THE EFFECTS OF A COURSE ON STATISTICAL LITERACY UPON STUDENTS' CHALLENGES TO STATISTICAL CLAIMS MADE IN THE MEDIA

A Dissertation
Presented to
the Graduate School of
Clemson University

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy
Curriculum and Instruction

---

by
Rose Martinez-Dawson
December 2010

Accepted by:
Robert Horton, Committee Chair
Robert Green
L Brent Igo
John Mittelstaedt

ABSTRACT

Statistical literacy involves critically evaluating and questioning statistics encountered in everyday life. The purpose of this study was to evaluate students' questions (challenges) concerning statistics encountered in everyday life and how these challenges differed before and after taking an undergraduate statistics course, entitled Statistics in Everyday Life 200 (SIEL), which focused on statistical literacy. One hundred forty-four students were given three media articles to indicate questions they had concerning statistics cited in the articles and explain why these questions were important to ask. Students' responses were categorized based on the topic and were assessed using the Statistics Literacy Components Rubric (SLCR). The SLCR is composed of eight components of statistical literacy; students' responses were scored based on the level of awareness of each component. In addition, at the end of the semester, students completed reflection papers pertaining to their past experiences and their experiences in SIEL. Responses were analyzed to determine associations between classifications of responses and sex and effort levels.

Minimal to essentially no awareness of any of the eight components of statistical literacy in SLCR were observed prior to and after taking SIEL. The highest levels of awareness were observed for the following components: Definition, Method, and Lurking variable. Significant changes in the level of statistical literacy were observed after taking SIEL for all statistical literacy components except Causality, but these changes were small and appear to be of little practical significance. Significant changes in the pattern of topic category responses were observed after taking SIEL for each media article, and differences in responses of topic categories were observed for males and females. Despite the evidence that low levels of statistical literacy were observed, reflection paper responses indicated that students believed they were statistically literate and that the course had changed the way they looked at statistics in everyday life. This study provides a foundation for future research in statistical literacy and aid in the further development of rubrics to assess components of statistical literacy.

DEDICATION

This dissertation is dedicated to the memory of my father, John H. Martinez, who taught me many things during his life including the value of education. From his illness and death, I learned compassion and forgiveness.

ACKNOWLEDGMENTS

I would first like to thank members of my committee, Dr. Bob Green, Dr. Brent Igo, and Dr. John Mittelstaedt, for their support and constructive criticisms. A special thank you is due to my committee chair, Dr. Bob Horton, whose patience with me during my eight year journey is greatly appreciated. I could not have accomplished this goal without his guidance, support, and friendship.

I would also like to thank Dr. Billy Bridges for putting up with me during the study and for his guidance during data analysis. In addition, I want to thank Cherylene Amidon for not only entering data and organizing my SAS files but for her unwavering support and friendship. I would also like to thank Dr. Jim Rieck who was always willing to give honest feedback. I would also like to thank family members, John and Patricia Martinez as well as Scott, Linda, Christine, and Jeffrey Kenison, for their support and belief that I could accomplish this goal.

Finally, I would like to thank the two most important people in my life, my husband, Dr. Paul Dawson and daughter, Elizabeth. Words cannot express how much your support and love during these eight years of juggling full time work, graduate school, and family have meant to me. Elizabeth, may you always strive to make your dreams come true.

TABLE OF CONTENTS

Table of Contents (Continued)

LIST OF TABLES

List of Tables (Continued)

Table Page

List of Tables (Continued)

Table Page

List of Tables (Continued)

Table Page

List of Tables (Continued)

Table Page

List of Tables (Continued)

Table Page

List of Tables (Continued)

LIST OF FIGURES

CHAPTER ONE

INTRODUCTION

Introduction

In today's data driven world, information is readily available and easily obtained. Some of this information may be misleading or inaccurate. Making informed decisions based on available information is an essential skill for success. People often assume a statistic appearing in print or reported on the news is true. According to Cerrito (1999), "the public can be duped by almost anyone capable of spouting numbers, percents and p-values."

Being statistically literate enables one to "consume and critically digest the wealth of information being produced in today's society" (Rumsey, 2002, p. 33). Therefore the need for a statistically literate citizenry is vital.

Milo Schield, a prominent researcher in the area of statistical literacy, suggests that "statistical literacy is more about questions than answers. It doesn't have many answers, but it should help one ask better questions and thereby make better judgments and decisions…Statistical literacy helps one answer the question asked of most statistics: "What does this mean?" (http:/www.augsburg.edu/ppages/schield). Challenging and critically evaluating statistics may increase awareness concerning the quality of information that is so readily available.

Statistical literacy involves the ability to challenge statistics encountered in everyday life. Past research on statistical literacy has not focused on evaluating the “questioning ability” of statistical literacy. This ability to question is critical in today’s world in which information is so easily accessible, voluminous and potentially unreliable.

Schools of higher education can play a role in promoting critical evaluation of these statistics. Creating courses to provide students opportunities to develop skills, attitudes, and concepts necessary to make informed decisions is important.

In this chapter, the definition and importance of statistical literacy are discussed. Furthermore, the format of a course used to address the research questions is described, and the significance of the study is explained. Research questions are also provided.

Statistics in Everyday Life 200

Aware of the significant gaps between students’ knowledge and experiences and demands of statistical literacy, statisticians developed a course to close these gaps. Statistics in Everyday Life 200 (SIEL) was created to enhance students’ capacities to challenge and question statistics, an important aspect of statistical literacy, through the critical evaluation of statistics encountered in everyday life. According to Rumsey (2002), the use of real world examples can be a motivator for students to ask questions.

Material in SIEL was structured to follow the sequence of a traditional introductory statistics course, but with an emphasis on concepts rather than formulae. Statistical concepts were contained within “everyday life modules,” reflecting aspects of everyday life. With this format, we hypothesized that students’ understanding of statistical concepts would be enhanced because of familiar contexts. These contexts served as the cornerstones of the modules in this course. Table 1.1 provides a summary of the everyday life modules and statistical topics addressed in each module.

Table 1.1: A summary of everyday life modules and course topics in SIEL

Everyday Life Module	Course Topics
Introduction to Statistics	Basic terminology (population, sample etc.)
Statistics in Society	Effect of definitions, Measuring and sampling issues on samples
Statistics in Government	Descriptive statistics, Robust statistics, Index numbers, CPI
Statistics in Surveys and Polls	Margin of error, Confidence intervals, Interpreting poll results, Sampling methods
Statistics and Probability in the Lottery	Basic probability concepts and combinatorics
Statistics and Probability in Sports	Random variables, Binomial probability, Simpson's Paradox
Statistics in Education	Validity, Reliability, Bias, Normal distribution, Correlation, Linear Regression, Simpson's Paradox
Statistics in the Liberal Arts	Constructing and interpreting graphical displays
Statistics in the Environment	Sampling distribution, Central Limit Theorem, Bias, Random error, Estimation techniques
Statistics in the Court Room	Type I and II errors
Statistics in Advertising and Marketing	Ways in which statistics are used and misused in advertising, Misuse of graphs, Hypothesis testing of quantitative and categorical information
Statistics in Medicine	Experimental design, Lurking and confounding variables

The foundation for this course format is David P. Ausubel's Meaningful Reception Theory (Ausubel, 1978). New knowledge becomes meaningful to the student if it is "anchored" to prior knowledge. In SIEL, each statistical concept is presented in a context which should be familiar to students through the use of everyday life module format. At the beginning of each everyday life module, background information concerning the context acts as an advance organizer to "set the stage" for statistical topics. To further develop the ability to critically challenge statistics in everyday life, students complete assignments or participate in activities related to the everyday life modules. Examples of these activities can be found in Table 1.2.

Table 1.2: Examples of activities performed in everyday life modules in SIEL

Everyday Life Module	Activities
Introduction to Statistics	Use class to illustrate observational studies, sample survey, census and experiment
Statistics in Society	Critique the who, why and how of student selected social statistics
Statistics in Government	Write opinion paper concerning article on unemployment using statistical concepts
Statistics in Surveys and Polls	Conduct poll, describe pros and cons of poll, and create confidence interval
Statistics and Probability in the Lottery	Determine strategy for playing the "lottery" for extra credit points using probability concepts
Statistics and Probability in Sports	Collect sports data and compute probabilities
Statistics in Education	Write opinion papers concerning high stakes testing and college rankings with a focus on validity and reliability
Statistics in the Liberal Arts	Determine authorship of manuscripts by creating and interpreting graphs
Statistics in the Environment	Collect data to compute confidence interval for totals (e.g. total emissions) Critique statistical information in *An Inconvenient Truth*
Statistics in the Court Room	Student "juries" evaluate evidence from discrimination case and make verdict
Statistics in Advertising and Marketing	Pepsi Challenge, Illustration of experimental design using advertisement
Statistics in Medicine	Discuss medical studies to identify lurking variables

A summary of statistical concepts covered in the everyday life modules that comprise SIEL can be found in Table 1.3. This table was updated from the work of Hill, Martinez-Dawson, and Bridges (2006).

