quiz-based structure, neither can I rule out the perceived value of the project-based assessment method in my courses. I have learned from this study that I need to make sure students in project-based sections have some opportunity in class to practice material in a way that is similar in format to the final exam. In particular, when our college moves toward a common final exam, I will need to work hard to make sure that students in project-based section are not at a disadvantage when compared with their peers in quiz-based sections in their preparation for that common final exam. Overall, I am pleased with the project-based assessment structure I've designed. I will continue to use this structure as I believe it provides students with a way to personally apply the content covered in the course to their everyday lives. I invite others who have used in-class quizzes and tests as their only assessment method to this point in their teaching experience to consider the benefits of introducing student-generated data projects into their courses. I believe they will find the result to be an enjoyable experience for both them and their students.

REFERENCES

Autin, M., Bateiha, S., and Marchionda, H. (2013). Power through struggle in introductory statistics. *PRIMUS, 23*(10), 935-948.

Bailey, B., Spence, D., & Simm, R. (2013). Implementation of discovery projects in statistics. *Journal of Statistics Education, 21*(3), 1-24.

Ben-Zvi, D., and Garfield, J. (2008). Introducing the Emerging Discipline of Statistics Education. *School Science and Mathematics, 108*(8), 355-361.

Carlson, K. and Winquist, J. (2011). Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *Journal of Statistics Education, 19*(1), 1-23. Retrieved from: https://ww2.amstat.org/publications/jse/v19n1/carlson.pdf

Carnell, L.J. (2008) The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education, 16*(1). Retrieved from: https://ww2.amstat.org/publications/jse/v16n1/carnell.html

Core Standards. (n.d.). Mathematics Standards. Retrieved from http://www.corestandards.org/Math/

Curran, E., Carlson, K., and Celotta, D.T. (2013). Changing attitudes and facilities understanding in the undergraduate statistics classroom: A collaborative learning approach. *Journal of the Scholarship of Teaching and Learning, 13*(2), 49-71.

Doehler, K, Taylor, L, and Smith, J. (2013). A study of faculty views of statistics and student preparation beyond an introductory class. *Journal of Statistics Education 21*(1). Retrieved from: https://ww2.amstat.org/publications/jse/v21n1/doehler.pdf

Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *The Mathematics Educator, 17*(2), 24-30.

Finney, S. J., & Schraw, G. (January 01, 2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28,* 2, 161-186.

Franklin, C.A., Kader, G., Mewborn, D., Moreno, J., Peck, R., Perry, M., & Scheaffer, R. (2007). *Guidelines for assessment and instruction in statistics education* (GAISE) *report: A pre-K-12 curriculum framework.* Alexandria, VA: American Statistical Association.

Garfield, J., Aliaga, M., Cobb, G., Cuff, C., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, R., Utts, J., Velleman, P., and Witmer, J. (2005). Guidelines for Assessment and Instruction in Statistics Education (GAISE) project: College report.

Groth, R. (2008). Assessing teachers' discourse about the pre-K-12 guidelines for assessment and instruction in statistics education (GAISE). *Statistics Education Research Journal, 7(1)*, 16-39.

Gundlach, E., Richards, K. A. R., Nelson, D., & Levesque-Bristol, C. (January 01, 2015). A comparison of student attitudes, statistical reasoning, performance, and perceptions for web-augmented traditional, fully online, and flipped sections of a statistical literacy class. *Journal of Statistics Education, 23,* 1. Retrieved from: https://ww2.amstat.org/publications/jse/v23n1/gundlach.pdf]

Lesser, L.M., Wall, A., Carver, R., Pearl, D., Martin, N., Kuiper, S., Posner, M., Erickson, P., Liao, S., Albert, J., and Weber, J. (2013). Using Fun in the Statistics Classroom: An Exploratory Study of College Instructors' Hesitations and Motivations. *Journal of Statistics Education 21*(1). Retrieved from: www.amstat.org/publications/jse/v21n1/lesser.pdf

Libman, Z. (2010). Integrating real-life data analysis in teaching descriptive statistics: A constructivist approach. *Journal of Statistics Education, 18*(1). Retrieved from: https://ww2.amstat.org/publications/jse/v23n1/gundlach.pdf

Love, T. E. (2000). A different approach to project assessment. *Journal of Statistics Education*, *8*(1). Retrieved from: http://www.amstat.org/publications/jse/secure/v8n1/love.cfm

