The Effect of Conceptualized Reading Assignments on Student Attitudes in Statistics

Matt Misterka

Committee:
Karsten Maurer (Advisor)
Lynette Hudiburgh
Michael O'Connell

Abstract

This case study investigates the effect of conceptualized reading assignments on student attitudes in an undergraduate introduction to modeling course (STA 363) and an upper-level design of experiments course (STA 466/566). Student attitudes in STA 363 were measured using a before-and-after Survey of Attitudes Toward Statistics (SATS). A new eight-item before-and-after survey was created to measure student attitudes in the STA 466/566 course. Principal component analysis and Cronbach's alpha were used to evaluate the internal consistency of the survey items. Hotelling's T^2 was used to compare the change in attitudes between treatment and control sections of each class. There is no difference in attitude scores for either class. This study gives valuable insight into the overall attitude standing of each course.

1. Introduction

Introductory statistics students have differing degrees of confidence when they enter the classroom for the first time. Some students are comfortable with their quantitative skills while others are petrified of the fraction bar. Many students think to themselves, "I didn't do well in math, so I won't do well in statistics." According to Iddo Gal, Lynda Ginsburg, and Candace Schau, these beliefs can affect the learning process, statistical behavior outside of the classroom, and the decision to enroll in future statistics classes. These authors go on to make a distinction between student emotions, beliefs, and attitudes. Emotions are fleeting responses to immediate experiences whereas beliefs are one's self-efficacy towards learning the course material. Lastly, attitudes are somewhat stable and they are developed through repeated emotional responses (Gal et al., 1997). Many people have investigated the impact of student attitudes on course achievement. Carmen Sorge built a structural model to investigate the impact of attitudes on course achievement for engineering students in statistics (Sorge 2001). Course achievement was measured using results from quizzes, midterms, and the final exam. Sorge found that achievement is related to the latent variables: previous success, difficulty, cognitive competence, and affect. These variables were measured using the 28-item version of the Survey of Attitudes Toward Statistics (SATS). It is important to note that there is a lot of variability in the measurement of these variables and between samples. In 2012, Emmioğlu and Capa-Aydin conducted a meta-analysis of 17 studies relating achievement to SATS student attitudes (Emmioğlu , E. & Capa-Aydin, Y. 2012). The authors used correlation as a measure of effect size. They found that the correlations between achievement and attitudes were fairly heterogeneous. However, the overall correlation coefficient between achievement and the attitude subscales was twice as large for U.S. studies compared to studies conducted in other countries. Below is part of a table taken from the study above. There were eight U.S. studies with sample sizes ranging from 91 to 342. There were nine non-U.S. studies with sample sizes ranging from 49 to 2031. These correlations are calculated using Fisher's r to Z transformation. This fixed-effects method is discussed in (Hedges and Vevea, 1998). According to Andy Field, this transformation leads to a slight bias (overestimation) of the coefficients (Field, 2001).

Column1	Overall	U.S.	non-U.S.
Affect	0.3	0.39	0.18
Cognitive	0.3	0.39	0.19
Value	0.21	0.27	0.14
Difficulty	0.2	0.26	0.13

Our question is, how can we positively change the student attitude to enhance this learning experience?

As mentioned previously, students often make a connection between math and statistics. There are preconceived notions that statistics includes an excessive amount of computation. However, we know that practicing statisticians are well-versed in a diverse set of skills, including data analysis, interpretation, and communication. The 2016 Guidelines for Assessment and Instruction in Statistics Education (GAISE) recommends that students focus on conceptual understanding (GAISE, 2016). By focusing on concepts, we allow the student to build a foundation necessary for proper implementation of statistical methods. Kendra Schmid discussed the impact that a conceptual article critique had on an introductory biostatistics class (Schmid, 2013). This assignment required the students to find an article from an academic journal that is closely related to their field of study. The students summarized statistical methods, made connections to in-class material, discussed their opinions of the methods, and finally wrote about how statistics is used in their career field. Schmid had a number of takeaways from this study. First, students were surprised at the number of articles related to their interests. Second, some articles helped reinforce classroom material while other articles introduced new statistical methods. Last, students understood the importance of the statistics in their field even if they did not actively use the methods themselves. Their study suggests that contextual assignments can increase student interest in the subject matter. We will implement a similar assignment and evaluate its impact on student attitudes.

This project introduces a set of reading assignments into two statistics classes at Miami University. Miami University is a moderate-sized, liberal arts university in Oxford, Ohio. The goals of the project are to help students learn the course material and positively change their attitudes about statistics. With readings that emphasize conceptual understanding in applied settings, we hope the assignments help students develop an appreciation for the role of statistics in their lives.

2. Literature Review

With new advancements in technology and developments in pedagogy, statistics educators are finding innovative ways to present the course material to students. For instance, instructors are increasingly introducing real datasets into the classroom, avoiding the artificial data used for convenience (GAISE, 2016). This allows the student to get hands-on guidance with exploration and data analysis in an authentic context. Perhaps these datasets are carefully chosen so that the content appeals to the class as a whole, but this is no guarantee.

A study investigated and evaluated ways to improve critical thinking skills for students in an introductory statistics class. Sherri Cheng, Mark Ferris, and Jessica Perolio introduced writings and small-group discussions into the classroom (Cheng et al., 2018). The assignments were coined "Statistics as Evidence" and required students to explore media from sources such as *The New York Times, Science*, TED Talks, and Gallup Polling. For instance, the first assignment explored the topic of gun ownership through the *Information is Beautiful* website (https://informationisbeautiful.net/). Students were asked to evaluate the statistical evidence while setting aside their personal opinions. The small-group discussion showed that students had a difficult time staying neutral in light of the data. The instructor measured critical thinking and student engagement with the assignments using two assessments: a before-and-after writing assignment and a self-reflective survey. The writing assignment was evaluated by writing professionals and it measured students' ability to explain issues, provide evidence, understand context, and provide a conclusion. Each of these four components were scored on a scale from 1 (low) to 4 (high). The self-reflective survey measured students' interest, motivation, curiosity, and ability to apply abstract thoughts to different situations. It was measured on a scale from 1 (strongly disagree) to 7 (strongly agree). A paired t-test showed a significant improvement in all dimensions of the critical thinking assessment. The self-reflective survey showed the class had an overall positive experience with averages of each component ranging from 5.83 to 6.05.

This idea of a reflective writing assignment is not a new one. Norean Radke-Sharpe argues that writing skills are essential in statistics education (Sharpe, 1991). It helps the student internalize the material, it encourages creativity, and it improves communication skills. Internalizing the material allows the student to "focus on the process, rather than the product of statistical approaches." Sharpe suggests a few ways to incorporate writing into the classroom. These include: requirements of formal written summaries for statistical procedures, reviews of journal articles, and written responses on exams. Andrew Gelman and his colleagues included the review of journal articles and newspapers as a project for an introductory statistics class (Gelman et al., 1998). The project required students to read a newspaper or journal article describing a statistical finding as well as the original report on which the article is based. Then, students summarized various components of the article and answered questions related to the reading. Gelman mentions that this project made students utilize their statistical knowledge as well as critical skills to understand the limitations of scientific studies. Students were also exposed to different ways of presenting statistical findings.

Jennifer Green and Erin Blankenship used low-stakes writing assignments to develop conceptual understanding of hypothesis tests in a mathematical statistics class (Green and Blankenship, 2015). Students selected one of four *Significance* articles to read for a class discussion. These articles were chosen not only because they demonstrated the statistical concept, but because they incorporated student interests. The authors used guided homework questions to help the students focus on important concepts. The students were graded on the thoughtfulness of the answers, as opposed to the accuracy. The discussions required groups of students to summarize each article in front of the class. In this way, students had an opportunity to learn about the articles they chose not to read.

Over the years, many instruments have been created to measure student attitudes in statistics. One of the most well-known surveys is the Student Attitudes Towards Statistics (SATS). The first version of the SATS was developed and validated by Candace Shau, Joseph Stevens, Thomas Dauphinee, and Ann Del Vecchio in 1992 (Schau et al., 1995). Initially, the survey included 28 items comprising four facets: affect, cognitive competence, value, and difficulty. The current version of the SATS includes 36 items with two additional facets: interest and effort. The survey is measured using a Likert scale. Possible answers range from 1 (strongly disagree) to 7 (strongly agree). Negatively-worded items are reverse-coded so that high scores represent positive attitudes. The score of an attitude component is measured as the average of all items in the set. In this way, the scores can be considered on a continuous scale. According to Candace Shau, an effective statistics attitude survey taps into the most important dimensions, applies to many introductory classes, and incorporates a balance of positive and negative items. Not only this, the survey should be developed with student feedback in mind and validated with confirmatory analysis techniques. In 1992, the SATS was the only survey to meet all of these criteria. The SATS was evaluated from a sample of 1,203 completed surveys. As an additional measure of validation, Schau calculated the correlation coefficients between the SATS and the Attitudes Toward Statistics scales (ATS). The ATS was created in 1985 by Steven Wise (Wise, 1985). The correlations were calculated from a sample of 230 completed ATS surveys. Cronbach's alpha values were calculated for the SATS survey as well. See appendix 6.1 for these values.The six component version of the SATS was validated in 2007 by Dirk Tempelaar (Tempelaar et. al, 2007). This study supports SATS use in academia.