Table 1.3: A summary of statistical concepts covered in everyday life modules in SIEL

Module	Producing Data	Basic Probability	Descriptive Statistics	Inferential Statistics
Society	X			
Government	X		X	
Survey & Polls	X		X	X
Lottery		X		
Sports		X	X	
Education	X		X	X
Liberal Arts	X		X	
Environment	X	X	X	X
Court room	X	X	X	
Advertising	X		X	X
Medicine	X		X	

Research Questions

The overall focus of the proposed research was to assess the effect of a statistical literacy course, SIEL, on students' challenges to statistical claims made in the media. Three main research questions were addressed in this study. Research Question 1 assessed the level of students' statistical literacy based on students' awareness to eight statistical literacy components based on the work of Utts (2002). Research Question 2 focused on what challenges students made to statistical claims in the media. For Research Questions 1 and 2, the levels of awareness of statistical literacy components and challenges made were assessed at the beginning and end of the course as well as changes during the course through the use of two advertisements and an article. Students were instructed to provide questions concerning claims made in the media articles and explain why these questions were important to ask. Research question 3 focused on responses from open-ended questions concerning students' experiences with the course as well as their past experiences with mathematics and statistics courses.

For Research Questions 1 and 2, media articles were used as a means to assess statistical literacy based on Watson (1997) who stated that "If evidence of the need for statistical literacy is found in the media, then the media is also an ideal vehicle to provide initial motivation for the study of statistics, applications of specific topics in the curriculum during instruction, and items for assessment in the final stages of learning" (p. 107). Since a connection between general literacy and statistical literacy is believed to exist (Gal, 2004), advertisements were used as two of the media articles to reduce the effects that deficiencies in general literacy might have.

For Research Question 1, the relationships between the level of awareness of statistical literacy components and demographic variables such as gender, aptitude, background attributes, and attitudes towards statistics were investigated. Research Question 2 focused on differences in areas of concern (topic categories) identified on claims made in the media prior to and after the course on statistical literacy and then further explored the relationship between gender and these claims. Research Question 3 explored the patterns that emerged as students described their experiences in the course on statistical literacy and also investigated the relationship between these patterns and both gender and effort. Effort was determined based on the percentage of assignments completed during the semester.

Although no published research was found that focused on variables that might affect statistical literacy, prior research has indicated that variables such as gender, attitudes towards statistics, and background attributes have an effect on course performance. Course performance is an important outcome of any course. Promoting statistical literacy was the goal of SIEL; therefore, enhancing statistical literacy and course performance may be related.

Gender effects were investigated since differences in course performance based on gender have been reported to exist. Females tend to outperform males when the criterion is overall course performance. Males tend to outperform females when the criterion was examination grades (Schram, 1996). Consequently, gender may have an effect on statistical literacy.

Gender has also been found to affect the number of mathematics courses that students take. Overall, females are underrepresented in pre-college mathematics courses which puts females at a disadvantage (Lane, 1990) since mathematical background has been correlated with performance in statistics (Lalonde & Gardner, 1993).

Furthermore, a positive relationship has been established between performance in statistics and basic mathematical ability (Feinberg & Halperin, 1978; Galagedera, 1998; Galagedera, Woodward, & Degamboda, 2000; Lalonde & Gardner, 1993; Nasser, 1999; Wooten, 1998). Cognitive factors such as mathematical ability, mathematical background, and cognitive dimensions of attitudes toward mathematics and statistics have been found to be related to statistics course performance (Feinberg & Halperin, 1978; Nasser, 1999**).** Additionally**,** a positive linear correlation has been observed between English ability and statistical literacy (Merriman, 2006). Verbal SAT scores may be used as a means to measure English ability. Therefore, the number of mathematics and statistics courses students taken as well as past performance indicators such as SAT scores may be factors that influence statistical literacy and were included in this study.

Course performance and the number of mathematics courses taken have been found to affect attitudes toward statistics. Roberts and Saxe (1982) reported that students who performed better on a basic mathematics test and took a greater number of mathematics courses had more positive attitudes toward statistics. Poor attitudes towards statistics may contribute to difficulties in learning basic statistical or probabilistic concepts (Shaughnessy, 1992) and may hinder the development of statistical thinking skills (Gal, Ginsburg, & Schau, 1997). Positive attitudes promote better appreciation and value for course material (Wise, 1985) and influence a student's willingness to take additional statistics courses (Gal et al., 1997; Zeidner, 1991). Attitudes toward statistics have been found to be correlated with achievement in a statistics course (Galli, Ciancaleoni, Chiesi, & Primi, 2008) which may influence statistical literacy. Therefore, it is important to consider pre-course and post-course attitudes toward statistics and their potential effects on statistical literacy.

Affective reactions to experiences in mathematics courses may also have an impact on how students relate to statistics learning (Gal & Ginsburg, 1994). This too may be a factor affecting statistical literacy and may be assessed through open-ended questions.

In addition, student effort has been shown to influence course performance. Lalonde and Gardner (1993) found that effort, defined as percentage of completed assignments, affected course achievement. Path analysis indicated that attitude had an effect on effort which in turn had an impact on course achievement. As with performance, effort may have influenced statistical literacy.

In order to assess the effects of a course on statistical literacy, the following research questions were addressed in this study:

Research question 1:

1a. What is the level of awareness of statistical literacy components for college students prior to a course on statistical literacy and what is the effect of gender, attitude, aptitude, and background upon this level?

1b. What is the level of awareness of statistical literacy components for college students after taking a course on statistical literacy and what is the effect of gender, attitude, aptitude, and background upon this level?

1c. What is the change in the level of awareness of statistical literacy components for college students who have taken a course on statistical literacy and what is the effect of gender, attitude, aptitude, and background upon this level?

Research question 2:

2a. Prior to taking a course on statistical literacy, what areas of concern (topic categories) do students make when viewing claims from the media that make generalizations?

2b. After taking a course on statistical literacy, what areas of concern (topic categories) do students make when viewing claims from the media that make generalizations?

2c. Is there a change in areas of concern (topic categories) that students make from pre-course to post-course?

2d. Are these areas of concern (topic categories) different for males and females?

Research question 3:

What patterns related to the experience of taking a course focused on statistical literacy and past experiences emerged from responses to a reflection paper?

3a. Were these patterns different for males and females?

3b. Were these patterns different for the three classifications of "effort?"

This research can provide the foundation for other studies to optimize factors such as teaching method or course format in order to help students develop the knowledge, strategies, and insights to challenge statistical claims and develop statistical literacy. Results from this study may facilitate development of future courses focused on enhancing the "questioning" aspect of statistical literacy as well as aid in the further development of rubrics to assess components of statistical literacy.

CHAPTER TWO

LITERATURE REVIEW

Introduction

In this chapter, definitions and models for statistical literacy as well as prior research on statistical literacy are discussed. Results from the literature related to assessing attitudes towards statistics and effects of achievement, gender, effort, prior experience, and aptitude are also discussed.

Definitions of Statistical Literacy

Statistical literacy is a new area of research in statistics education, and common themes can be found among definitions in the literature. Statistical literacy has been compared to a thick rope made up of two necessary components: mathematics/statistical understanding of the content interwoven with comprehending and "connecting" with the context (Tognolini, 1996). According to Watson and Callingham (2003), "statistical literacy is not just knowing curriculum-based formulas and definitions but integrating these with an understanding of the increasingly sophisticated and often subtle settings within which statistical questions arise" (p. 20). Likewise, Callingham and Watson (2005) suggested that statistical literacy is more than "number crunching" but includes understanding numbers within a context. Watson and Callingham (2003) advocated that the format of statistics courses should "move from non-context based application of statistical skills, such as 'add them up and divide' interpretations of average, to an appreciation of context" (p. 22). In addition, Watson and Callingham (2003) believed that promoting an awareness of the importance of statistics in decision making and developing the ability to identify bias and misrepresentation were important skills to address in statistics courses.

According to Townsend (2006), four abilities are necessary for statistical literacy. These abilities include understanding and interpreting statistical information, critically evaluating statistical information, applying this information to everyday life situations, and being able to communicate reactions or concerns about the information (Townsend, 2006). Likewise, Gal (2002) and Watson and Kelly (2003) stated that statistical literacy involves interpretation and critical evaluation of statistical information within varied contexts coupled with the ability to discuss and provide an opinion or concern regarding statistical claims.

Critical evaluation of statistical information, a necessary component of statistical literacy, also involves the ability to know what questions to ask (Snell, 1999), questions that are external to the data (Schield, 2004). Watson (1997, p. 110) indicated that an important aspect of statistical literacy is "to move students from a situation where they automatically believe everything they read in the media to one where they intelligently question data and claims." The highest level in the statistical literacy hierarchy is having the confidence to challenge what is read in the media (Watson, 1997). What is needed are "the critical thinking skills required to question claims made without proper justification" (Watson, 2002, p. 29). These questions include what population is chosen, which variables are chosen, which variable definitions are used, what effects lurking variables have on results, and what effect sample size has on statistical significance (Schield, 2004).

Statistical literacy skills are grouped into two main clusters: effects of variation and how data were obtained (Moore, 2001). For evaluating the effects of variation, it should be understood that variation is ever-present, conclusions are never certain, and inferences from coincidence should be avoided. When evaluating how data were obtained, one should consider whether data were obtained from an experiment or observational study, who conducted the study, that correlation can be misinterpreted as implying causation, and what effects lurking variables have on inferences. A statistically literate person should know the importance of replication and randomization, recognize attributes of good and bad graphs, and understand that evaluation of statistical significance requires information such as the size of the sample (Snell, 1999). Statistical literacy skills also include understanding basic statistics from tables and charts found in the media, interpreting political poll results, making decisions from data, and comprehending results from medical studies (Barbieri & Giacche′, 2006).