Mills, J. & Raju, D. (2011). Teaching Statistics Online: A decade's review of the literature about what works. *Journal of Statistics Education, 19,* 2. Retrieved from: www.amstat.org/publications/jse/v19n2/mills.pdf

Mvududu, N. (2003). A cross-cultural study of the connection between students' attitudes toward statistics and the use of constructivist strategies in the course. *Journal of Statistics Education 11*(3). Retrieved from: http://amstat.org/publications/jse/v11n3/mvududu.html

Neumann, D., Neumann, M., and Hood, M. (2010). The development and evaluation of a survey that makes use of student data to teach statistics. *Journal of Statistics Education, 18*(1). Retrieved from: www.amstat.org/publications/jse/v18n1/neumann.pdf

Neumann, D., Hood, M., and Neumann, M. (2013). Using real-life data when teaching statistics: Student perceptions of this strategy in an introductory statistics course. *Statistics Education Research Journal 12*(2), 59-70.

Preparing America's Students for Success (2017). Retrieved from: www.corestandards.org

Prins, S. (2009). Student-Centered instruction in a theoretical statistics course. *Journal of Statistics Education, 17*(3). Retrieved from: https://ww2.amstat.org/publications/jse/v17n3/batesprins.html

Schau, C. (2003). *Survey of Attitudes Toward Statistics* (SATS-36). Retrieved from: http://evaluationandstatistics.com/

Schau, C. & Emmioglu, E. (2012). Do Introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal, 11*(2), 86-94.

Sizemore, O.J., & Lewandowski, G.W. (2009). Learning might not equal liking: Research methods course changes knowledge but not attitudes. *Teaching of Psychology, 36*(2), 90-95.

Spence, D. J., Sharp, J. L., and Sinn, R. (2011). Investigation of factors mediating the effectiveness of authentic projects in the teaching of elementary statistics. *Journal of Mathematical Behavior (30)*, 319-332.

Total Registration. (n.d.). AP Exam Score Distribution. Retrieved from https://www.totalregistration.net/AP-Exam-Registration-Service/2013-AP-Exam-Score-Distributions.php

Ward, B. (2004), "The Best of Both Worlds: A Hybrid Statistics Course," *Journal of Statistics Education. 12*(3). Retrieved from: http://www.amstat.org/publications/jse/v12n3/ward.html

Warton, D. (2007). How many words do you know? An integrated assessment task for introductory statistics students. *Journal of Statistics Education, 15*(3). Retrieved from: https://ww2.amstat.org/publications/jse/v15n3/warton.html

Wilson, S. (2013). The flipped class: A method to address the challenges of an undergraduate statistics course. *Teaching of Psychology, 40*(3), 193-199.

Woodward, R. & McGowan, H. (2012). Redesigning a Large Introductory Course to Incorporate the GAISE Guidelines. *Journal of Statistics Education 20*(3). Retrieved from: www.amstat.org/publications/jse/v20n3/woodward.pdf

If you want to number your bibliographic entries, change the style of the items to *Bib Entry - numbered.*

Appendix A

BASELINE MATH SKILLS SURVEY

Circle your choice for each multiple choice question.

1. What is 30% of 60?

   a. 30 b. 180 c. 20 d. 18 c. 9

2. Convert the decimal to a fraction in simplest terms: 0.4

   a. $\frac{2}{50}$ b. $\frac{2}{5}$ c. $\frac{4}{10}$ d. $\frac{1}{10}$ e. $\frac{4}{100}$

3. Simplify the following expression: -4-5-6+6+5-4

   a. -4 b. -8 c. -18 d. 0 e. 4

4. At the grocery store you notice that the soup that you would like to buy is on sale for 40% off. If the original price for one can was $2.00, what is the sale price?