In 2010, Candace Schau and Ann Michele Millar discussed the good and bad ways to analyze the SATS survey (Millar and Shau, 2010). T-tests are robust against normality violations, but independence between students is assumed. This is not plausible because students share common experiences within the same class. These experiences will not impact them in the same way, but their scores on the SATS will be correlated. On the other hand, linear models can account for dependence among factors (fixed or random effects). Lisa Carnell illustrates the use of t-tests for SATS-36 data (Carnell, 2008). Carnell introduced a group project into one section of an introductory statistics class while the other section did not complete a group project. Carnell ran independent samples t-tests for the equality of mean differences (post - pre) between the classes. These results were not significant. Carnell also ran separate independent samples t-tests on each of the subscales to see if the mean differences (post - pre) was different from zero. She found that the project group had significant decreases in interest and effort while the control group had a significant decrease in interest. In this case, the findings are contrary to what the investigator would expect to see. However, these results should be taken with caution because the investigator does not account for the dependence between students or multiple comparisons.

In the same study described above, Schau and Millar used linear models to analyse 200 sets of SATS component scores. One such model they investigated was the gain score of affect as a function of the pretest score while controlling for lecture section. The authors show that a student with a pretest score of 2 will likely have an increase in gain score while a student with a pretest score of 6 will likely have a decrease in gain score. The models are presented below with associated confidence intervals. These results show that a student with a pretest score of 2 still holds a negative attitude while a student with a pretest score of 6 still holds a positive attitude. This example illustrates one of the drawbacks to the SATS. It's not known whether the negative slope is due to the true dependence of gain score on the pretest score or regression to the mean.

Mean Gain = 1.8 – 0.40*Pretest , lecture as fixed effects
Mean Gain = 1.8 – 0.39*Pretest , lecture as random effects

	95% CI fixed effects	95% CI random effects
Pretest = 2	(+0.68, +1.33)	(+0.60, +1.42)
Pretest = 6	(-0.90, -0.31)	(-0.96, -0.18)

The SATS survey has been used in many statistics classes and analyzed with a variety of methods. Keith Carlson and Jennifer Winquist evaluated the effect of a student workbook in the classroom on student attitudes (Carlson and Winquist, 2011). Students completed short, conceptual readings and associated questions before coming to class. During the class, students were encouraged to work in groups to complete a set of questions related to the reading. Examples include: the distribution of sample means and the logic of hypothesis testing. Carlson and Winquist used the Wilcoxon signed rank test to determine if attitudes changed throughout the semester. Students had significantly higher affect and cognitive competence at the end of the course compared to the beginning. Additionally, students' effort was significantly lower than at the beginning. This finding is contrary to what the investigator would expect. Lastly, students thought the class was more difficult at the end of the course than at the beginning. A Mann Whitney U test was used to compare these students' attitudes to a group of students without the workbook curriculum. The data for this control group was supplied by Candace Schau. Compared to the control group the workbook curriculum had greater positive changes in affect and cognitive competence. On the other hand, the treatment group thought the class was more difficult by the end of the semester. Carlson and Winquist also found that these positive changes correlated positively with final exam performance and GPA.

In a randomized experiment, Lawrence Lesser, Dennis Pearl, and John Weber III investigated the effect of student fun items on learning outcomes, student anxiety, and student attitudes (Lesser et al, 2016). Lesser hypothesized that student attitudes would not show significant change because attitudes are a "trait" rather than a "state". Lesser's experiment introduced small readings into a learning management system (an electronic learning environment) with an accompanying song, cartoon, or joke to help reinforce the material. The treatment (fun item included) and control (fun item omitted) were randomly assigned to half of the students in each section. The instructors informed the students that there would be an associated multiple choice question on the exam about the reading to incentivize completion. The authors argue that this is an effective way to evaluate fun items because it is easy to insert the song in an e-book or video while keeping all other conditions unchanged. Results showed that the multiple choice questions corresponding to a song were answered correctly 50% of the time, on average, compared to the control groups' 42.3%. The cartoon and joke items did not show a strong effect. The sample size in this study was not large enough to provide adequate power for detecting a difference in anxiety or attitude subscales. For more examples of SATS research see Candace Shau's website.

Many studies have used the SATS. We briefly describe two more examples here. Beth Chance, Jimmy Wong, and Nathan Tintle incorporated the SATS survey into a study of a simulation-based curriculum (Chance et al., 2016). Thomas Devaney investigated the change in graduate student attitudes between online and traditional classrooms (DaVaney, 2010). For more examples of SATS research see Candace Shau's website.

3. Methods

This study involved two statistics classes at Miami University, STA 363 and STA 466/566. In this study, we introduced reading assignments into these classes and evaluated the change in student attitudes between treatment and control sections. In STA 363, student attitudes were measured using the SATS-36. We developed a new survey to measure student attitudes in STA 466/566.

3.1 Course Descriptions

STA 363 is an undergraduate statistical modeling course. This class can be thought of as the introduction to the statistics major. It covers several topics in regression, design, and the implementation in R. STA 466/566 is a design of experiments class, composed of undergraduate and graduate students. Each of these courses has two sections. Section A of STA 363 had 52 students: 8 students were admitted to Miami in 2015, 9 admitted in 2016, 19 admitted in 2017, and 16 admitted in 2018. Seventeen of the 52 students in section A are enrolled in either statistics or the math/stat program. Seventeen students are enrolled in computer science or software engineering. Eight students are enrolled in biology-type major. The remaining students are enrolled in a variety of majors from kinesiology, earth science, and economics among others. Nineteen students in section A are enrolled in a double major. These majors are distributed across neuroscience, premed, analytics, statistics, math education, entrepreneurship, computer science, leadership, and sustainability.

Section B of STA 363 has 49 students: 2 students were admitted to Miami in 2014, 3 admitted in 2015, 12 admitted in 2016, 25 admitted in 2017, and 7 admitted in 2018. Ten of the 49 students are enrolled in either statistics or math programs. Eighteen students are enrolled in computer science or software engineering. Five students are enrolled in a biology-type major. The remaining students are enrolled in a variety of majors from the social sciences, engineering, and language studies among others. There is one non-matriculated student. Eleven students in section B are enrolled in a double major. These majors are distributed across neuroscience, premed, education, interactive media studies, marketing, math, microbiology, and software engineering.

Section A of the STA 466/566 class includes 30 students. Seventeen of these students are undergraduates. The graduate students are all in the statistics master's program. Fourteen of the 17 undergrad students are enrolled in either the statistics or the math/stat program. The other three undergrad students are enrolled in economics, business economics, and chemical engineering. Section B of STA 466 is composed of 27 undergraduates. Twenty of the 27 are enrolled in the statistics, math/stat, or mathematics programs. Three of the remaining seven students are enrolled in economics. The other four students are enrolled in psychology, linguistics, marketing, and international studies.

3.2 Reading Assignments

The majority of the reading assignments were selected from various issues of the *Significance* magazine. *Significance* is a general audience magazine of the Royal Statistical Society and the American Statistical Association. It publishes articles about all aspects of statistics including history, contemporary issues, statistical ethics, and practical application among many other topics. We chose this magazine for the readings because the content was more conceptual in nature and less computational. This allows the students to read through the articles without having to stop and think about how the author moved from one line to the next. Additionally, the students can focus on where statistics is applied in everyday life.

Not only do the two courses cover different topics, but they are taught to students of different experience levels. Thus, it was important to consider the content as well as the purpose of the readings within each class. We needed to help students meet the course requirements while engaging them in a new way. The readings for the STA 363 focus on fundamental concepts: the development of modern statistics, understanding p-values, exploring research questions, and understanding the consequences of biased data. The STA 466/566 students have already developed a strong foundation in statistics, so these readings focused more narrowly on design principles as well as ethical practices in statistics. Ethical practice is essential for any statistician because misinterpreted results can have devastating consequences in the broader scientific community. It is also important to handle data responsibly and take care to document decisions made so that results are reproducible.

The STA 363 class completed four readings. The assignments required them to complete up to five questions about the articles. These questions were written to help guide the students through the reading and assist in their comprehension of the material. Below is a list of the readings. See appendix 6.5 for the individual questions.

- “How One Woman Used Regression to Influence the Salary of Many” by Amanda Golbeck. This article is about how Elizabeth Scott used regression to investigate the difference between the salaries of men and women in academia. The article discusses the steps taken to answer a research question including data cleaning, variable selection, and model building.
- “Where the Seeds of Modern Statistics Were Sown” by Robert Langkjaer-Bain. This article explores the agricultural studies at Rothamsted by Ronald Fisher and Frank Yates. Fundamental ideas in experimental design such as ANOVA, one-way completely randomized designs, and factorial designs are introduced in this reading.
- “The ASA’s P-Value Statement One Year On” by Robert Matthews. This article was written in 2017, one year after the ASA released a statement about the improper use and interpretation of p-values.
- “To Predict and Serve” by Kristian Lum and William Isaac. This article introduces the use of predictive modeling in the police department. The reading uncovers the dangers of fitting models to biased data.