Statistical literacy not only affects the individual but has broader implications for society in general for it involves not only understanding and critically evaluating statistical information encountered in everyday life but also appreciating the importance that statistical thinking can have on decisions that are made (Watson & Moritz, 2000). These decisions not only concern public and private issues but also professional and personal concerns (Wallman, 1993). “Statistical literacy is not only important to our society as a whole; it is relevant to the individual members of society as they make decisions in their personal lives based on information and risk analysis provided by others in the community" (Watson & Callingham, 2003, p. 6). Statistically literate people have "a lasting appreciation of the value of statistics in their everyday lives as decision makers and citizens" (Schield, 2004 p. 62). Barbieri and Giacche′ (2006) referred to statistical literacy as a "tool for democracy, as a skill that should be in the cultural baggage of every good citizen" (p. 3). The focus of statistical literacy should be on what is necessary for "survival in the world outside the classroom" (Watson & Callingham, 2003).

In the many definitions of statistical literacy presented in this section, there is a common theme; statistical literacy concerns understanding statistics within a context and questioning those statistics. For example, statistical literacy not only involves computing means but understanding what averages represent within a context. Statistical literacy also concerns questions such as: "Is the average the best representation for a given sample?" "How was this average obtained?" and "How large a sample was used to obtain the average?"

Along with this understanding is the ability to question statistical claims. This ability requires a deep understanding that involves integration of statistical concepts. The focus of this research is on this aspect of statistical literacy - questioning statistics encountered in everyday life. This aspect concerns what questions students ask when presented with statistical claims made in the media before and after taking a course focused on statistical literacy.

Models of Statistical Literacy

An Overview of Two Types of Models

According to Sanchez (2007), statistical literacy models can be grouped into two types, sequential models and longitudinal models. Figure 2.1 provides an illustration of these models.

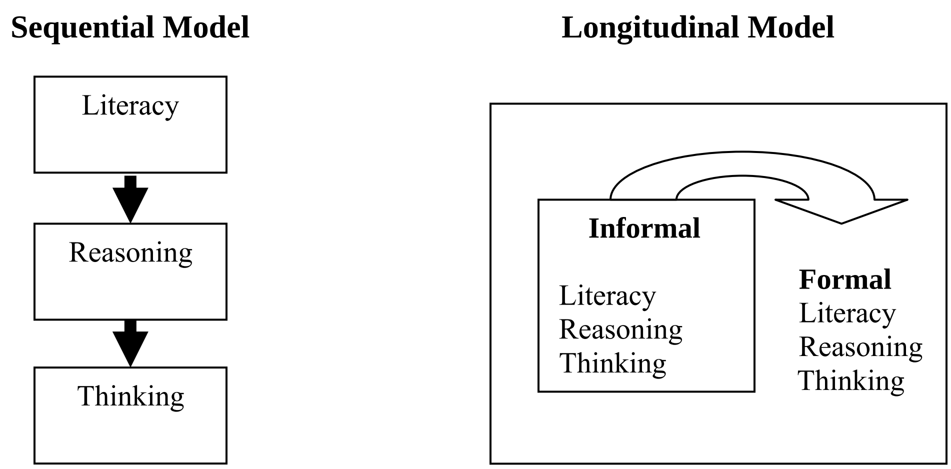


Figure 2.1: General Models of Statistical Literacy (Sanchez, 2007, p.1)

The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) web site (http://app.gen.umn.edu/artist/publications.html) is an example of a sequential model. In this model there is a sequential path from statistical literacy to statistical reasoning to statistical thinking. According to the ARTIST web site, statistical literacy is defined as understanding basic statistical terminology. Statistical reasoning refers to the ability to understand, explain, and interpret statistical information while statistical thinking involves the capacity to understand the research process from conception of research questions, data collection procedures, selection and use of appropriate of inferential tests, and drawing conclusions within the context of a situation.

A concern with this definition of statistical literacy as described in the ARTIST web site is that it is a narrowly focused definition. It does not include aspects such as understanding statistics within a context or questioning statistics.

In contrast, statistical literacy consists of two levels in longitudinal models, informal and formal. The informal level involves understanding the entire statistical process from collection of data, description of data, and summarization of data. The formal level includes understanding statistical concepts such as sampling distributions, hypothesis testing, and confidence intervals.

The following section focuses on three often referenced models of statistical literacy in the literature. These models include the hierarchical model described in Watson (1997), the two element model from Gal (2004), and the components of statistical literacy model explained in Watson (2006). The 1997 model has characteristics of the sequential model while the 2006 model and Gal (2004) model are examples of longitudinal models.

The Hierarchical Model (Watson, 1997)

Watson (1997) developed a theoretical model for statistical literacy based on the models of learning from developmental psychology. This model involves a "three tiered hierarchy." The first tier of statistical literacy includes a basic understanding of statistical terminology including terms related to numerical and graphical descriptive statistics. The second tier builds on the first tier through the understanding of this terminology within a social context such has having the ability to read and interpret commonly encountered statistical messages found in newspapers and magazines. The requirement for the second tier is to be able to make decisions and draw conclusions based on everyday statistics. The third tier, the highest level in the statistical literacy hierarchy, pertains to the development of a “questioning attitude” and involves the use of more sophisticated concepts in order to contradict claims made without proper statistical basis along with the ability to confidently challenge claims made in the media. Watson and Moritz (2000) have found the Hierarchical Model useful in studies in which students judge claims in the media.

The Two Element Model (Gal, 2004)

A model of statistical literacy proposed by Gal (2004) involves the interrelationship between interpretation and critical evaluation of statistical information from diverse contexts. Statistical literacy also involves the ability to discuss and communicate concerns about statistical information (Gal, 2004). This philosophical model consists of two elements: knowledge elements and dispositional elements. Components of the knowledge and dispositional elements are described in greater detail in the following sections.

Knowledge elements.

Knowledge elements are composed of literacy skills, statistical knowledge, mathematical knowledge, context/world knowledge, and critical questions. These components affect the ability to understand, interpret, and critically evaluate statistical information.

Literacy Skills.

Since most statistical messages are presented through written or oral means or in tabular format, general literacy skills represent an important component of the knowledge element of statistical literacy (Gal, 2004). Survey results from the International Adult Literacy Survey (Statistics Canada and OECD, 1996) indicated that a large percentage of adults have only basic comprehension skills. Therefore, a lack of literacy skills including basic comprehension skills may impede skills necessary for statistical literacy (Gal, 2004).

Statistical Knowledge.

A basic understanding of statistical and probabilistic concepts and procedures is necessary for statistical literacy. Gal (2004, pp. 56-57) proposed a list of essential topics that should be taught in high school. These topics are listed in Table 2.1.

Table 2.1: Essential statistical topics to be taught in high school as identified by Gal (2004)

Essential Statistical Topics
Number sense
Understanding variables
Interpreting tables and graphs
Planning a survey or experiment: Good sample attributes, sampling methods and questionnaire design
Data analysis processes: Descriptive statistics and detecting patterns
Relationship between probability and statistics: Characteristics of random samples, significance testing background
Inferential statistics and reasoning

Gal (2004) divided statistical knowledge into five components. These components include knowing why obtaining data are important and ways in which data can be obtained; understanding basic terminology and concepts concerning graphical and descriptive statistics; understanding basic probabilistic concepts; and knowing how statistical inferences are made. Each of these components of statistical knowledge is discussed in the following paragraphs.

The first component of statistical knowledge involves understanding that samples are the basis for inferences concerning populations and that unrepresentative samples can adversely affect such inferences. The sampling method used to obtain the sample should be critically evaluated. Understanding advantages of probability sampling methods is important since results obtained from convenience sampling should be interpreted with caution. In addition, the size of the sample should be considered when interpreting statistical results.

The second component of statistical knowledge pertains to understanding what descriptive statistics such as proportions, means, and medians measure. When selecting statistics to describe data, consideration should be given to the distribution of data. For example, it may be more appropriate to describe a sample containing an outlier with the median rather than the mean (Gal, 2004).

Being able to read and interpret data presented in graphical or tabular form is an important element in the third component of the statistical knowledge base. This component also includes understanding how tables and graphs can be presented to mislead or give false impressions (Gal, 2004). Understanding basic concepts of probability and random phenomena and the role they play in statistical inference is the key element of the fourth component of statistical knowledge (Gal, 2004).

The fifth component of statistical knowledge involves knowing how statistical conclusions or inferences are reached. This includes understanding the effect of errors or biases on inferences and the role that designed experiments play in controlling such errors. Differentiating between statistical significance and practical significance is important because despite trends appearing to exist and results being statistically significant, these trends may not be large enough to be of practical significance (Gal, 2004).

Mathematical Knowledge.

Another component of the knowledge element of statistical literacy is mathematical knowledge. This knowledge involves understanding how to compute statistics such as percentages, arithmetic means, and medians (Gal, 2004). Mathematical knowledge also includes understanding of the effect of extreme values on the mean and appropriate use of measures of central tendency when data are skewed. Gal (2002) indicated that too much emphasis placed on mathematical theory may inhibit understanding of important statistical concepts.