   a. $1.80 b. $1.60 c. $1.40
   b. d. $1.20 e. $0.80

5. Which of the following fractions is the largest: $\frac{2}{3}, \frac{4}{5}, \frac{7}{10}, \frac{5}{6}$

   a. $\frac{2}{3}$ b. $\frac{4}{5}$ c. $\frac{7}{10}$ d. $\frac{5}{6}$

6. What is $\frac{1}{4}$ of 20?

a. 8 b. 5 c. $\frac{1}{5}$ d. $\frac{1}{8}$ e. 80

7. If soup is on a sale of 4 cans for $5.00, how much would 6 cans cost?

a. $7.00 b. $7.10 c. $7.50 d. $4.80 $8.00

8. Compute: $\frac{3}{4} \div \frac{1}{3}$

a. $\frac{1}{4}$ b. $\frac{9}{4}$ c. $\frac{4}{9}$ d. $\frac{4}{7}$ e. 4

9. Simplify the following expression. Write your final answer in simplest form: $\frac{1}{3} + \frac{3}{5}$

a. $\frac{14}{15}$ b. $\frac{4}{15}$ c. $\frac{4}{8}$ d. $\frac{1}{2}$ e. $\frac{1}{5}$

10. You are making a recipe that calls for ¾ cup butter. You would like to cut this amount in half. How much butter will you need?

a. $\frac{1}{2}$ cup b. $\frac{3}{8}$ cup c. $\frac{1}{3}$ cup d. $\frac{2}{5}$ cup e. $1\frac{1}{2}$ cups

11. Find the slope of the line between the points (0,4) and (4,0)

a. -4 b. 4 c. -1 d. 1 e. $-\frac{1}{2}$

12. Carly and Kyle each got lots of candy for Halloween. Together they received 185 pieces of candy, and Kyle's pile contained 15 more pieces than Carly's pile. How many pieces of candy did Kyle get for Halloween?

    a. 105 b. 100 c. 90 d. 85 e. 80

13. Solve the following equation: $2x + 1 = 11$

    a. $x=10$. $x=5$ c. $x=6$ d. $x=20$ e. $x=24$

14. Suppose you took a quiz and got 16 questions correct out of 20. What percent of the questions did you get correct?

    a. 60% b. 70% c. 80% d. 85% 90%

15. If you scored 60, 60, 80, and 100 on 4 quizzes in a course, what would your quiz average be?

    a. 60 b. 70 c. 75 d. 77.5 e. 80

16. Convert the following fraction to a decimal: $\frac{1}{5}$

    a. 0.2 b. 0.5 c. 1.5 d. 2.0 e. 2.5

17. Did you take a statistics course in high school?

    Yes No

18. Have you taken a statistics course in college before this semester?

    Yes No

Appendix B

SURVEY OF ATTITUDES TOWARD STATISTICS PRE-TEST

DIRECTIONS: The statements below are designed to identify your attitudes about statistics. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

	Strongly disagree			Neither disagree nor agree			Strongly agree
I plan to complete all of my statistics assignments.	1	2	3	4	5	6	7
I plan to work hard in my statistics course.	1	2	3	4	5	6	7
I will like statistics.	1	2	3	4	5	6	7
I will feel insecure when I have to do statistics problems.	1	2	3	4	5	6	7
I will have trouble understanding statistics because of how I think.	1	2	3	4	5	6	7
Statistics formulas are easy to understand.	1	2	3	4	5	6	7
Statistics is worthless.	1	2	3	4	5	6	7
Statistics is a complicated subject.	1	2	3	4	5	6	7
Statistics should be a required part of my professional training.	1	2	3	4	5	6	7
Statistical skills will make me more employable.	1	2	3	4	5	6	7
I will have no idea of what's going on in this statistics course.	1	2	3	4	5	6	7

I am interested in being able to communicate statistical information to others.	1	2	3	4	5	6	7
Statistics is not useful to the typical professional.	1	2	3	4	5	6	7
I plan to study hard for every statistics test.	1	2	3	4	5	6	7
I will get frustrated going over statistics tests in class.	1	2	3	4	5	6	7
Statistical thinking is not applicable in my life outside my job.	1	2	3	4	5	6	7
I use statistics in my everyday life	1	2	3	4	5	6	7
I will be under stress during statistics class.	1	2	3	4	5	6	7
I will enjoy taking statistics courses.	1	2	3	4	5	6	7
I am interested in using statistics.	1	2	3	4	5	6	7
Statistics conclusions are rarely presented in everyday life.	1	2	3	4	5	6	7
Statistics is a subject quickly learned by most people.	1	2	3	4	5	6	7
I am interested in understanding statistical information.	1	2	3	4	5	6	7
Learning statistics requires a great deal of discipline.	1	2	3	4	5	6	7
I will have no application for statistics in my profession.	1	2	3	4	5	6	7
I will make a lot of math errors in statistics.	1	2	3	4	5	6	7
I plan to attend every statistics class session.	1	2	3	4	5	6	7
I am scared by statistics.	1	2	3	4	5	6	7