The STA 466/566 class completed three readings with a similar structure described above. They completed reading assignments with two of the same articles as listed above -- agricultural studies at Rothamsted and the p-value statement -- along with one other:

- “Video Game Violence and Aggression: A Proven Connection?” by Joseph Hilgard. This article discusses why it is impossible to establish a causal connection between video games and aggression in children.

3.3 Design and Data Collection

Both sections of each course completed the reading assignments; however, within each course one section completed the assignments in the first half of the semester while the other section completed the assignments in the second half of the semester. We used a coin flip to randomly determine which section of the class would complete the assignments first. Section A of STA 363 was randomized to the treatment. Section B was the control. Section A of STA 466/566 was randomized to the treatment. Section B was the control. A start-of-semester and mid-semester survey was administered at the same time for each section to evaluate student attitudes. In this way, one section acted as the treatment group (readings assigned between surveys) whereas the other section acted as the control group (no readings assigned between surveys). The instructors administered the surveys during class time for 5-10 minutes. The names of the students were recorded on each survey for matching purposes. The 363 class completed the first survey in week two and the last survey in week 11. The four readings were assigned to the treatment class in this timeframe. Similarly, the 466 class completed the first survey in week four and the last survey in week 11. The three readings were assigned to the treatment class in this timeframe. This design allows us to collect all of the data for the study after the mid-semester survey. We will compare the survey responses between the treatment group and the control group.

In the STA 363 course, 36 students in the treatment group and 28 students in the control group provided consent to be included in the analysis. In the STA 466/566 course, 29 students from the treatment group and 9 students from the control group provided consent to be included in the analysis.

3.4 Survey Details

The SATS-36 is comprised of 6 components: affect, cognitive competence, value, difficulty, interest, and effort. Candace Schau defines each of these components as follows:

- Affect - students' feelings concerning statistics (6 items).
- Cognitive competence - students attitudes about their intellectual knowledge and skills when applied to statistics (6 items).
- Value – students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life (9 items).
- Difficulty – students' attitudes about the difficulty of statistics as a subject (7 items).

- Interest – students' level of individual interest in statistics (4 items).
- Effort - amount of work the student expends to learn statistics (4 items).

We designed and implemented a new 8-item survey in the STA 466/566 course because the SATS survey is intended for introductory statistics students and is not as applicable to students at higher levels. Our survey is measured using a Likert scale from 1 (strongly disagree) to 7 (strongly agree). Attitude scores are calculated as an average of all items in each component. Our survey included the following components:

- Value - similar to the SATS value component described above (4 items).
- Statistical Ethics - student awareness of ethical data practices in the statistics field (4 items).

The survey items for both courses can be found in Appendix 6.1 and 6.4

3.5. Survey Validation

Brian Everitt and Graham Dunn describe principal component analysis (PCA) as a method to "describe the variation of a set of multivariate data in terms of a set of uncorrelated variables, each of which is a particular linear combination of the original variables" (Everitt and Dunn, 2001). PCA was introduced by Karl Pearson in 1901 (Pearson, 1901). We use principal components to investigate the factor loadings of the survey items. PCA is described below.

Let X be an n x p matrix of independent variables. The jth principal component:

$$Z_j = a_j^T \mathrm{X} = a_{j1}\, x_1 + a_{j2}\, x_2 + \ldots + a_{jp}\, x_p$$

Z_j has the largest variance subject to $a_j' \, a_j = 1$ and $a_j' \, a_i = 0$ (i < j)

Let S be the p x p sample covariance matrix. Then Var(Z_j) = $a_j' \,\mathrm{S}\, a_j$ Maximizing this variance uses the method of Lagrange multipliers to find the eigenvector of S. First, find λ_j such that: det(S - λ_j I) = 0. Then, find the unit eigenvector, e_j , that satisfies S e_j = $\lambda_j e_j$. The coefficients for the jth principal component are $a_j = e_j$. Since $a_j' \, a_j = 1$, the variance of the jth component is λ_j .

The total variance of the p components is $\sum_{i=1}^{p} \lambda_i$ = trace(S). The p x p matrix of eigenvectors Q is orthogonal. Thus, the p principal components are uncorrelated. Since S is a symmetric, positive definite matrix it can be decomposed as S = $Q \Lambda Q^T$ where Λ is a p x p diagonal matrix of the eigenvalues of S and Q is as defined above. Everitt and Dunn discuss a rescaling of the coefficients $a_j^* = \sqrt{\lambda_j}\, a_j$ such that $a_j' a_j = \lambda_j$. These coefficients are the factor loadings which give correlations between the components and the independent variables (Everitt and Dunn, 2001).

We will use principal component analysis to evaluate the dimensionality and internal consistency of a given item set. We will investigate scree plots of the proportion of variability explained by each principal component as well as the factor loadings. There is a case for unidimensionality and internal consistency if all of the items in a set load strongly onto the first principal component. We're also looking for the first principal component to explain a large proportion of variability.

Cronbach's alpha can be used as a measurement for reliability. Coefficient alpha was created in 1951 by Lee Cronbach (Cronbach,1951). We can use Cronbach's alpha to measure the internal consistency of a given set of survey items. Alpha takes a value between 0 and 1 with higher values indicating greater internal consistency. For example, a value of 0.7 for the interest items of the SATS tells us that the items measure the construct of interest fairly well. Cronbach's alpha is given below. K is the number of items, σ_X^2 is the variance of observed total test scores, $\sigma_{Y_i}^2$ is the observed variance of component i. $X = Y_1 + Y_2 + \ldots + Y_k$:

$$\alpha = \frac{K}{K-1}\left(1 - \frac{\sum_{i=1}^{K} \sigma_{Y_i}^2}{\sigma_X^2}\right)$$

Taken from (University of Virginia, 2019).

3.6 Data Analysis

Hotelling's T^2 statistic is a multivariate t-test to test equality of mean vectors:

Ho: $\mu_1 = \mu_2$
Ha: $\mu_1 \neq \mu_2$

Suppose we have two groups with sample mean vectors $\overline{x_1}$ and $\overline{x_2}$ for groups 1 and 2.

Let S be the pooled within-group sample covariance matrix (Everitt and Graham, 2001).

That is,
S = $[(n_1 - 1)S_1 + (n_2 - 1)S_2]/(n_1 + n_2 - 2)$

Then,
$T^2 = [(n_1 * n_2)/(n_1 + n_2)]\ (\overline{x_1} - \overline{x_2})^T\ S^{-1}\ (\overline{x_1} - \overline{x_2})$.

Of note, $(n_1 + n_2 - p - 1)/[(n_1 + n_2 - 2)p]\ T^2$ is distributed F with p and $(n_1 + n_2 - p - 1)$ degrees of freedom under the null hypothesis of no group difference. For our analysis, the student responses within a section are very likely to be correlated with one another. Given the violation of independence, it is likely that the type 1 error rate (rejecting the null hypothesis when the null hypothesis is true) is inflated. If our results are significant, we can run a simulation to investigate the inflation of type 1 error rates under different covariance structures. If the results are not significant under this assumption violation, it suggests they will not be significant with a different covariance structure.

The two groups are the treatment and control sections of a class. The SATS survey has 6 components and each section $i \in \{1, 2\}$ will have the mean difference (post - pre) vector as follows

$$\overline{x_i} = \left[\overline{Affect\ Diff}\quad \overline{Cognitive\ Diff}\quad \overline{Value\ Diff}\quad \overline{Difficulty\ Diff}\quad \overline{Interest\ Diff}\quad \overline{Effort\ Diff}\right]^T.$$

These difference vectors for treatment and control sections will then be compared with the Hotelling’s T as defined.

4. Results

4.1 STA 363

Pre Survey Cronbach’s and PCA

Principal component analysis and Cronbach’s alpha measures were calculated separately for the start-of-semester surveys and mid-semester surveys. We included the responses from both the treatment and control sections for this part of the analysis because we want to measure whether the students are responding similarly across all items within a set. Ideally, a good survey will be internally consistent no matter what group (treatment or control) the students are in. For example, perhaps one group of students answer positively to a set of items while another group of students answer negatively to the same set of items. Thus, we can consider treatment and control sections responses together when evaluating the internal consistency of survey items. The Cronbach alpha measures are shown in figure 1. The 95% bootstrapped confidence intervals are as follows:

Affect: (0.73, 0.88),
Cognitive: (0.73, 0.88)
Difficulty: (0.41, 0.73)
Effort: (0.52,0.75)
Interest: (0.87,0.94)
Value: (0.84, 0.93)