Context Knowledge.

In addition to mathematical knowledge, context knowledge is another component of the knowledge element of statistical literacy. Since data are context-based, context can affect correct interpretation of data. An unfamiliar context can affect understanding of sources of variation, associations, and factors that might adversely affect statistical conclusions. Therefore, it is my belief that familiar contexts should be used in teaching statistical literacy. This can be accomplished by using current events or using contexts from everyday life situations that students already understand.

Critical Questions.

Gal's final component of the knowledge element of statistical literacy involves critical evaluation of statistics encountered in everyday life. Gal (2004) listed ten "worry questions" that should be considered when confronting statistical messages. These "worry questions" can be found in Table 2.2.

Table 2.2: Ten "worry questions" to consider for statistical messages (Gal, 2004, p. 67)

Ten "worry" questions
Where did the data come from? What kind of study was it? Is this kind of study reasonable in this context?
Was the sample large enough? Did the sample include people/units which are representative of the population? Is the sample biased in some way? Overall, could this sample reasonably lead to valid inferences about the target population?
How reliable or accurate were the instruments or measures used to generate the

reported data?
What is the shape of the underlying distribution of raw data (on which this summary statistic is based)? Does it matter how it is shaped?
Are the reported statistics appropriate for this kind of data? E.g., was an average used to summarize ordinal data; is a mode a reasonable summary? Could outliers cause a summary statistic to misrepresent the true picture?
Is a given graph drawn appropriately, or does it distort trends in the data?
How was this probabilistic statement derived? Are there enough credible data to justify the estimate of likelihood given?
Overall, are the claims made here sensible and supported by the data? E.g., is correlation confused with causation, or a small difference made to loom large?
Should additional information or procedures be made available to enable one to evaluate the sensibility of these arguments? Is something missing? E.g., did the writer "conveniently forget" to specify the base of a period of a reported percent-of-change, or the actual sample size?
Are there alternative interpretations for the meaning of the findings or different explanations for what caused them, e.g., an intervening or a moderator variable affected the results? Are there additional or different implications that are not mentioned?

Dispositional Elements of Statistical Literacy.

In addition to knowledge elements, dispositional elements represent another element in the statistical model Gal (2004). Dispositional elements include beliefs and attitudes as well as critical stance. Statistical literacy involves the ability to interpret and critically evaluate statistical messages as well as communicate concerns about these messages.

These abilities are dependent on two interrelated dispositional elements: critical stance *and* beliefs and attitudes (Gal, 2004). Critical stance refers to the propensity to question statistical messages without prompting (Gal, 2004). These questions may be similar to those listed in Table 2.2.

Beliefs and attitudes can influence critical stance by affecting the "willingness to invest mental effort or occasionally take risks as part of acts of statistical literacy" (Gal, 2004, p. 69). Attitudes are stable, intense feelings that develop over time from "gradual internalization of repeated positive or negative emotional responses" (Gal, 2002, p. 18). Attitudes can be represented on a continuous scale such as like to dislike and can refer to feelings towards people, objects, or topics. Like attitudes, beliefs develop over time. Beliefs are ideas or opinions about people including oneself, social matters. Beliefs are less emotionally intense than attitudes.

The two element model from Gal (2004) is the first model of statistical literacy that takes into account the effect of dispositional elements on statistical literacy. In addition, literacy skills are considered to be an important component of the knowledge element. These components are also incorporated into the model described in Watson (2006).

Components of Statistical Literacy Model (Watson, 2006)

In addition to the previous models of statistical literacy, Watson (2006) described six components which contribute to statistical literacy. These components include mathematical/statistical skills, context, task motivation, task format, literacy skills, and knowledge concerning variation. The relationships among these components are presented in Figure 2.2 and are discussed in the following sections.

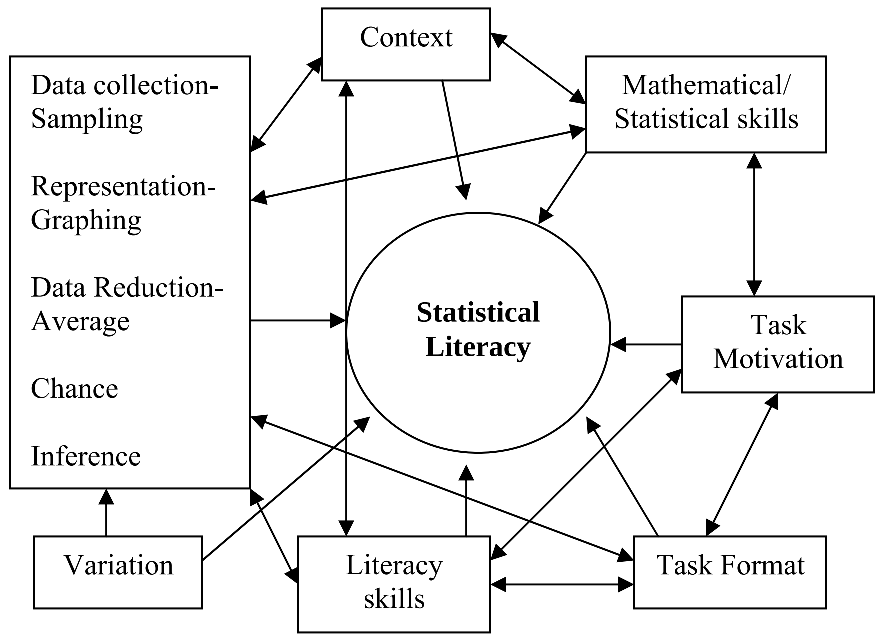


Figure 2.2: Relationship between components of statistical literacy (Watson, 2006, p. 48)

The mathematical/statistical skills needed for statistical literacy include understanding and calculating averages and performing basic probability calculations including probabilities involving compound independent events. Variation is another component of statistical literacy and understanding the effect of variation on data collection, data reduction, chance, and inference are also important (Watson, 2006).

Watson (2006) also believed that context plays a role in statistical literacy. According to Watson (2006), three levels of context exist, each with increasing complexity. The first level of context involves examples typically seen in an introductory statistics class such as tossing a die or reading tables with “mathematical, isolated contexts” (Watson, 2006, p. 249). The second level of context relates to more personal contexts such as school issues or experiences, and the third level of context concerns potentially unfamiliar media-based contexts. Watson (2006) believed that not until the context is understood can critical questioning occur. This assertion was based on several research studies described in Watson (2006) including one study in which students were asked to graph information from various contexts. The results indicated that graphical displays from contexts which were of interest and familiar to students were more advanced than those from unfamiliar or interesting contexts.

Additional components of statistical literacy include task format and motivation. Task formats are grouped into two types, open-ended questions and multiple choice questions. Watson (2006) preferred multiple-choice questions to assess statistical literacy because these types of questions permit students to "show recognition rather than creation of an appropriate answer" (Watson, 2006, p. 251). Task motivation refers to students' dispositions toward the statistical literacy task. Gal (2004) refers to these dispositions in his model as "critical stance" and "beliefs and attitudes." Other dispositions needed for statistical literacy include skepticism, curiosity, imagination, logic, and persistence (Wild & Pfannkuch, 1999).

For any task format, general literacy skills are another important component of statistical literacy. Four strategies needed for general literacy as described by Luke and Freebody (1997) are discussed in Watson (2006) with respect to statistical literacy. The first strategy involves the "code breaking" aspect of literacy. In terms of statistical literacy, this strategy pertains to understanding and interpreting graphs. The second strategy pertains to context-based understanding which is related to understanding different interpretations of averages. The emphasis of the third strategy concerns understanding how information is used which parallels understanding concepts such as "samples" (Watson, 2006). The fourth strategy involves "reading between the lines" and focuses on understanding the underlying meaning of text as well as what inferences can be made. This strategy focuses on questioning statistical claims (Watson, 2006).

Since the research for this dissertation focuses on challenges students made concerning statistical claims encountered in everyday life, articles from the media were used in this research to assess what questions students asked concerning these claims. Therefore, research involving media articles in assessing statistical literacy is described in the following section.

Using the Media to Assess Statistical Literacy

"If evidence of the need for statistical literacy is found in the media, then the media is also an ideal vehicle to provide initial motivation for the study of statistics, applications of specific topics in the curriculum during instruction, and items for assessment in the final stages of learning" (Watson, 1997, p. 107). Media and current event articles with authentic social contexts should be used to assess statistical literacy (Gal, 1997; Watson, 1997; Cerrito, 1999; Watson & Callingham, 2003). Gelman, Nolan, Men, Warmerdam, and Bautista (1998) suggested using examples from the media in an introductory statistics course as a way to make the "connection between statistics and the outside world" (p. 160) and to "shake the students out of complacent thinking" (Cerrito, 1999, p. 13). One such example from Cerrito (1999) focused on the safety of immunizations. Students were asked to evaluate the claims from an article with negative views on immunizations. According to Cerrito (1999, p. 10), “most students react negatively to the article because it contradicts their long held beliefs.”

By using these types of articles, students are challenged to evaluate strengths and weaknesses of the information in the article. Context should be familiar to students because “students seem to give more credibility to examples gleaned from the real world, these examples can serve as a powerful motivator for students to ask more questions” (Rumsey, 2002, p. 34).