I am interested in learning statistics.	1	2	3	4	5	6	7
Statistics involves massive computations.	1	2	3	4	5	6	7
I can learn statistics.	1	2	3	4	5	6	7
I will understand statistics equations.	1	2	3	4	5	6	7
Statistics is irrelevant in my life.	1	2	3	4	5	6	7
Statistics is highly technical.	1	2	3	4	5	6	7
I will find it difficult to understand statistical concepts.	1	2	3	4	5	6	7
Most people have to learn a new way of thinking to do statistics.	1	2	3	4	5	6	7

Please notice that the labels for each scale on the rest of this page <u>change</u> from item to item.

How well did you do in mathematics courses you have taken in the past?	Very poorly 1	2	3	4	5	6	Very well 7
How good at mathematics are you?	Very poor 1	2	3	4	5	6	Very good 7
In the field in which you hope to be employed when you finish school, how much will you use statistics?	Not at all 1	2	3	4	5	6	Great deal 7
How confident are you that you can master introductory statistics material?	Not at all confident 1	2	3	4	5	6	Very confident 7
Are you required to take this statistics course (or one like it) to complete your degree program?	Yes 1			No 2			Don't know 3

	Not at all likely						Very likely
If the choice had been yours, how likely is it that you would have chosen to take any course in statistics?	1	2	3	4	5	6	7

Scoring the SATS

Subscale scores on the SATS are formed by reversing the responses (1 becomes 7, 2 becomes 6, etc.) to the items indicated with an * and summing the items within each subscale. Using our 7-point response scale, higher scores then correspond to more positive attitudes.

The following lists the items in our subscale structure.

Affect - positive and negative feelings concerning statistics (6 items):

3. I will like statistics.
4.* I will feel insecure when I have to do statistics problems.
15.* I will get frustrated going over statistics tests in class.
18.* I will be under stress during statistics classes.
19. I will enjoy taking statistics courses.
28.* I am scared by statistics.

Cognitive Competence - attitudes about intellectual knowledge and skills when applied to statistics (6 items):

5.* I will have trouble understanding statistics because of how I think.
11.* I will have no idea of what's going on in statistics.
26.* I will make a lot of math errors in statistics.
31. I can learn statistics.
32. I will understand statistics equations.
35.* I will find it difficult to understand statistics concepts.

Value - attitudes about the usefulness, relevance, and worth of statistics in personal and professional life (9 items):

7.* Statistics is worthless.
9. Statistics should be a required part of my professional training.
10. Statistical skills will make me more employable.
13.* Statistics is not useful to the typical professional.
16.* Statistical thinking is not applicable in my life outside my job.
17. I use statistics in my everyday life.
21.* Statistics conclusions are rarely presented in everyday life.
25.* I will have no application for statistics in my profession.
33.* Statistics is irrelevant in my life.

Difficulty - attitudes about the difficulty of statistics as a subject (7 items):

6. Statistics formulas are easy to understand.
8.* Statistics is a complicated subject.
22. Statistics is a subject quickly learned by most people.
29.* Learning statistics requires a great deal of discipline.
30.* Statistics involves massive computations.
34.* Statistics is highly technical.
36.* Most people have to learn a new way of thinking to do statistics.

Interest – students' level of individual interest in statistics (4 items):

12. I am interested in being able to communicate statistical information to others.
20. I am interested in using statistics.
23. I am interested in understanding statistical information.
29. I am interested in learning statistics.

Effort - amount of work the student expends to learn statistics (4 items):

1. I plan to complete all of my statistics assignments.
2. I plan to work hard in my statistics course.
14. I plan to study hard for every statistics test.
27. I plan to attend every statistics class session.

Appendix C

QUIZ #1

All questions worth one point each.