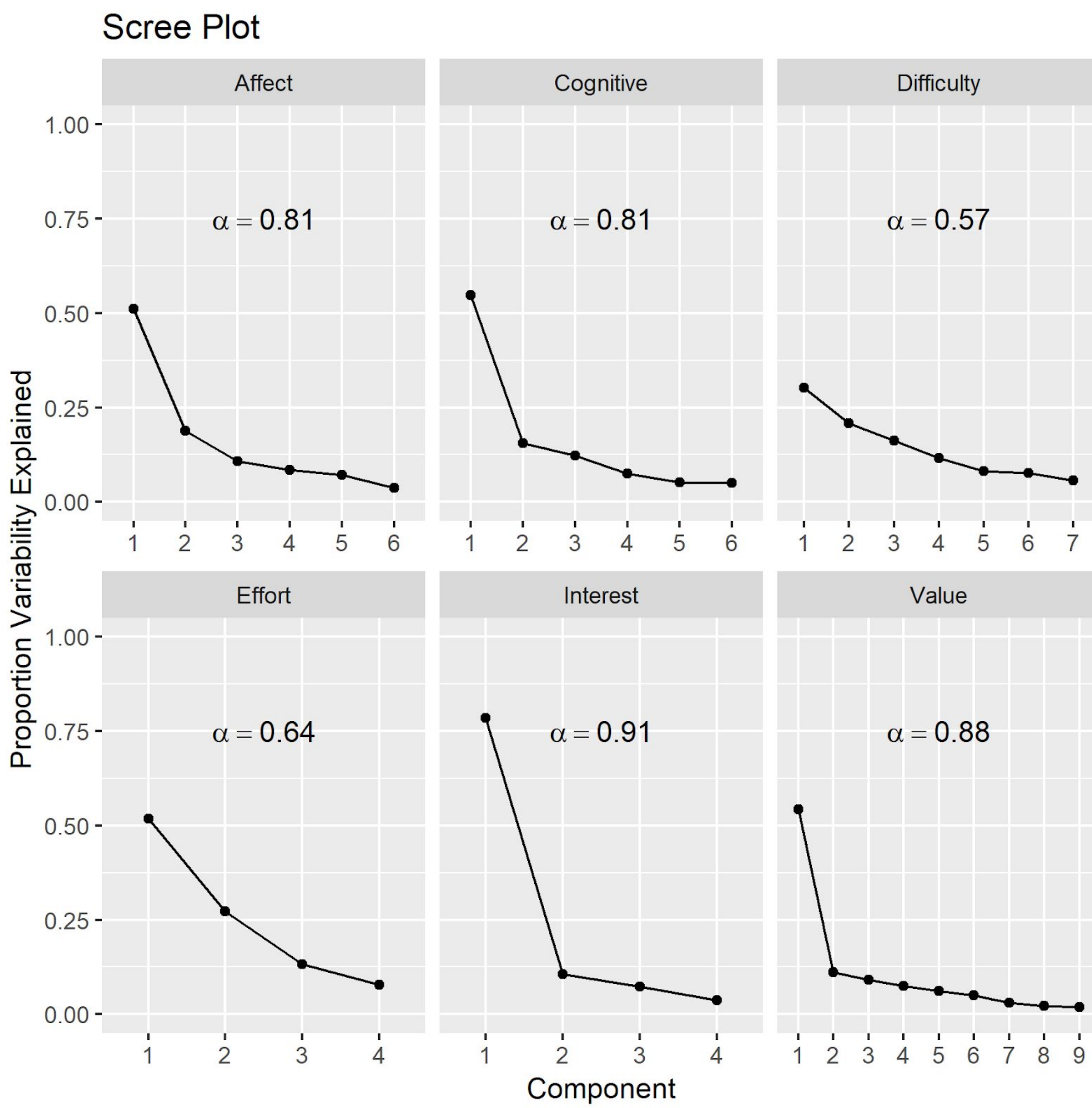

Figure 1: Scree plots of the six attitude components for the first survey

The difficulty and effort components are less internally consistent than the rest. This suggests that students did not respond similarly across all items within difficulty and effort. Below are the pca loadings for the difficulty and effort components. See appendix 6.2 for the other loadings. The positively-worded items do not load as strongly onto the first principal component as the negatively-worded items. The second principal component for difficulty may distinguish between how students learn statistics and the difficulty of individual statistical concepts. The second principal component for effort may distinguish between the effort needed to pass the class and the effort needed to earn a good grade in the class. We can investigate the effort component further by considering the correlation matrix given below in table 3. We see that item 1 and item 27 are moderately correlated (r = 0.44). Similarly, item 2 and item 14 are moderately correlated (r = 0.66). This gives more evidence for the interpretation above.

Difficulty:

Item	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
s6	0.17	0.567	0.24	0.585	0.07	0.471	0.152
s8N	0.309	0.178	0.698	-0.024	-0.014	-0.601	-0.152
s22	0.176	0.6	-0.194	-0.577	-0.046	0.181	-0.452
s24N	0.537	-0.121	0.099	-0.377	-0.349	0.223	0.612
s30N	0.384	-0.481	0.249	-0.048	0.451	0.437	-0.403
s34N	0.443	0.137	-0.473	0.086	0.605	-0.359	0.243
s36N	0.461	-0.149	-0.35	0.416	-0.55	-0.118	-0.392

Table 1: Loadings for the difficulty items

Note on notation: s8N = “SATS item 8, negative”

Effort:

Item	Comp.1	Comp.2	Comp.3	Comp.4
s1	0.508	0.374	0.766	0.126
s2	0.577	-0.316	-0.339	0.672
s14	0.533	-0.492	0	-0.688
s27	0.353	0.72	-0.547	-0.241

Table 2: Loadings for the effort items

Correlation	s1	s2	s14	s27
s1	1	0.36682	0.331943	0.435428
s2	0.36682	1	0.661987	0.221073
s14	0.331943	0.661987	1	0.05496
s27	0.435428	0.221073	0.05496	1

Table 3: Correlation matrix of effort items

Post Survey Cronbach's and PCA

The Cronbach's alpha measures for the post survey are given in figure 2 below. The 95% bootstrapped confidence intervals are as follows:

Affect: (0.71, 0.87),
Cognitive: (0.72, 0.87)
Difficulty: (0.64, 0.83)
Effort: (0.53,0.81)
Interest: (0.91,0.96)
Value: (0.88, 0.94)

Figure 2 and table 4 give the scree plots as well as pca loadings. See appendix 6.3 for the other loadings. Similar to the pre survey, we see lower alpha values for difficulty and effort. The second component for effort mimics the findings from the pre-survey. The associated effort items are slightly more correlated.

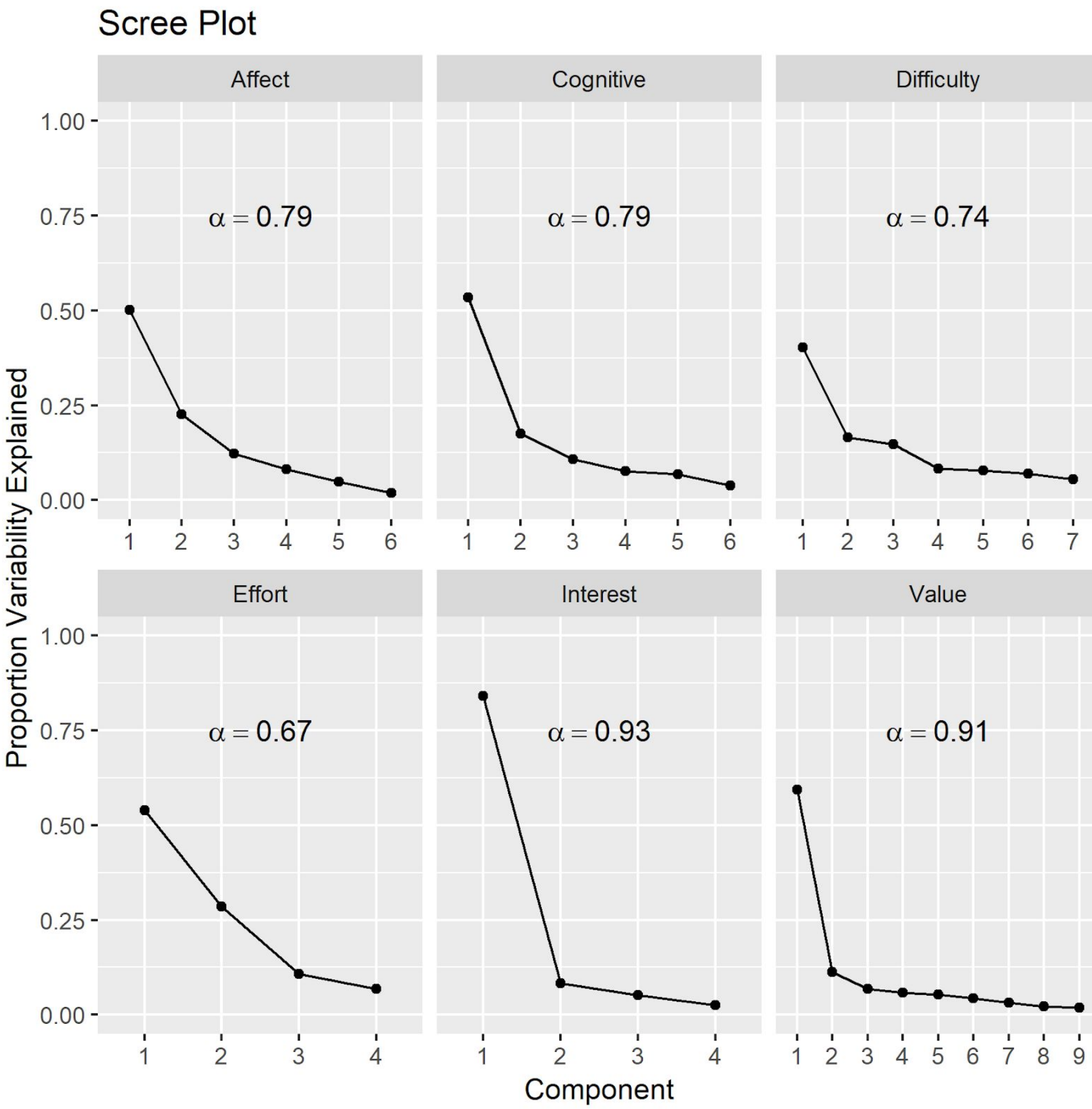

Figure 2: Scree plots of the six attitude components

Difficulty:

Item	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6	Comp.7
s6	0.259	0.426	0.642	0.363	0.004	0.442	0.109
s8N	0.404	0.281	0.223	-0.679	-0.414	-0.273	0.013
s22	0.243	-0.673	0.433	-0.067	0.065	0.064	-0.536
s24N	0.44	-0.384	-0.087	-0.18	0.299	0.211	0.696
s30N	0.388	0.004	-0.539	0.096	-0.432	0.542	-0.265
s34N	0.449	-0.088	-0.054	0.596	-0.224	-0.611	0.096
s36N	0.406	0.362	-0.223	-0.071	0.706	-0.12	-0.37

Table 4: Loadings of difficulty items

Effort:

Item	Comp.1	Comp.2	Comp.3	Comp.4
s1	0.563	0.385	0.109	0.723
s2	0.537	-0.355	0.688	-0.333
s14	0.314	-0.757	-0.521	0.237
s27	0.544	0.391	-0.492	-0.556

Table 5: Loadings of effort items

Correlation	s1	s2	s14	s27
s1	1	0.462091	0.070918	0.698081
s2	0.462091	1	0.496262	0.376896
s14	0.070918	0.496262	1	0.103954
s27	0.698081	0.376896	0.103954	1

Table 6: Correlation matrix of effort items

Hotelling and Class Comparison

We ran the Hotelling's T^2 test to see if there is a difference in at least one of six component scores (post - pre) between the two sections of STA 363. Since the p-value of 0.8218 is greater than any reasonable level of significance, we fail to reject the null hypothesis. We don't have sufficient evidence to suggest that there is a difference in at least one of the attitude components between treatment and control. The test statistic was 0.47782 on 6 and 53 degrees of freedom. Four students were removed from the data because they either did not fill out a post survey or did not complete all of the items.

Figure 3 is a comparison of the two classes between pre and post survey. These are faceted line plots for each of the six components on the SATS. The composite score of a student is rescaled from 0 to 1 because not every component has the same number of items. The the blue lines (~36 students) represent the change in composite score for a single student in the reading section (pre survey to post survey) while the red lines (~28 students) represent the change in composite score for a single student in the control class (pre survey to post survey). While there is no difference between sections, it is interesting to see the individual trajectories of these scores. It appears that both sections thought the class would be somewhat difficult. Overall, the affect and cognitive scores appear to increase across both sections. The effort scores tell us that students planned to work hard, study hard, complete assignments, and attend class. These scores didn't change that much. On the other hand, there is a lot of variability in the scores for interest and value.

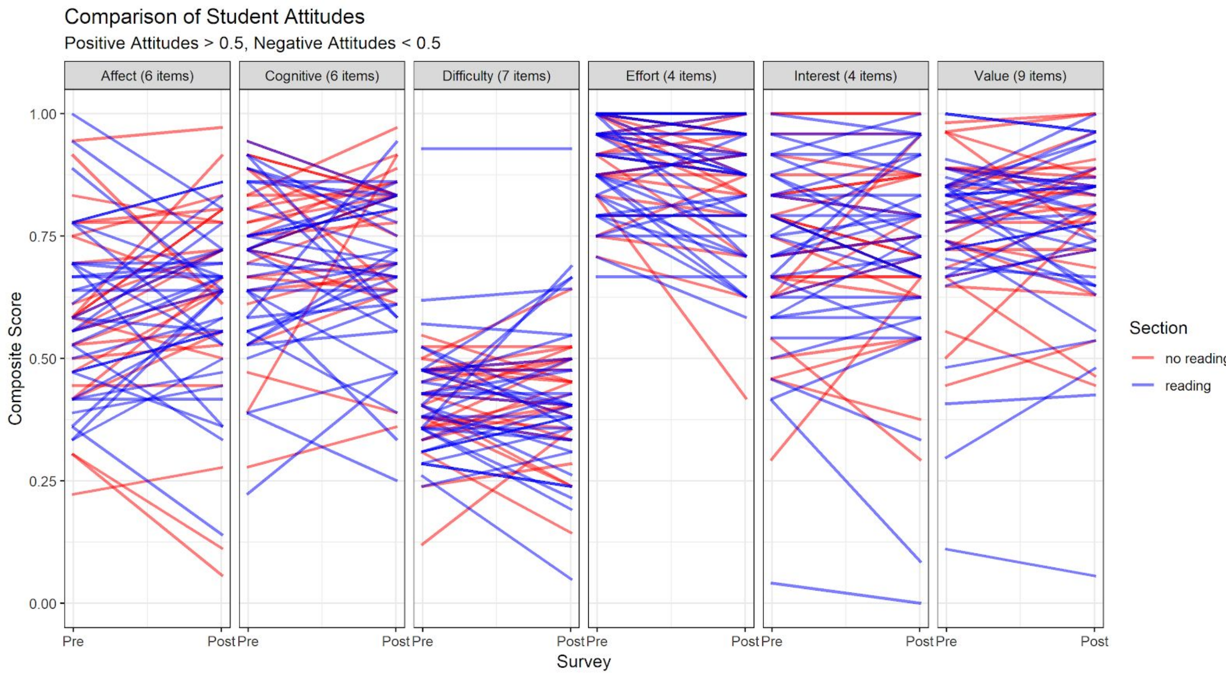

Figure 3: Comparison of attitudes between treatment and control groups

4.2 STA 466/566

Pre Survey Cronbach's and PCA

Principal component analysis and Cronbach's alpha measures were calculated separately for the start-of-semester surveys and mid-semester surveys. The Cronbach alpha measures are given in figure 4. The 95% bootstrapped confidence intervals are as follows:

Ethics: (0.72, 0.91)
Value: (0.43, 0.84)

Note that when the item seven ("I think that statistics is not useful in many areas of society.") is removed the alpha measure increases to 0.85. This suggests that students answered item seven differently than the other items. Item seven does not load onto the first component giving more evidence that this item is less internally consistent than the other items. The other three items relate to the students' personal feelings about the value of statistics in their life whereas this item speaks of statistics on the societal level. It’s plausible that students think statistics is useful, but not in “many” areas of society. Below are the pca loadings and scree plots.

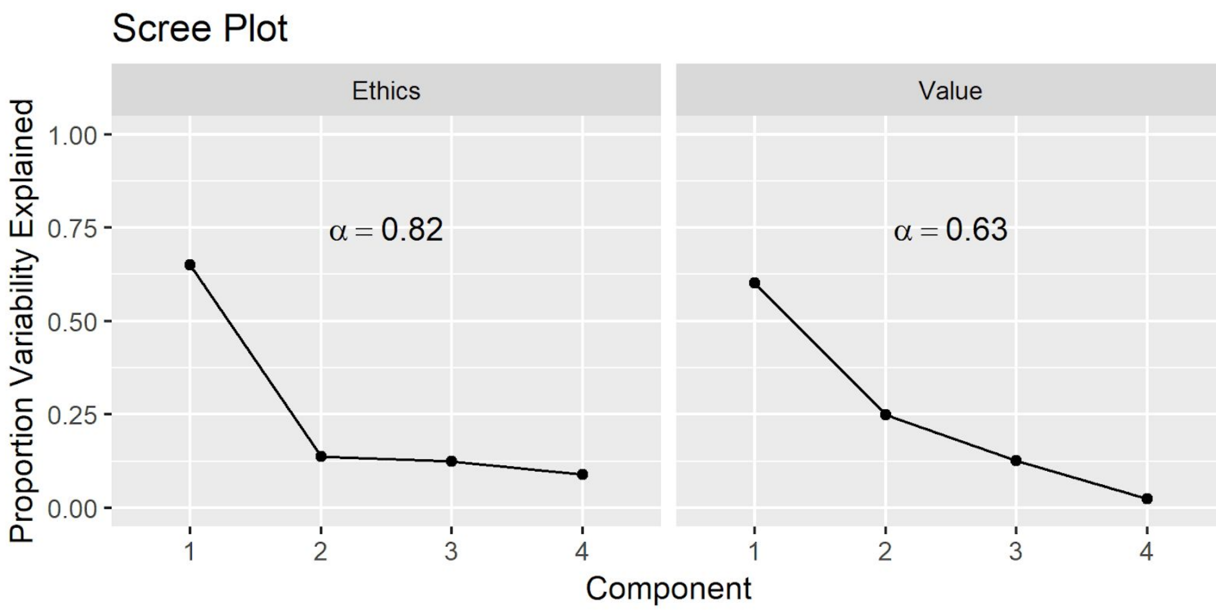

Figure 4: Scree plots of the two components in the STA 466/566 survey

Ethics:

Item	Comp.1	Comp.2	Comp.3	Comp.4
item4N	0.493	0.001	0.834	0.248
item5	0.531	0.254	-0.075	-0.805
item6	0.502	0.506	-0.456	0.533
item8N	0.472	-0.824	-0.301	0.079

Table 7: Loadings of ethics items

Value:

Item	Comp.1	Comp.2	Comp.3	Comp.4
item1N	0.506	0.131	0.853	0.018
item2	0.604	0.06	-0.382	0.697
item3	0.609	-0.014	-0.344	-0.715
item7N	0.095	-0.99	0.094	0.055

Table 8: Loadings of value items

Post Survey Cronbach's and PCA

The Cronbach alpha measure are given in figure 5. The 95% bootstrapped confidence intervals are as follows:

Ethics: (0.46, 0.83)
Value: (0.38, 0.83)

Once again, if item seven is removed the alpha measure for value increases to 0.85. Interestingly, the alpha measure for ethics decreased quite a bit. It's important to note that the survey items between pre and post surveys had identical wording. Thus, might be a cause for concern if the survey was longer. Students might get frustrated that they are answering the exact same items twice. This might cause them to answer the items in a way that's not truthful.

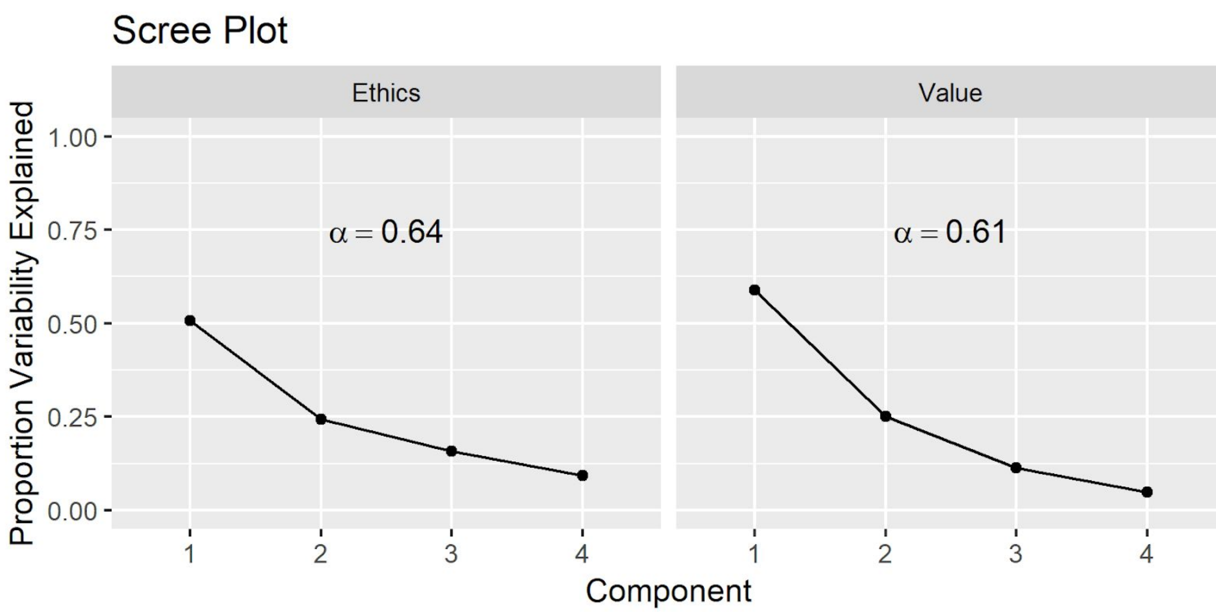

Figure 5: Scree plots of the two components in the STA 466/566 survey

Ethics:

Item	Comp.1	Comp.2	Comp.3	Comp.4
item4N	0.507	0.299	0.761	0.273
item5	0.319	0.807	-0.474	-0.153
item6	0.581	-0.359	0.016	-0.73
item8N	0.552	-0.362	-0.442	0.607

Table 9: Loadings of ethics items

Value:

Item	Comp.1	Comp.2	Comp.3	Comp.4
item1N	0.557	0.012	0.724	0.407
item2	0.61	0.07	-0.026	-0.789
item3	0.562	-0.009	-0.69	0.456
item7N	0.044	-0.997	0.013	-0.054

Table 10: Loadings of value items

Hotelling and Class Comparison

We ran the Hotelling's T^2 test to see if there is a difference in at least one of the two component scores (post - pre) between the two sections of STA 466/566. Since the p-value of 0.7863 is greater than any reasonable level of significance, we fail to reject the null hypothesis. We don't have sufficient evidence to suggest that there is a difference in at least one of the attitude components between treatment and control. The test statistic was 0.2421 on 2 and 35 degrees of freedom. One student's survey contained three responses that required adaptation due to the student circling two numbers (i.e. if 1 and 2 circled, then an imputed value of 1.5 is used).

While there is no significant difference between the sections of the class, figure 6 shows most students have positive attitudes in both components. This is not surprising since these students are majoring in statistics. Four students in the control class have overlapping scores with respect to the value component. In particular, they answered all questions with a maximum score and their responses did not change from pre to post survey. For the ethics component, the average difference in the treatment group is 0.33 (sd = 0.85) compared to 0.23 (sd = 1.02) in the control group. For the value component, the average difference in the treatment group is 0.06 (sd = 0.64) compared to -0.11 (sd = 0.79) in the control group. These averages were computed using the original scaling. On average, the control group had a decrease in value attitude scores from pre to post survey.

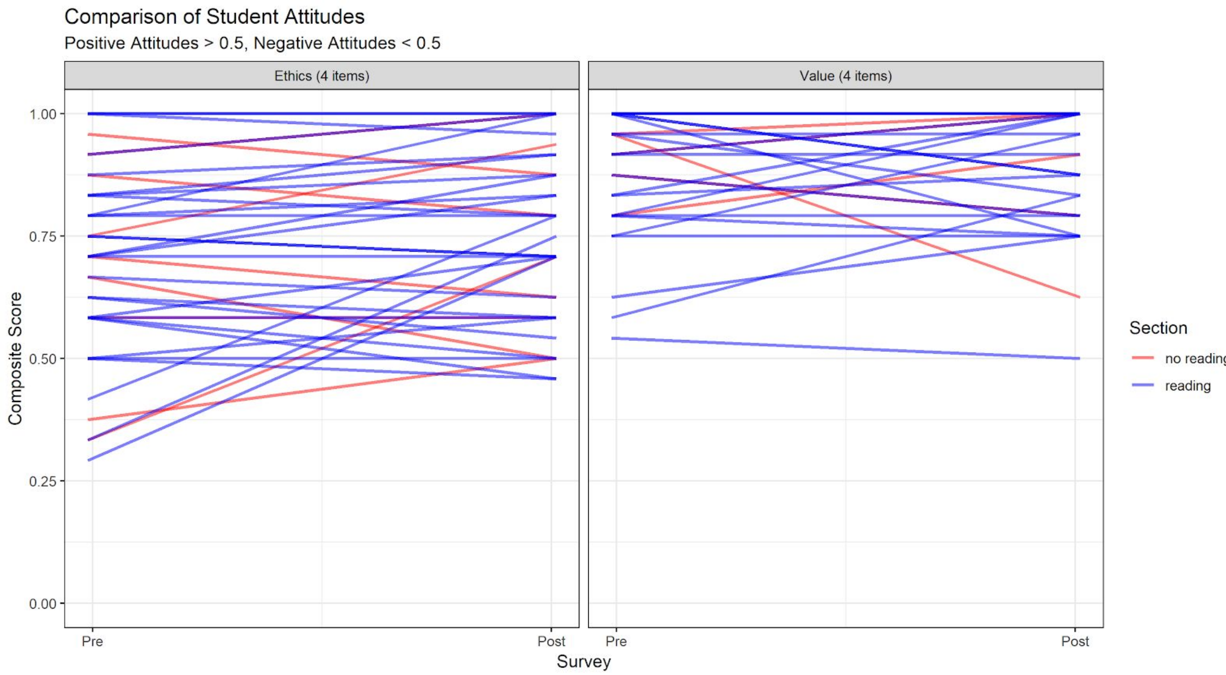

Figure 6: Comparison of attitudes between treatment and control groups

5. Discussion/ Recommendations

Student attitudes are important to consider in a statistics course because they can influence the overall learning experience as well as statistical decision-making outside of the classroom. Student attitudes have been associated with course achievement as well. It's important to recognize that attitudes are developed over a long period of time. Thus, they tend to be stable traits that do not change much except under "extreme interventions" (Lesser et al, 2016). Student attitudes are latent variables. This means that they are not directly observed or measured. We studied them using survey items that capture the constructs of attitude. However, there is a lot of variability in student attitudes from sample to sample. Not only this, conclusions can change depending on the wording of the item because the students may interpret it differently than intended.