Gelman et al. (1998) described assignments in which students were asked to raise questions regarding newspaper articles. Questions that students were to consider concerned the type of study conducted, the study’s protocol, the statistical methods used, the stated conclusions, and the generalizations that were made. In addition, questions were asked concerning the accuracy of the study’s results and appropriateness of conclusions drawn in the study. This approach directed students to ask particular questions about the article instead of seeing what questions students would ask if not prompted.

Delmas, Garfield, Ooms, and Chance (2007) found that after completing a first course in statistics students did not show improvement in understanding the importance of randomization on sample selection and that correlation does not imply causation. Also, results indicated that there were actually more misconceptions from students after the statistics course. For example, students believed that random assignments and random sampling were equivalent; sampling error can be reduced by random assignments; and correlation implies causation (Delmas et al., 2007). These results may have been influenced by the inclusion of students from colleges, universities, and two-year or technical colleges in which course rigor and/or quantitative abilities may be lacking. Another issue with the research was that students were given extra credit for completing the assessment instrument. Students may have not applied themselves which could have resulted in lower post-course assessment scores.

In recent years, interest in understanding statistical literacy has increased. An overview of current research on statistical literacy is discussed including several instruments developed to assess statistical literacy.

Prior Research on Statistical Literacy

Schield's Statistical Literacy Inventory (www.statlit.org/survey) is an instrument used to assess statistical literacy through evaluation of statistics in graphs or tables similar to those which might be found in newspaper articles. This instrument is composed of 69 questions concerning proportions and rates. Response options include "yes," "no," and "don't know." Although the number of correct answers can be tabulated, insight into thought processes involved in response selection or the questioning nature of statistical literacy cannot be determined from this format.

In Schield (2006), the Statistical Literacy Inventory was administered to 169 participants including college students in the United States, college teachers from throughout the world, and professional data analysts in the United States and South Africa. Based on occupation, highest overall scores were obtained by college faculty (71%), and lowest overall scores were obtained by K-12 math teachers (45%). The average score for college students was 51.2%. For data analysts, the average score was 56.4%. Native English speakers scored higher (57%) than those learning English (48%).

Of the college students in the study, 19% misinterpreted a pie chart, and 62% did not correctly compare pieces of a pie chart. Eighty-two percent of students incorrectly compared percentages while 44% did not recognize Simpson's paradox, a situation in which a comparison that holds for all of several groups disappears or even reverses directions when the data are combined to form a single group.

Another instrument, The Statistical Literacy Assessment Scale (SLAS) (http://course1.winona.edu/cblumberg/reston.pdf), was developed to assess statistical literacy of graduate students in a statistics course for education majors (Reston, 2005). SLAS is composed of 15 items and focuses on two dimensions: understanding statistical terminology and concepts associated with real-world contexts and understanding claims and arguments presented in the media. For evaluation of tables, response choices included "yes-no-cannot tell." One point was assigned to each correct answer. For open ended-questions concerning claims found in the media, student responses were evaluated based on a three-point rubric scale. A score of 2 was assigned to responses in which the reasoning was correct and justified based on statistical concepts. A score of 1 was given for reasoning that was partially correct. If no attempt to reason was observed, a score of 0 was given.

A concern with SLAS is that the context used in this instrument is specific to current events in the Philippines. If SLAS were administered elsewhere, the unfamiliar context may affect students' abilities to respond. In addition, SLAS does not address the "questioning attitude" of statistical literacy and the challenges students make concerning media claims.

In Reston (2005), 56 graduate students from the Philippines enrolled in an introductory graduate statistics course for education majors were given SLAS after the statistics course. The mean score on the 15-item survey was 6.93 with a standard deviation of 2.39 and range of 10. The correlation between SLAS scores and final course grade was 0.74. Although the mean SLAS score was low, Reston (2005, p. 2) suggested that the discussions showed that "graduate students were able to construct their own meaning of statistical literacy and realize its importance in their professional and personal lives."

One of the major researchers in the area of statistical literacy is Jane Watson. Over time, her approach to assessing statistical literacy has evolved. What follows is a description of the progression of her research on statistical literacy including the development of instruments to assess statistical literacy.

Watson and Moritz (2000) conducted a longitudinal study to assess understanding of the concept of "samples" over three years involving students who were in grades 3, 6, and 9 at the beginning of the study. An 11-item media survey focused on questions concerning newspaper articles was administered to students. Responses were evaluated based on four levels of increasing understanding (pre-structural, unistructural, multistructural, and relational) as described in Watson and Moritz (2000). Results indicated that the development of understanding the concept of "sample" increased with grade following the unistructural to multistructural to relational structure. Overall, students' understanding of "sample" developed to higher levels over time. The percentage questioning claims based on biased samples increased significantly from 22% to 66%. Over time, it was observed that some students reverted to a lower level but it was hypothesized that this might be due to a lack of motivation from the students to demonstrate their understanding (Watson & Moritz, 2000). This finding was a motivation for accounting for "effort" in this dissertation.

Contexts of media articles may have affected students' understanding of sample size (Watson & Moritz, 2000). For example, girls were more likely to give the incorrect interpretation for an item involving cars. Identification of potential sources of bias in sample selection may have been affected by real life contexts which Watson and Moritz (2000) believed may have been a distraction for students. Even though media articles were used in the survey, the questions asked in the survey may have led students to challenge certain aspects of the article. Watson and Moritz (2000) suggested that future research focus on the effect context and questions asked may have on students' responses. Furthermore "assessment of student outcomes needs to distinguish between simple 'recognition' of the meaning of the concept and the level of structuring of that recognition when applied in social contexts" (Watson & Moritz, 2000, p. 130). To assess growth in understanding of statistical concepts, Watson and Moritz (2000) advocated the use of very open-ended questions followed by more directed questions. This result was the impetus for using open-ended questions from a variety of contexts in this research.

Watson and Callingham (2003) conducted two large scale research projects involving over 3000 students in grades 3 through 9. Unlike Watson and Moritz (2000), Watson and Callingham (2003) used open-ended questions from media articles because they believed these would allow students to show their understanding at a higher level. Responses were evaluated and grouped into six levels of understanding. These levels included Level 1: Idiosyncratic, Level 2: Informal, Level 3: Inconsistent, Level 4: Consistent non-critical, Level 5: Critical and Level 6: Critical mathematical. Responses allocated to Levels 1 or 2 focused on a single element and indicated that students had difficulty understanding the contexts. For Levels 3 and 4, context was considered in the response. Responses for Levels 5 and 6 denoted more advanced reasoning and understanding of nuances of language and context. Students' responses were characterized into levels of statistical literacy as described in Table 2.3. The use of characterizing students' responses into levels of understanding was the basis of SLCR, the instrument developed and used in this research. This instrument is described in more detail in Chapter 3.

Table 2.3: Statistical literacy levels and descriptions (Watson and Callingham, 2003, p. 14)

Level	Brief characterization of step levels of tasks
6. Critical Mathematical	Task-steps at this level demand critical, questioning engagement with context, using proportional reasoning particularly in media or chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language.
5. Critical	Task-steps require critical, questioning engagement in familiar and unfamiliar contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology, qualitative interpretation of chance, and appreciation of variation.
4. Consistent Non-critical	Task-steps require appropriate but non-critical engagement with context, multiple aspects of terminology usage, appreciation of variation in chance settings only, and statistical skills associated with the mean, simple probabilities, and graph characteristics.
3. Inconsistent	Task-steps at this level, often in supportive formats, except

	selective engagement with context, appropriate recognition of conclusions but without justification, and qualitative rather than quantitative use of statistical ideas.
2. Informal	Task-steps require only colloquial or informal engagement with context often reflecting intuitive non-statistical beliefs, single elements of complex terminology and settings, and basic one-step straightforward table, graph, and chance calculations.
1. Idiosyncratic	Task-steps at this level suggest idiosyncratic engagement with context, tautological use of terminology, and basic mathematical skills associated with one-to-one counting and reading cell values in tables.

Based on Rasch modeling techniques, results indicated that the construct of statistical literacy is a one-dimensional construct since "mathematical skills and understanding of contexts, as well as content from the school curriculum, were all aspects of the same construct" (Watson & Callingham, 2003, p. 19). The six levels of the statistical literacy construct as described in Watson and Callingham (2003) are related to the tiers of the hierarchical model from Watson (1997). For example, Tier 3 which involves questioning statistical claims can be found in Levels 5 and 6. Higher levels of statistical literacy were not always related to higher ability because students at the same ability level may respond differently to a particular question.

Similar characteristics were found between the six levels of statistical literacy as described in Watson and Callingham (2003) and the three-tiered hierarchical model from Watson (1997) discussed earlier in this chapter. Skills necessary in Tier 1 of the hierarchical model are comparable to the mathematical and statistical skills listed in Table 2.3. Tier 2 and Level 3 are similar as well as Tier 3 is similar to Levels 5 and 6.

Since the sample only involved students from the Australian state of Tasmania, caution should be taken to make inferences to all students since factors such as culture, gender, instruction, and prior experiences may have affected responses. Watson and Callingham (2003) recommended incorporating more socially-based examples from the media in curricula to promote and enhance statistical thinking and reasoning within a variety of contexts. “Statistical literacy is incomplete without the opportunity to engage with genuine social contexts, particularly such as those found in the media items” (Watson & Callingham, 2003, p. 21).