A. Use the codes below to categorize the following variables by type:
QL = Qualitative D = Discrete quantitative C = Continuous quantitative

______ 1. Position of football player (quarterback, running back, etc…)
______ 2. Number of passes a quarterback makes in a game
______ 3. Distance of a pass
______ 4. Number of points the winning team has at the end of the game
______ 5. Color of the uniforms for each team

B. I want to sample people attending a Philadelphia Eagles' football game. Use the codes below to choose the best answer for the following sample descriptions:

SRS = Simple Random Sample
CLS =Cluster Sample
VRS = Voluntary Response Sample
SYS = Systematic Sample
COS = Convenience Sample
STR = Stratified Random Sample

_____ 6. I divide the stadium by its section numbers. From every section in the stadium, I use the first and last person seated in each row for my sample.
_____ 7. A message is posted on the large screen that posts my short survey questions and invites people to text in their answers.
_____ 8. I divide the stadium by its section numbers. I randomly select 6 section numbers from a hat. I go to each of those 6 sections and use everyone seated in those selected sections for my sample.
______ 9. As people enter the stadium, I ask every 20^{th} person my question.

C. Use the codes below to categorize the following variables into their appropriate measurement scales:

N = Nominal O = Ordinal I = Interval R = Ratio

_______ 10. Marital status of patients in a doctor's office
_______ 11. Time required to get to campus daily
_______ 12. Daily high temperature
_______ 13. Age of students in this class
_______ 14. Months of the year

The following are possible research questions for your semester long project. Identify each as either a qualitative (QL) question or quantitative (QN) question.

_______ 15. What is the most common method of payment for customers at Walmart? (cash, debit, credit, etc…)
_______ 16. How many RCGC students have been to a Phillies game this season?
_______ 17. How many wins do you predict the Eagles will have this season?
_______ 18. What is your favorite fast food restaurant?
_______ 19. What is the average winning amount of money for a Wheel of Fortune contestant each night?
_______ 20. What language would you most like to learn?

Appendix D

QUIZ #2

1. (8 points) Using the set of data below, identify the values listed at the bottom of this page. Show your work and place a box around your final answer for each value. Round answers to the nearest tenth, as necessary. You do not need to include a boxplot. These values represent final scores from selected games the Sixers played last season:

78
110
114
115
92
98
103
104
128
117
106
103
89
99
111

Mean:
Mode:
Range:

Five Number Summary: (fill in table below)

Minimum	
Q1	
Median	
Q3	
Maximum	

2. (10 points) Fill in the table regarding the standard deviation. Also, below the table show your work to complete the final two steps for calculating the standard deviation. <u>You must show your work to receive credit for this problem.</u> Round your final answer to the nearest tenth. Place a box around your final answer. You do not need to include a sentence of interpretation. The data concerns the number of points scored per game for 7 randomly selected Sixers games last season. The average for the sample is 101.9 points.

Number of Points	$x - \bar{x}$	$(x-\bar{x})^2$
89	-12.9	166.41
99		
110	8.1	65.61
95		
120		
88	-13.9	193.21
112		
	Sum =	

<u>Additional Work for Calculating Standard Deviation:</u>

Appendix E

PROJECT #1

Step 1: Choose a **quantitative** question to investigate. For this project you may obtain your data from real people (direct contact or Facebook/social media site), observation, or any published source (internet, magazine, television, etc…)

> Examples: What is the average cost of a loaf of bread? How many miles does a GCC student live away from campus? What is the average number of saves for closers in Major League Baseball for the 2011 season? What is the average number of wins for an NHL team last season?

Step 2: Define your population (the entire group of individuals you want to know something about). Decide on your method of sampling. Clearly identify the sampling method you are using to gather data from your population.

> Examples: All loaves of bread sold at Shop-Rite in Brooklawn, NJ; all GCC students; all closers in MLB during the 2011 season; all NHL teams

Step 3: Gather your data. You must have at least **20 values** for your sample size (unless your population has certain limitations – see me if this applies). **You must include a list of your data values in the final report.**

Step 4: By hand, calculate the following things (include hand calculations on a separate page or pages, neatly labeled, at the end of your report). You can then use a computer or calculator to double-check these values, if you wish.

- Five Number Summary
- Mean
- Mode
- Midrange
- Range
- Standard Deviation (*include a sentence of interpretation using the context of your data!)

Step 5: Include 2 charts in the final report:

- Chart #1: Boxplot of your data, well-labeled. This can be done by hand (unless you find a computer program which has this application) and should be included on the final pages with your hand calculations

- Chart #2: Use the computer to display your data in an appropriate chart – histogram or other creative picture. You should not be displaying all the values listed in step 4 (mean, median, mode, etc); simply display your original data (the 20+ original values) in a clear and readable way. **<u>Remember that with quantitative data it is often most effective to group similar values (unless your range of data or sample size is very small).</u>** Include a descriptive title for your chart, labels for your x- and y-axis and a legend (if necessary). One should be able to draw conclusions about your data by looking just at your chart alone! Your chart may be on a separate page of your report, or contained in the body of your report – your choice.