The reading assignments were created to help students reflect on course ideas in real world applications. With just three or four reading assignments, we did not see a change in student attitudes compared to a control group from the beginning to the middle of the semester. Additionally, we would not expect to see a substantial improvement in attitudes if the students have positive attitudes to begin with. A modification to the reading assignments which allows the student to choose from a selection of articles is one consideration for future research. With this approach, we know the student is reading an article that he/she finds at least somewhat interesting.

We gained valuable insights into the attitude standing of each class. In STA 363, the affect and cognitive scores appear to increase slightly. Most students felt the class was somewhat difficult and that they would expend a lot of effort to learn statistics. There is a lot of variability in the scores for interest and value, but most students held positive attitudes. The STA 466/566 students generally had positive attitudes across both components. This is not surprising since these students are majoring in statistics.

There are a number of drawbacks related to our design and the method used to analyze the data. First, our case study worked with only two sections from each class. This does not allow for replicates of treatment and control. Thus, we cannot distinguish the treatment effect from the section effect. That is, these two effects are confounded. The second drawback is that we have assumption violations for the multivariate t-test. Even though the Hotelling's T test is robust to moderate normality violations, we have a fairly small sample of students including only nine students from STA 466. Thus, we are hesitant to make any generalizable conclusions based on this case study. Additionally, the t-test assumes a common covariance structure between classes. We know that students within a class section are correlated with one another. If the results had been significant, it's likely that we rejected the null hypothesis when the null is true. In this case, we would simulate student responses under different covariance structures to investigate the inflation of type 1 error. A third drawback is that students are not randomized to each section of a class. They signed up to be in the early or late sections. Perhaps students in the earlier class are more motivated to participate and engage in class material than the later class.

To ameliorate these issues, in any future studies it would be preferable to have replication of treatment and control. This would allow estimation of the treatment effect and also model the section properly as a random effect. Suppose we wish to eliminate the section effect on the attitude scores, we could potentially introduce the readings into an online class. In the online setting it would be feasible to randomize the readings to students within the same section. Students in an online class have opportunities to interact with the instructor through google hang-out sessions and emails. They can also interact with other students through discussion boards. However, it's more reasonable to assume the covariance structures of two sections of an online class are similar. This is because the assignments and modes of interaction are similar.

In conclusion, student attitudes are an important consideration in an introductory statistics class. Attitudes are also difficult to measure. In this case study, we did not see a change in student attitudes between treatment and control sections. However, this study sheds light on the overall attitude standing of the class. STA 363 felt that the class was difficult, but seemed to improve in affect and cognitive competence. These students expended a lot of effort in the class. The attitude scores for interest and value were more variable than the other facets. STA 466/566 valued the class and had a general awareness of statistical ethics.

6. Appendix

6.1 SATS-36 Items

Affect – students' feelings concerning statistics (6 items; .80 to .89):
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 class.
19. I will enjoy taking statistics courses.
28.* I am scared by statistics.

Cognitive Competence – students' attitudes about their intellectual knowledge and skills when applied to statistics (6 items; .77 to .88):
5.* I will have trouble understanding statistics because of how I think.
11.* I will have no idea of what's going on in this statistics course.
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 statistical concepts.

Value – students' attitudes about the usefulness, relevance, and worth of statistics in personal and professional life (9 items; .74 to .90):
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 – students’ attitudes about the difficulty of statistics as a subject (7 items; .64 to .81):

6. Statistics formulas are easy to understand.
8.* Statistics is a complicated subject.
22. Statistics is a subject quickly learned by most people.
24.* 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, new component):

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, new component):

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.

6.2 PCA Loadings (Pre Survey)

Affect:

Item	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
s3	0.3	0.7	0.411	0.007	0.132	0.482
s4N	0.423	-0.245	-0.035	0.865	0.03	0.1
s15N	0.38	-0.37	0.656	-0.233	-0.466	-0.128
s18N	0.458	-0.189	0.038	-0.274	0.783	-0.254
s19	0.419	0.491	-0.308	0.005	-0.312	-0.625
s28N	0.448	-0.19	-0.55	-0.349	-0.23	0.534

Cognitive:

Item	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
s5N	0.42	0.195	0.497	0.541	0.33	0.37
s11N	0.437	0.104	-0.514	-0.211	0.682	-0.158
s26N	0.186	-0.973	0.037	0.088	0.093	0.04
s31	0.468	0.03	-0.02	-0.548	-0.352	0.596
s32	0.432	0.055	-0.476	0.526	-0.523	-0.183
s35N	0.439	0.037	0.511	-0.278	-0.143	-0.669

Value:

Item	Comp .1	Comp .2	Comp .3	Comp .4	Comp .5	Comp .6	Comp .7	Comp .8	Comp .9
s7N	0.342	0.387	0.265	0.229	0.368	0.295	0.153	0.478	0.371
s9	0.332	0.077	-0.591	-0.199	0.104	-0.22	0.657	0.015	-0.058
s10	0.298	0.561	-0.203	-0.418	0.125	-0.031	-0.552	-0.242	-0.015
s13N	0.316	0.136	-0.228	0.338	-0.812	0.085	-0.113	0.131	0.136
s16N	0.361	-0.358	-0.08	0.152	0.201	-0.454	-0.387	0.48	-0.293
s17	0.279	-0.537	-0.062	-0.519	-0.063	0.555	-0.083	0.126	0.161
s21N	0.282	0.051	0.655	-0.414	-0.311	-0.41	0.228	0.001	0.014
s25N	0.393	0.028	0.212	0.241	0.053	0.376	0.126	-0.308	-0.698
s33N	0.377	-0.303	0.043	0.312	0.189	-0.181	-0.055	-0.595	0.491

Interest:

Item	Comp.1	Comp.2	Comp.3	Comp.4
s12	0.465	0.859	0.179	0.118
s20	0.518	-0.282	-0.438	0.678
s23	0.491	-0.417	0.763	-0.056
s29	0.524	-0.092	-0.441	-0.723

6.3 PCA Loadings (Post Survey)

Affect:

Item	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
s3	0.4	0.567	0.162	0.094	0.124	0.684
s4N	0.275	-0.462	0.786	0.046	0.3	-0.026
s15N	0.431	-0.228	-0.446	-0.549	0.511	0.025
s18N	0.397	-0.303	-0.377	0.779	0.038	-0.005
s19	0.43	0.523	0.115	0.033	0.042	-0.725
s28N	0.486	-0.22	0.023	-0.283	-0.794	0.075

Cognitive:

Item	Comp.1	Comp.2	Comp.3	Comp.4	Comp.5	Comp.6
s5N	0.395	0.086	0.811	0.136	0.349	0.194
s11N	0.502	0.125	-0.021	-0.14	-0.025	-0.844
s26N	0.142	-0.907	0.027	-0.37	0.14	0.007
s31	0.441	0.293	-0.109	-0.661	-0.297	0.427
s32	0.431	0.039	-0.572	0.263	0.609	0.214
s35N	0.438	-0.259	-0.045	0.564	-0.631	0.148

Value:

Item	Comp .1	Comp .2	Comp .3	Comp .4	Comp .5	Comp .6	Comp .7	Comp .8	Comp .9
s7N	0.311	0.023	0.818	0.143	0.221	0.192	0.177	0.183	0.25
s9	0.307	-0.438	-0.143	-0.661	0.138	0.193	-0.033	0.441	-0.065
s10	0.352	-0.345	-0.148	0.031	0.12	-0.31	0.658	-0.435	0.046
s13N	0.373	-0.016	-0.102	0.039	0.242	0.518	-0.418	-0.588	-0.036
s16N	0.335	-0.189	-0.207	0.575	-0.426	0.327	0.119	0.335	-0.254
s17	0.315	0.291	0.163	-0.402	-0.754	-0.027	0.014	-0.214	0.111
s21N	0.233	0.725	-0.366	-0.067	0.288	0.175	0.339	0.211	0.094
s25N	0.376	-0.047	-0.215	0.209	0.027	-0.44	-0.421	0.177	0.603
s33N	0.373	0.201	0.17	0.012	0.16	-0.48	-0.229	0.052	-0.693

Interest:

Item	Comp.1	Comp.2	Comp.3	Comp.4
s12	0.472	0.856	0.207	0.046
s20	0.504	-0.432	0.516	0.541
s23	0.503	-0.091	-0.824	0.244
s29	0.52	-0.27	0.109	-0.803

6.4 STA 466/566 Survey Items

Statistical Ethics - Awareness of ethical issues in statistics:
4.* I am not aware of the ethical responsibilities related to working with data.
5. I know the ethical data practices used in my career field
6. I know of cases where incorrect interpretations of data were consequential.
8.* I am not aware of common misconceptions that lead to incorrect statistical results.

Value – similar to the SATS value component:

1.* Statistics is not valuable for my personal development
2. My career will benefit from the use of statistics
3. Statistics is helpful for my professional development.
7.* I think that statistics is not useful in many areas of society

6.5 Article Questions

To Predict and Serve:

Link: https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00960.x

1.) The author mentions a specific type of sampling procedure that is used to reduce bias within a sample. What sampling procedure is this? Using information from this reading, describe two ways in which bias is introduced into police records.