Using a 4-point coding system, Watson and Kelly (2003) evaluated students’ definitions and student-provided examples of “sample,” “random,” and “variation” before and after lessons on these topics. A code of 0 was assigned if the student’s response did not indicate an understanding of the term. A code of 1 was given if the response included a “single idea,” and a code of 2 was assigned if the student gave a straightforward explanation of the term. Responses with complete explanations and examples were given a code of 3. In comparing pre-test and post-test responses, an improvement in understanding “sample” was observed for grades 3, 5, and 7. At grade 7, improvement in understanding “random” and “variation” was observed. For 9th graders, improvement in understanding was only seen for “random.” The study concluded that instruction over time can improve students’ understandings of terms associated with statistical literacy.

Callingham and Watson (2005) developed the Statistical Literacy Scale, an instrument to assess statistical literacy. The instrument consisted of 50 items that focused on three subgroups of statistical concepts: average/chance (AC), sample/inference (SI), and graphing/variation (GV).

For the AC subgroup, items pertained to measurements of chance and average. The following is an example of an item from the AC subgroup.

> Nine students in a science class weighed a small object separately on the same scale. The weights (in grams) recorded by each student are shown below.
>
> 6.3   6.0   6.0   15.3   6.1   6.3   6.2   6.15   6.3
>
> The students had to decide on the best way to summarize these values. Ben said, "I'd use the most common value to get the mode. That's 6.3." Is Ben's way a good way to summarize the information? Explain your answer.

Items in the SI subgroup focused on how samples were obtained and drawing appropriate inferences based on samples. The following is an example of an item from the SI group.

> A class wanted to raise money for their school trip to Movieworld. They could raise money by selling raffle tickets for a Nintendo Game system. Before they decided to have a raffle they wanted to estimate how many students in the whole school would buy a ticket. They decided to do a survey to find out first. The school has 600 students in grades 1-6 with 100 students in each grade. How many students would you survey? How would you choose them? Explain your answers.

Items concerning data presented in tabular or graphical format and issues related to variation comprised the GI subgroup. Items in this instrument focused on interpretation of graphs and discussion of relationships between variables which were graphed. Some graphs appeared in newspapers but most of the items concerned context of interest to the sample group in this study, children in grades 5 through 10. An example of an item from the GI subgroup follows.

A primary school had a sports day where every student could choose a sport to play. Here is what they chose.

	Netball	Soccer	Tennis	Swimming	Total
Boys	0	20	20	10	50
Girls	40	10	15	10	75

How many girls chose tennis? What was the most popular sport for boys? How many children were at the sports day? One of the tennis players was late. Was this player a boy or a girl? Explain your answer.

Since the context of some items in the Statistical Literacy Scale pertained to younger age groups, use of this instrument for older age groups may not be appropriate. Also, the contexts of several items were contrived. For example, several items dealt with determining values that would be expected when rolling a die.

The Statistical Literacy Scale was administered to 673 students in the above mentioned grades from five schools in the Catholic system in the Australian state of Tasmania. Different grades were used to represent different abilities. Responses were coded as 0-1 or 0-5 based on the complexity of the response.

Results of Callingham and Watson (2005) indicated that the Statistical Literacy Scale had high item (0.96) and case (0.87) reliabilities which suggested that the scale defines a single construct. The three subgroups were each strongly correlated with the overall scale. AC and GV were moderately correlated with a correlation coefficient of 0.68, while GV and SI were more strongly correlated with a correlation coefficient of 0.79. The lowest correlation (0.62) was seen between subgroups SI and AC. This correlation between GV and SI may have occurred since both subgroups are of a qualitative nature while AC consists of items that are quantitative in nature. This indicated that there may be "overlapping components of the statistical literacy construct" which "implies that there may be different developmental pathways for students with respect to different subgroups" (Callingham & Watson, 2005, p. 30).

Factors That May Affect Statistical Literacy

Determining factors that may influence the level of statistical literacy is important. Factors such as attitudes toward statistics, gender, previous experience with mathematics and statistics courses, and effort may have an effect on statistical literacy. Each is discussed in the following sections.

Attitudes toward Statistics

Attitude can be defined as an "affect for or against, evaluation of, like or dislike of, or positiveness or negativeness toward a psychological object" (Mueller, 1986, p. 1). Aiken (1980) described attitude as follows:

> Attitudes may be conceptualized as learned predispositions to respond positively or negatively to certain objects, situations, concepts, or persons. As such they possess cognitive (beliefs or knowledge), affective (emotional, motivational), and performance (behavior or action tendencies) dimensions (p. 2).

Attitudes are fairly stable, intense feelings that develop over a period of time and are expressed along a positive-negative continuum (Gal et al., 1997).

Attitude toward statistics has been defined as the "summation of emotions and feelings experienced over time in the context of learning statistics" (Gal et al., 1997). Olson and Zanna (1993) referred to attitudes toward statistics as a multidimensional concept – including affective, cognitive, and behavioral dimensions.

Researchers believe that it is important to study student attitudes because of their potential impact on students' performance, achievement, and willingness to take more courses in the subject area (Gal et al., 1997; Zeidner, 1991). Positive attitudes promote better appreciation and value for course material (Wise, 1985).

Gal et al. (1997, p. 40) listed key beliefs associated with statistics. These beliefs included "beliefs about mathematics; beliefs about the extent to which statistics is a part of mathematics, or requires mathematical skills; beliefs about what should happen or transpire in a statistics course, or expectations as to the culture of a statistics classroom; beliefs about oneself as a learner of statistics or mathematics; and beliefs about the usefulness or value of statistics and its importance in one's future life or career."

Attitudes toward statistics influence the teaching/learning process (Gal et al., 1997), and poor attitudes towards statistics may contribute to difficulties in learning basic statistical or probabilistic concepts (Shaughnessy, 1992). Furthermore, students' attitudes toward statistics may affect the development of statistical thinking skills (Gal et al., 1997), and negative attitudes can be a major obstacle to effective learning (Waters, Maretelli, Zakrajsek, & Popovich, 1988).

According to Gal et al. (1997), there may be three sources of attitudes towards statistics for students enrolled in an introductory undergraduate statistics course. These include past experience with statistics in school, out-of-school experience, and beliefs that statistics is mathematics. If students have had a negative experience with statistics, this will adversely affect their current attitude (Gal et al.). Students may have a negative attitude towards statistics because they do not have a clear understanding of what statistics is. Another source of attitudes toward statistics is the belief that statistics is mathematics. Negative attitudes toward mathematics are often transferred to statistics (Gal et al.).

Research indicated that attitudes have an effect on learning statistics. Gal and Ginsburg (1994) believed that assessing attitudes toward statistics is valuable because it can be used to monitor changes throughout the course.

Attitude assessment in statistics education

Several instruments have been developed to measure students' attitudes toward statistics. These instruments include the Statistics Attitude Survey (SAS) (Roberts & Bilderback, 1980), Statistical Anxiety Rating Scale (STARS) (Cruise, Cash, & Bolton, 1985), Attitudes Toward Statistics (ATS) (Wise, 1985), Statistics Test Anxiety (STA) (Benson, 1989), Statistics Attitude Scale (McCall, Belli, & Madjidi, 1990), Multifactorial Scale of Attitudes Toward Statistics (Auzmendi, 1991), Students' Attitudes toward Statistics (STATS) (Sutarso, 1992), Statistics Anxiety Inventory (SAT) (Zeidner, 1991) , and the Survey of Attitudes Toward Statistics (SATS-28© and SATS-36©) (Schau, Stevens, Dauphinee, & Del Vecchio, 1995; www.flaguide.org/tools/attitude/attitudes_toward_statistics.php).

Survey of Attitudes toward Statistics, SATS, is one of the most widely used instruments in the literature to assess attitudes towards statistics. This instrument is discussed in the following section.

Survey of Attitudes toward Statistics (SATS).

The development of each component of SATS was based on educational theories. These theories included expectancy-value (Atkinson, 1957; Eccles & Wigfield, 1995), attribution (Weiner, 1979), social cognition (Bandura, 1977), and goal theories (Maehr, 1984). Three expectancy-value factors of importance in statistics education include expectancy for success, task difficulty, and task value (Schau, 2003). Expectancy for success is related to students' beliefs that they can successfully do statistics. Task difficulty is associated with students' perceptions of the complexity of statistics, and relevance of performing statistics successfully is defined as task value (Schau, 2003).

SATS-28© (Schau et al., 1995) is a 28-item instrument that focuses on four aspects of student attitudes toward statistics. These include Affect, Cognitive Competence, Value, and Difficulty. The Affect component of the instrument measures the positive and negative feelings concerning statistics. Cognitive Competence concerns attitudes about intellectual knowledge and skills that are needed for statistics. Value focuses on attitudes about the usefulness, relevance, and worth of statistics in personal and professional life. Difficulty pertains to attitudes about the difficulty of statistics as a subject.

In the most recent version of SATS, SATS-36©, a 36-item instrument, consists of six components of students' attitudes toward statistics. In addition to the four components already mentioned in SATS-28©, Interest and Effort are also assessed. Pre-course and post-course versions of SATS-36© are available at www.flaguide.org/tools/attitude/attitudes_toward_statistics.php. More information concerning SATS-36© is discussed in Chapter 3.