Step 6: Write a report **<u>(minimum 2 full pages, not including charts and hand calculations)</u>** describing this experience. Include information on the following questions:

- How did you choose your topic? Why/how did you choose your population?
- What process did you go through to identify and obtain your sample? Was your sample method 'scientific' (i.e. one of the 'good' sampling methods)? Be sure to name your sampling method (simple random, systematic, convenience, etc…)
- Do you think the data you obtained from your population is a good representation of the population you defined in step 2? Why or why not?
- Were there any outliers in your data? Identify them, if so. Also, how did the outliers affect your calculations in step 4?
- What difficulties (if any) did you encounter completing this assignment (at any point in this assignment, either in the collection of your data or the calculations stage)?
- How did this experience with quantitative data compare with your assignment #1 experience with qualitative data? Which did you enjoy more? Why?
- What new research questions did you think of as you completed this project?
- What would you do differently if you were re-doing this assignment?

Make sure your report is double-spaced use size 12 Times New Roman font. Proof-read and spell check your work! See grade breakdown on the following page.

Put your full name and section number at the top of your assignment.

Graded Components for Assignment #2	Points
Content Following the instructions in each of the 6 steps listed above. Be sure to cover all questions in Step 6.	30
Grammar/Writing Style/Professional Appearance Using complete sentences, correct grammar and readable writing style. Your final product should be ready to hand over for publication with no by-hand corrections or notations except for those allowed in the project description	15
By-Hand Calculations Accurate answers, neatly labeled	20
Charts Follow instructions in project description – be sure your charts display the data clearly and appropriately and include titles and labels as described above	Boxplot: 5 Chart #2: 10
Total	**80 points**

Appendix F

SEMESTER PROJECT

The goal of this project is for you to get first-hand experience with statistics in the "real world". You are free to choose whatever you want to study – this is your opportunity to study something that really interests you, so choose wisely! This project is worth a significant portion of your grade for this semester, so you will want to invest time into this and be thorough in your work. This is your chance to be creative and apply in a fun way all the "book stuff" you've learned. I look forward to seeing your individuality expressed in this project. This is NOT something that should be done at the last minute! It is my hope that all the things we have studied and will be studying in the next several weeks this semester will make sense to you (if they haven't already!) by the end of this project.

Initial Guidelines

1. You may work either individually or with one other person in the class on this project. Working with a partner is strongly encouraged (so you can learn together and help each other) but not required. You may work with someone in another of my statistics sections if you wish. If you work together, both people will receive the same grade.
2. All work handed in for this project must be typed.
3. You must use new data – that is, you cannot use data you collected for any of your previous assignments (if applicable). You may choose a similar question to investigate, but must gather new data.

Your project has been broken down into a number of steps (defined below). Step 1 must be completed by **____________________.** I will review your proposal, grade it, make suggestions for change if necessary, and give it back to you to complete steps 2-6. It is important that you do not move past Step 1 until I have checked your work. (That is, I would not want you to go take your sample and then have to redo it because there was a flaw in your design).

The final project is due _______________. Late projects will receive a 50% penalty. <u>You must hand in a fresh copy of Step 1 (with any changes you have made) along with your final project</u>. Your final project should be in the form of a report – complete sentences, paragraphs, etc. Each step does not need to take up a separate page, but the steps should be presented in order in your paper. Your visual display (Step 3), however, could be on a separate page.

Graded Components of the Project – 80 points total

Your Proposal: 15 points
Content of Final Project (thoroughly covering each step): 30 points
Grammar: 10 points
Writing Style/Explanation/Word Choice: 15 points
(It is ok to use statistical terms, but each one should be completely explained. Your material should be presented in such a way that any non-statistician would both understand and learn from.)
Professional Appearance: 10 points
(Your final project should be ready to hand over for publication!)