2.) What kind of results are produced from a predictive model that is built with biased data? In the context of predictive policing, what are some consequences of using such a model?

3.) The author discusses a model used to investigate the effect of police-recorded data on predictive policing models. What are the three predictors used in this model? Subsequently, what are the findings from using this model to predict drug-related crime rate?

4.) An implicit assumption in the model is, "the presence of additional policing in a location does not change the number of crimes that are discovered in that location". To address this scenario, the author uses simulation to increase the number of observed crimes in targeted locations by 20%. What are the findings from using this model to predict drug-related crime rate?

5.) Upon reflection, what ethical issues in statistics are presented in this reading? What are your thoughts? You may use answers from previous questions to motivate your response.

How one woman used regression to influence the salary of many:

Link: https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2017.01092.x

1.) While working at Berkeley, which professors did Elizabeth Scott have an opportunity to collaborate with? What projects were these professors working on?

2.) Describe the events that preceded Elizabeth Scott's primary research at Berkeley. What was the single research question that motivated her research? What steps did she take with the data to prepare it for use in the analysis?

3.) For what reasons did Scott eliminate the "academic rank of a professor" as a possible predictor variable in the analysis?

4.) The author states that Scott's method for modeling salary was not perfect. Ultimately, what model did she use to predict a woman's salary? How did she use this model to answer the research question? What are the limitations of using this model?

5.) Consider the components of data preparation, variable selection, and statistical analysis discussed in this reading. What are some important components that you can incorporate in your own statistical analyses? You may use answers from previous questions to motivate your response.

Where the seeds of modern statistics were sown:

Link: https://rss.onlinelibrary.wiley.com/doi/10.1111/j.1740-9713.2018.01144.x

1.) Describe Rothamsted's two most famous experiments. Why did Lawes keep these experiments going?

2.) According to Andrew Mead, what was the most important statistical development to come out of Rothamsted? How does Ronald Fisher describe this development in his own words?

3.) Describe the "lady tasting tea" experiment. Be sure to include the claim that was tested, the factors in the experiment, and the randomization scheme. You can think of the factor as the recipe for tea. How many ways of making tea are there?

4.) At that moment in history, what were the two important concepts in experimental design? What other contributions did Fisher make in this area of statistics?

5.) Consider the "lady tasting tea" experiment. Design your own small-scale experiment. Be sure to include the claim that is tested, the factors in the experiment, and the randomization scheme.

The ASA's p-value statement, one year on:

Link: https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2017.01021.x

1.) Statistical significance testing is used in many areas of society. Give examples from this reading of professions or fields of study that use statistical significance testing. Can you think of any other areas of society that have used this approach?

2.) The author mentions one of the "biggest scientific controversies of our time". What is this controversy? In the context of advancing science, why is this controversy a problem?

3.) Please read the box labeled, "P-value problem in a nutshell". Extrasensory perception is the claim that humans have more than the known number of senses. Write out the p-value in words. Describe how p-values "spring their trap".

4.) There are scientific studies that have overstated the statistical significance of the results. Statistics is a scientific, evidence-based field. This means that statisticians consider the evidence from a multitude of perspectives before making conclusions. Below is a press release from the *American Statistical Association* about the p-value statement:

https://www.amstat.org/asa/files/pdfs/P-ValueStatement.pdf

5.) What are some consequences of overstating significance and misinterpreting p-values? You may use answers from previous questions to motivate your response.

Video Game Violence and Aggression: A Proven Connection?:

Link: https://rss.onlinelibrary.wiley.com/doi/full/10.1111/j.1740-9713.2016.00955.x

1.) What were the two types of studies that the American Academy of Pediatrics used to establish a "scientific connection between virtual violence and real-world aggression"?

2.) The reading discusses an experiment to examine the effect of violent media on aggression. Describe this experiment and how the investigator would make a conclusion.

3.) Consider the experiment in the previous question. Describe the challenge of running such an experiment. In your response, discuss the difficulty of controlling variability in this situation.

4.) The author states that there is no way to buy two copies of the same game, one with violence and one without. How do the researchers address this problem? What is the drawback of this approach?

5.) Many of these experiments are small-scale with few subjects. Discuss how a misinterpreted p-value can lead researchers to incorrect conclusions about the effect of violent video games on aggression.

Reference List

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 www.amstat.org/publications/jse/v16n1/carnell.html.

Kieth A. Carlson & Jennifer R. Winquist (2011) Evaluating an Active Learning Approach to Teaching Introductory Statistics: A Classroom Workbook Approach, Journal of Statistics Education, 19:1, , DOI: 10.1080/10691898.2011.11889596

Beth Chance, Jimmy Wong & Nathan Tintle (2016) Student Performance in Curricula Centered on Simulation-Based Inference: A Preliminary Report, Journal of Statistics Education, 24:3, 114-126, DOI: 10.1080/10691898.2016.1223529

Sherri Cheng, Mark Ferris & Jessica Perolio (2018) An Innovative Classroom Approach for Developing Critical Thinkers in the Introductory Statistics Course, The American Statistician, 72:4, 354-358, DOI: 10.1080/00031305.2017.1305293

Cronbach LJ (1951). "Coefficient alpha and the internal structure of tests". *Psychometrika*. 16 (3): 297–334.

DeVaney, T. A. (2010). Anxiety and attitude of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education, 18(1)*.

Emmioğlu , E. & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal, 11(2)*, 95-102. Retrieved from http://iase-web.org/documents/SERJ/SERJ11(2)_Emmioglu.pdf.

Everitt, S. B. & Dunn G. (2001). Applied Multivariate Data Analysis. London, ENG: Arnold.

Field, P. Andy. (2001). Meta Analysis of Correlation Coefficients: A Monte Carlo Comparison of Fixed and Random Effects Methods. *Psychological Methods, 6(2)*, 161-180.

GAISE College Report ASA Revision Committee (2016). "Guidelines for Assessment and Instruction in Statistics Education College Report 2016," http://www.amstat.org/education/gaise.

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 37-51). Netherlands: IOS Press.

Gelman, A. et al (1998). Student Projects on Statistical Literacy and the Media, The American Statistician, 52:2, 160-166

Green, J. & Blankenship, E. (2015). Fostering Conceptual Understanding in Mathematical Statistics, The American Statistician, 69:4, 315-325, DOI: 10.1080/00031305.2015.1069759

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. Psychological Methods, 3(4), 486–504.

Hotelling, H. (1931). "The generalization of Student's ratio" *Ann. Math. Stat.* , 2 pp. 360–378

Lesser, L. M., Pearl, D. K. & Weber III, J. J. (2016). Assessing Fun Items' Effectiveness in Increasing Learning of College Introductory Statistics Students: *Results of a Randomized Experiment. Journal of Statistics Education, 24(2)*, 54-62. Retrieved from http://tandfonline.com/doi/full/10.1080/10691898.2016.1190190.

Millar, A. M., & Schau, C. (2010). *Assessing students' attitudes: the good, the bad, and the ugly*. Joint Statistical Meetings, Vancouver, available at www.evaluationandstatistics.com\JSM2010.pdf.

Pearson, K. (1901). "On Lines and Planes of Closest Fit to Systems of Points in Space". *Philosophical Magazine*. 2 (11): 559–572

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The Development and Validation of the Survey of Attitudes toward Statistics. *Educational and Psychological Measurement*, *55*(5), 868–875. https://doi.org/10.1177/0013164495055005022

Schau, Candace. *"SATS Scoring".* 2019. Retrieved from: https://www.evaluationandstatistics.com/scoring

Schmid, K. (2013), “Virtual Discussion for Real Understanding: The Use of an Online Discussion Board in an Introductory Biostatistics Course,” Technology Innovations in Statistics Education, 7, 1.

Radke-Sharpe, N. (1991). Writing as a Component of Statistics Education, The American Statistician, 45:4, 292-293

Sorge, C. (2001). Impact of engineering students’ attitudes on achievement in statistics: A structural equation model analysis. Unpublished doctoral dissertation, University of New Mexico, Albuquerque, NM.

Tempelaar, D. T., Schim van der Loeff, S. & Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities, and course performance. *Statistics Education Research Journal, 6(2)*, 78-102. Retrieved from http://iase-web.org/documents/SERJ/SERJ6(2)_Tempelaar.pdf.

University of Virginia (2019). Using and Interpreting Cronbach’s Alpha. Retrieved from https://data.library.virginia.edu/using-and-interpreting-cronbachs-alpha/

Wise, S. L. (1985). The Development and Validation of a Scale Measuring Attitudes toward Statistics. *Educational and Psychological Measurement*, *45*(2), 401–405. https://doi.org/10.1177/001316448504500226

Software References:

R Core Team (2018). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Hadley Wickham (2017). tidyverse: Easily Install and Load the 'Tidyverse'. R package version 1.2.1.https://CRAN.R-project.org/package=tidyverse

Revelle, W. (2018) psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, https://CRAN.R-project.org/package=psych Version = 1.8.12.

James Curran (2018). Hotelling: Hotelling's T^2 Test and Variants. R package version 1.0-5. https://CRAN.R-project.org/package=Hotelling