According to Schau (2003), scores for the four components in SATS-28© have the same meaning for both genders at pre-course and post-course. These four components of SATS-28© strongly correlated with Wise's ATS course scale, and SATS-28© value component was highly correlated with ATS Field scale (Schau et al., 1995).

SATS-28 © is the only statistics attitude instrument with gender equivalence (Hilton, Schau, & Olsen, 2004). This instrument was found to be factorially invariant across gender, administration time, and for the gender-administration time interaction with respect to factor loadings and factor correlations. “These invariance results suggest that mean differences by gender, by administration time, and by their interaction in the SATS’ components can be examined meaningfully” (Hilton et al., 2004, p.102).

Greater variation in attitudes toward statistics was observed post-course as compared to pre-course, but this result was not unusual since variances “usually increase across a period of growth as should occur in introductory statistics course” (Hilton et al., 2004, p. 103). Correlation between SATS components Affect and Cognitive Competence was 0.94 which may suggest redundancy due to the large value of this correlation coefficient. Hilton et al. (2004) believed that Cognitive Competence and Affect represented different attitude constructs since Cognitive Competence was strongly positively correlated to successful completion of introductory statistics for both males and females, and Affect was unrelated to successful completion of introductory statistics for females and weakly related for males.

SATS-28© was used to assess the attitudes toward statistics of Arabic speaking pre-service teachers enrolled in an introductory statistics course (Nasser, 2004). The participants were predominantly female with strong mathematical abilities. Results indicated that course achievement improved with stronger mathematical aptitude and more positive attitudes toward statistics. Attitudes toward statistics were more positive when motivation to succeed was stronger and mathematics anxiety was lower. Mathematical aptitude, mathematics anxiety, attitudes toward statistics, attitudes toward mathematics, and motivation accounted for 36% of the variance in achievement in an introductory statistics course. Based on the results, the best predictor of course achievement was mathematical aptitude which in turn was related to more positive attitudes toward statistics. To measure course achievement, a limited number of open-ended questions were used. Nasser (2004) recommended the use of multiple methods of assessment to evaluate course achievement.

Schau (2003) found that after taking a traditional statistics course attitudes were more polarized than before taking the course. Students' perceived value of statistics decreased after taking the traditional course. In addition, Schau (2003) found that attitudes toward statistics expressed verbally were more negative than responses to the SATS instrument. Students believed that their attitudes toward statistics were a result of their achievement and instructors. Attitudes toward statistics and achievement were positively correlated.

Previous experience in mathematics, gender, cognitive and affective factors as well as effort have been reported in the literature as having an effect on attitudes towards statistics and/or course performance. Currently, no research has focused on the effects of these variables on statistical literacy, but course performance may be related to statistical literacy.

Effect of Previous Mathematics Courses.

Attitudes toward mathematics may influence attitudes toward statistics. Affective reactions to experiences in mathematics courses may affect how students relate to learning statistics (Gal & Ginsburg, 1994). Students expect that statistics will include complex mathematical formulas (Simon & Bruce, 1991). Roberts and Saxe (1982) reported that students who performed better on a basic mathematics test and took a greater number of mathematics courses had more positive attitudes toward statistics. Lalonde and Gardner (1993) found that mathematical background, defined as highest grade level of mathematics attained, and basic mathematical ability, based on a 10-item test, correlated with exam performance in a statistics course. Male students had taken more mathematics courses than female students and scored higher than females on quizzes (Lalonde & Gardner, 1993). Other researchers including Adams and Holcomb (1986) and Feinberg and Halperin (1978) found that basic mathematical ability and statistics course performance were positively related. Using a math skills test as a covariate, initial math ability and attitudes were important to successful course completion (Harlow, Burkholder, & Morrow, 2002).

Gender Effects

Many studies have investigated the effect of gender on attitudes toward statistics. In several studies, males were found to have more positive attitudes toward statistics than females (Auzmendi, 1991; Roberts & Bilderback, 1980; Waters et al., 1988) although the differences were small. Other research indicated no differences in attitudes toward statistics for males and females (Cherian & Glencross, 1997; Faghihi & Rakow, 1995; Schau, Dauphinee & Del Vecchio, 1992; Sutarso, 1992; Tomazic & Katz, 1988; Wisenbaker & Scott, 1997). Based on results using the Fenema-Sherman Mathematics Attitude Scale (Fenema & Sherman, 1977), female students had more positive attitudes toward math than males and were less likely to view mathematics as a male domain (Elmore & Vasu, 1980). Females also earned significantly higher mean number of points in the course than males. These results were in contrast to previously mentioned studies. The subjects in this study were graduate students, and Elmore and Vasu (1980, p. 221) believed that the females in the study may have been more "highly motivated and success-oriented" than females in general and that may explain the study's results.

Effects of Cognitive and Affective Factors

Cognitive factors such as mathematical ability, mathematical background, and cognitive dimensions of attitudes toward mathematics and statistics have been found to be related to statistics course performance. Affective factors such as mathematics and statistics anxiety, motivation, and affective dimensions of attitudes toward mathematics and statistics have also been found to affect statistics course performance (Feinberg & Halperin, 1978; Nasser, 1999**).** Post-course attitudes toward statistics were a predictor of course achievement but pre-course attitudes were not (Wisenbaker, Scott & Nasser, 2000).

In Feinberg and Halperin (1978), the sample consisted of 278 students enrolled in an introductory statistics course jointly offered by the Psychology and Education Departments at Syracuse University. Males and females were equally represented in this sample. Students' attitudes, perception of mathematical ability, anxiety, and basic mathematics achievement were monitored four times during the semester, at the beginning, end, and two times in between. Since students were told that participation would not affect their grade and instruments were self administered, some students might not have been serious about completing the instruments.

The sample in Nasser (1999) consisted of 169 Arab college students who were enrolled in statistics course for education majors in an Israeli university. A vast majority of students, 96%, were female. This sample is very restrictive and can only be generalized to a specific population. Similarly, the sample in Wisenbaker et al. (2000) consisted of Arab students at an Israel university. In addition to the 111 Arab students, 136 Americans students who attended a small public college in the United States were included in the sample. SATS-28© was administered at the beginning and end of the semester. In both groups of students, the vast majority of students were female, 91% of the Arab students and 80% of the American students. The sample used in this study was not representative of both genders and may have affected results.

In Dempster and McCorry (2009), attitudes toward statistics of 82 undergraduate psychology majors of which 78% were female were assessed at the beginning and end of a four-course statistics sequence. Prior experiences in mathematics and statistics courses and students' attitudes toward statistics measured using SATS-28© were important predictors of course performance. Cognitive Competence scores were more strongly related to performance than previous experiences with mathematics and statistics courses. These results may have been obtained since attitude assessment was measured at the beginning and end of a two year time period. Students who were in the sample "survived" the four-course sequence and as a result may have more positive attitudes toward statistics. This time period may also have eliminated those students who did not perform well in the courses and not have positive attitudes toward statistics.

Effect of Effort

In Lalonde and Gardner (1993), effort, defined as the number of completed assignments, was found to have been affected by attitudes toward statistics. Effort was also found to affect course achievement. Neither mathematical background nor basic mathematical ability correlated significantly with effort (Lalonde & Gardner, 1993).

In Bude′, Van de wiel, Imbos, Candel, Broers, & Berger (2007), effort was determined based on class attendance and tutors' perceptions of how well students were prepared and involved in tutorial group meetings. Results from Bude′ et al. (2007) indicated that effort did not have a significant effect on achievement, but Affect positively influenced effort. These results provided the motivation for collecting information concerning students' effort and influenced how effort was defined in this research.

Summary

Statistical literacy involves critical evaluation and understanding of statistics encountered in everyday life. To understand and evaluate statistics within a context involves understanding the integration of many important statistical concepts such as choice of sampling method, size of sample, variable definitions, and sources of potential bias to name a few. To fully evaluate and understand statistics in context involves questioning these statistical concepts. The questions that were asked when confronted with statistical messages in the media were the focus of this research.

Although several models for statistical literacy were discussed in this chapter, the Hierarchical Model (Watson, 1997) had the most influence on this research. This model involves three tiers. Progression from one tier to the next tier occurs in a sequential manner. The first tier of statistical literacy involves basic understanding of statistical terminology such as sample and statistic. Understanding this terminology and making decisions and conclusions concerning statistics within a social context is the focus of the second tier of the Hierarchical Model. The third tier, the highest level of statistical literacy, involves questioning statistics. This tier was the focus of this research.

In addition, Utts (2002) was a major influence on this research, specifically the development of the Statistical Literacy Components Rubric. The eight components of statistical literacy were based on Utts (2002).

Research from the literature indicated the effect of gender, attitude toward statistics, background attributes, and effort on course achievement. These variables might also have an effect on statistical literacy and are therefore investigated in this research.

In the next chapter, the methods used for each research question in the current study are discussed. A description of the statistical analyses used for each research question is also provided.

CHAPTER THREE

METHODS

Introduction

In this chapter, the setting, sample, and research methods are described. For each research question, independent and dependent variables are identified, and procedures for obtaining these variables are discussed. Finally, the techniques used to analyze the data from each research question are explained.