Step 1 – This should be about one page, typed, in length. You should address all 4 parts of this step **in paragraph format**. You can also include why this topic is of interest to you.

a. Construct a question of interest.
   Examples:
   What is the mean cost of an 18oz box of cereal?
   What percentage of college students skip class at least once a week?
   What is the average number of cars that go thru McDonald's drivethru over lunch?
b. Once you have your question, you need to define your population. Be as specific and complete as possible! Examples on the following page…
   All kinds of cereal sold at the Pathmark in Deptford, NJ
   All RCGC students
   All cars that go through the McDonalds in Deptford between 11:30 and 1pm on weekdays
c. Write down a hypothesis – a guess at the answer to your question. This could be either one number of a range of numbers.
   Examples:
   $3.25 for a box of cereal
   15% of RCGC students
   Between 45 and 55 cars
d. Design a way to take a good sample from the population you have defined. Be specific and complete in your description. In earlier assignments you were free to use less scientific methods of sampling, but this project should incorporate a more scientific method of sampling.

Step 2

Go take your sample! Carefully record each data observation you collect. You should present a list of your data somewhere in your final report so that your analysis can be verified if necessary.

Step 3
Using the computer, organize your data visually in an appropriate, clear, and concise way. You should use a bar chart, pie chart, histogram, or pictograph. See me if you have other creative ideas for a display. All items on your graph should be clearly labeled and the display should have an appropriate title. You can use more than one display if you wish.

Step 4
Construct a confidence interval for your parameter of interest. You should cover each of the steps we did in class for confidence intervals. <u>It is best to present the steps in paragraph, rather than list form. If you are more comfortable, you can use the list form we do in class, however, you must make sure that either as a part of the list, or in a paragraph before or after the list, you completely explain all the statistical terms you use.</u> Basically, explain what you do, and why! Be sure to define what a confidence interval is in your work. You can choose whatever level of confidence you wish. Does the confidence interval give you insight toward your guess from Step 1? Explain.

Step 5.
Perform a hypothesis test in order to test your guess from Step 1. You should cover each of the steps we did in class for hypothesis tests. <u>It is best to present the steps in paragraph, rather than list form. If you are more comfortable, you can use the list form we do in class, however, you must make sure that either as a part of the list, or in a paragraph before or after the list, you completely explain all the statistical terms you use.</u> Again, explain what you do at each step, and why! Also, include a conclusion that clearly explains how the hypothesis test confirms or contradicts your guess from Step 1. Finally, do the results from your confidence interval in Step 4 support the results from your hypothesis test? (That is, how are confidence intervals and hypothesis tests related?) Explain your answers.

Step 6
What have you learned from this project and the semester as a whole? A broad topic, I know, but simply tell me what insights this whole process has given you into the field of statistics. This is your chance to sum up the semester! Also, feel free to include memorable adventures you had while gathering your sample. Be creative with this step, as I look forward to reading about how this project and course have helped you understand statistics! This step should be about one page in length, minimum.

Appendix G

SURVEY OF ATTITUDES TOWARD STATISTICS POST-TEST

DIRECTIONS: The statements below are designed to identify your attitudes about statistics. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

	Strongly disagree			Neither disagree nor agree			Strongly agree
I tried to complete all of my statistics assignments.	1	2	3	4	5	6	7
I worked hard in my statistics course.	1	2	3	4	5	6	7
I like statistics.	1	2	3	4	5	6	7
I feel insecure when I have to do statistics problems.	1	2	3	4	5	6	7
I have trouble understanding statistics because of how I think.	1	2	3	4	5	6	7
Statistics formulas are easy to understand.	1	2	3	4	5	6	7
Statistics is worthless.	1	2	3	4	5	6	7
Statistics is a complicated subject.	1	2	3	4	5	6	7
Statistics should be a required part of my professional training.	1	2	3	4	5	6	7
Statistical skills will make me more employable.	1	2	3	4	5	6	7
I have no idea of what's going on in this statistics course.	1	2	3	4	5	6	7

I am interested in being able to communicate statistical information to others.	1	2	3	4	5	6	7
Statistics is not useful to the typical professional.	1	2	3	4	5	6	7
I tried to study hard for every statistics test.	1	2	3	4	5	6	7
I get frustrated going over statistics tests in class.	1	2	3	4	5	6	7
Statistical thinking is not applicable in my life outside my job.	1	2	3	4	5	6	7
I use statistics in my everyday life	1	2	3	4	5	6	7
I am under stress during statistics class.	1	2	3	4	5	6	7
I enjoy taking statistics courses.	1	2	3	4	5	6	7
I am interested in using statistics.	1	2	3	4	5	6	7
Statistics conclusions are rarely presented in everyday life.	1	2	3	4	5	6	7
Statistics is a subject quickly learned by most people.	1	2	3	4	5	6	7
I am interested in understanding statistical information.	1	2	3	4	5	6	7
Learning statistics requires a great deal of discipline.	1	2	3	4	5	6	7
I will have no application for statistics in my profession.	1	2	3	4	5	6	7
I make a lot of math errors in statistics.	1	2	3	4	5	6	7
I tried to attend every statistics class session.	1	2	3	4	5	6	7
I am scared by statistics.	1	2	3	4	5	6	7

I am interested in learning statistics.	1	2	3	4	5	6	7
Statistics involves massive computations.	1	2	3	4	5	6	7
I can learn statistics.	1	2	3	4	5	6	7
I understand statistics equations.	1	2	3	4	5	6	7
Statistics is irrelevant in my life.	1	2	3	4	5	6	7
Statistics is highly technical.	1	2	3	4	5	6	7
I find it difficult to understand statistical concepts.	1	2	3	4	5	6	7
Most people have to learn a new way of thinking to do statistics.	1	2	3	4	5	6	7

NOTICE that the labels for the scale on each of the following items differ from those used above.

How good at mathematics are you?	Very poor 1	2	3	4	5	6	Very good 7
In the field in which you hope to be employed when you finish school, how much will you use statistics?	Not at all 1	2	3	4	5	6	Great deal 7
How confident are you that you have mastered introductory statistics material?	Not at all confident 1	2	3	4	5	6	Very confident 7
As you complete the remainder of your degree program, how much will you use statistics?	Not at all 1	2	3	4	5	6	Great deal 7

	Not at all likely						Very likely
If you could, how likely is it that you would choose to take another course in statistics?	1	2	3	4	5	6	7

	Very easy						Very difficult
How difficult for you is the material currently being covered in this course?	1	2	3	4	5	6	7

Appendix H

COURSE EVALUATION

Spring 2016
Section Number ________

1. (Answer in 2-3 sentences) The goal of this course in statistics was to teach me….

2. (Answer in 2-3 sentences) When people ask me what statistics is about, I will give them this response:

3. I think the most important thing in statistics is …

4. Use the following scale to answer the questions below. Put an X in each row that corresponds to your level of confidence about each of the items on the table.

   If you had to use statistics in the following ways, how confident are you that you could:

	Not confident at all	Below average confidence	Average confidence	Above average confidence	Very confident
Write a research question					
Design a good or scientific way to gather data					
Collect my own data					
Use the computer to make an appropriate chart or graph for a set of data					
Analyze my own data					
Write a report about data I analyzed for a research project					

5. What did you like most about statistics?

Appendix I

IRB APPROVAL LETTER

RESEARCH OFFICE

210 Hullihen Hall
University of Delaware
Newark, Delaware 19716-1551
Ph: 302/831-2136
Fax: 302/831-2828

DATE: January 26, 2016

TO: Sarah Baxter

FROM: University of Delaware IRB

STUDY TITLE: [849898-1] Comparing Assessment Methods in Undergraduate Statistics Courses

SUBMISSION TYPE: New Project

ACTION: APPROVED

APPROVAL DATE: January 26, 2016

EXPIRATION DATE: January 25, 2017

REVIEW TYPE: Expedited Review

REVIEW CATEGORY: Expedited review category # (7)

Thank you for your submission of New Project materials for this research study. The University of Delaware IRB has APPROVED your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

This submission has received Expedited Review based on the applicable federal regulation.

Please remember that informed consent is a process beginning with a description of the study and insurance of participant understanding followed by a signed consent form. Informed consent must continue throughout the study via a dialogue between the researcher and research participant. Federal regulations require each participant receive a copy of the signed consent document.

Please note that any revision to previously approved materials must be approved by this office prior to initiation. Please use the appropriate revision forms for this procedure.

All SERIOUS and UNEXPECTED adverse events must be reported to this office. Please use the appropriate adverse event forms for this procedure. All sponsor reporting requirements should also be followed.

Please report all NON-COMPLIANCE issues or COMPLAINTS regarding this study to this office.

Please note that all research records must be retained for a minimum of three years.

Based on the risks, this project requires Continuing Review by this office on an annual basis. Please use the appropriate renewal forms for this procedure.

If you have any questions, please contact Nicole Farnese-McFarlane at (302) 831-1119 or nicolefm@udel.edu. Please include your study title and reference number in all correspondence with this office.