Setting and Sample

The focus of this study was to understand the effect of a statistical literacy course on the challenges students made to statistical claims in the media. Data were collected from undergraduate students at a public university in the southeastern region of the United States who were enrolled in any of four sections of Statistics in Everyday Life 200 (SIEL) taught during the spring 2009 semester. Three of the four sections were taught by the same instructor. There were no prerequisites for the course, and students were unable to receive credit for the course if they had already received credit for a higher level statistics course although about 41% of students had taken a statistics course in high school and about 17% had taken one in college prior to this course. Approval for this study was granted by the Institutional Review Board at the University, and student participation was voluntary.

One hundred forty-four of the 195 students (73.85%) who completed the course also completed all components of this study. Majors such as Psychology, Communications, Political Science, Business, General Engineering, History, Visual Arts, English, and Biological Sciences enrolled in this course and participated in this study. Table 3.1 provides a comparison of demographic variables for the study's sample and the University in 2009.

The University from which the sample was obtained is a land grant university which has strong programs in engineering, mathematics, and sciences. Approximately 19,000 students attend this institution.

There were differences between the sample and the University. The majority of students in the sample were female while the majority of students at the University were male. In the sample, freshmen made up the highest class percentage. Although percentages of each classification at the University were about the same, seniors comprised a slightly higher percentage. Most students in the sample had majors in Arts, Architecture, and Humanities *and* Business and Behavioral Sciences while at the University Business and Behavioral Sciences *and* Engineering and Sciences had the highest percentage of majors.

Table 3.1: Demographic comparison of the study sample and university

	Sample	University
Female	58.33%	45.6%
Male	41.67%	54.4%
Freshmen	41.67%	23.24%
Sophomore	36.11%	24.2%
Junior	13.89%	22.98%
Senior	8.33%	29.58%
Arts, Architecture & Humanities	31.94%	13.55%
Agriculture and Life Sciences	13.19%	19.68%
Business and	28.47%	26.76%

Behavioral Sciences		
Engineering and Sciences	16.67%	26.49%
Health, Education & Human Development	9.72%	13.52%

Table 3.2 provides a summary of averages and standard deviations for demographic variables for the 144 participants in this study. Students in this study were similar to students from the University as a whole. In 2007, the average Verbal SAT score for the University's freshmen students was 595 which is the same average as observed for this study's participants. In 2007, the average Math SAT score for the University's freshmen was 626, a slightly higher average than that observed for this study's participants. The reported Total SAT score in this study was 1246 which was slightly higher than the University's total SAT average of 1221. The inconsistency in the reported Math and Verbal SAT scores not summing to the reported Total SAT score may be attributed to self-reporting test scores or reporting only the Total SAT and not individual scores or vice versa.

Table 3.2: Averages and standard deviations for demographic variables of participants

Demographic variable	**Average**	**Standard deviation**
Reported Math SAT score	618	79
Reported Verbal SAT score	595	84
Reported Total SAT score	1246	203
Number of high school math courses taken	4.4	1.1
Number of high school statistics courses taken	0.43	0.5
Number of college math courses taken	1.3	1.0

Number of college statistics courses taken	0.26	0.8
Cumulative attempted college credits	28.2	26
Cumulative earned college credits	41.4	28.7
Cumulative college points	86.3	82.9
Cumulative GPA	2.90	1.0
Reported GPA	3.14	0.7

Since the University recommends four units of mathematics for incoming students, it is not surprising that the average number of high school mathematics courses taken was 4.4. The average number of high school and college statistics courses taken was less than one. Although there was a discrepancy between the cumulative grade point average (GPA) obtained from the University's database and student-reported GPA, this may be explained by the exclusion of transfer students' GPA's from the University's data base. Cumulative attempted, earned, and college points were collected from the sample because transfer students and first semester students did not have a GPA in the University's data base and inclusion of these variables was a means to measure past coursework experience.

Procedure

Overview

On the first and last days of classes during the spring 2009 semester, students enrolled in SIEL were asked to complete the Survey of Attitudes toward Statistics (SATS-36©) developed by Schau et al. (1995). This instrument assessed attitudes toward statistics and included a section of demographic items. Since the consent form was attached to SATS-36©, students' responses were matched pre-course and post-course. The reliability and validity of this instrument are discussed later in this chapter.

On the second day of classes and on the final exam, students were asked to provide questions (challenges) concerning statistical claims made in two advertisements and an article and to explain the importance of their questions. These media articles can be found in Appendix A. No time limit was given for students to complete this assignment within the constraints of class time. The same two advertisements and article were given to students at the beginning and end of the course, and the order of these media articles was randomized to avoid potential order bias. The same order of media articles was given to each student prior to and after SIEL. Responses were matched by students to evaluate response changes during the semester.

At the end of the semester, students completed a reflection paper (Appendix B) that accounted for five percent of their final course average. The reflection paper consisted of open-ended questions that pertained to students' experiences in the course, their current attitudes toward statistics, and their prior experiences in mathematics and statistics courses.

A description of the course, Statistics in Everyday Life 200 (SIEL), is provided in Chapter 1. More detailed information concerning this course can be obtained by contacting the author.

To address the research questions, quantitative and qualitative data were collected and analyzed. Quantitative data consisted of data from an attitude assessment instrument, SATS-36©, and scores from the Statistical Literacy Components Rubric (SLCR), a rubric developed to assess levels of awareness of statistical literacy components (Appendix C). Students' challenges to the two advertisements and article and responses from reflection paper questions comprised the qualitative data. These challenges were grouped into topic categories prior to and after taking SIEL, and patterns from reflection paper questions were observed.

In the following sections, the data collected and analyses performed for each research question are discussed. Details concerning instruments used are also presented.

Data Collection for Research Question 1

Research Question 1 focused on the level of awareness of statistical literacy components demonstrated by students prior to and after a course on statistical literacy (SIEL). In addition, the effect of gender, attitude, aptitude, and background attributes upon the level of awareness of statistical literacy components (pre-course, post-course and changes from pre-course to post-course) were investigated. The dependent variables consisted of statistical literacy component scores obtained from SLCR for each of eight identified components of statistical literacy. The independent variables included gender, aptitude, background attributes, and attitude component scores from SATS-36©. In the next sections, the independent variables are discussed followed by a discussion of the dependent variables.

Attitudes toward Statistics

Students completed SATS-36©, a valid and reliable instrument, that assessed students' attitudes toward statistics pre-course and post-course (Schau et al., 1995, Schau, 2003, Hilton et al., 2004.) Two versions of the survey were administered; one assessed attitudes prior to the course and the other after the course. The only difference between the two versions was tense, with the pre-test asking students to indicate what they expected from the course where appropriate, and the post-test indicating what they experienced in the course. Both versions of the survey consisted of 36 items and included a 7-point Likert scale with response options that ranged from "strongly disagree" to "strongly agree."

SATS-36© assesses six attitude components: Affect, Cognitive Competence, Value, Difficulty, Interest, and Effort. A description of each attitude component, survey items pertaining to each component and Cronbach's alpha for each attitude component based on previous research are presented in Table 3.3 (Schau, http://www.evaluationandstatistics.com/Final36scoring.pdf). Since Interest and Effort were recently added to the instrument, Cronbach's alpha values were not reported for these components.

Table 3.3: A summary of information concerning attitude components of the Survey of Attitudes toward Statistics (http://www.evaluationandstatistics.com/Final36scoring.pdf)

Attitude Component	Description	Pre-SATS Items (Item #)	Cronbach's alpha
Affect	Students' feelings about statistics	I will like statistics. (3) I will feel insecure when I have to do statistics problems. (4) I will get frustrated going over statistics tests in class. (15) I will be under stress during statistics class. (18) I will enjoy taking statistics courses. (19) I am scared by statistics. (28)	0.8 to 0.89
Cognitive Competence	Students' attitudes about their intellectual knowledge and skills when applied to statistics	I will have trouble understanding statistics because of how I think. (5) I will have no idea of what's going on in this statistics course. (11) I will make a lot of math errors in statistics. (26) I can learn statistics. (31) I will understand statistics equations. (32) I will find it difficult to understand statistical concepts. (35)	0.77 to 0.88
Value	Students' attitudes about the usefulness, relevance and worth of statistics in personal and professional life	Statistics is worthless. (7) Statistics should be a required part of my professional training. (9) Statistical skills will make me more employable. (10) Statistics is not useful to the typical professional. (13) Statistical thinking is not applicable in my life outside my job. (16) I use statistics in my everyday life. (17) Statistics conclusions are rarely presented in everyday life. (21) I will have no application for statistics in my profession. (25) Statistics is irrelevant in my life. (33)	0.74 to 0.90
Difficulty	Students' attitudes about the difficulty of statistics as a subject	Statistics formulas are easy to understand. (6) Statistics is a complicated subject. (8) Statistics is a subject quickly learned by most people. (22) Learning statistics requires a great deal of discipline. (24) Statistics involves massive computations. (30) Statistics is highly technical. (34) Most people have to learn a new way of thinking to do statistics. (36)	0.64 to 0.81

Interest	Students' level of individual interest in statistics	I am interested in being able to communicate statistical information to others. (12) I am interested in using statistics. (20) I am interested in understanding statistical information. (23) I am interested in learning statistics. (29)	New component
Effort	Amount of work needed to learn statistics	I plan to complete all of my statistics assignments. (1) I plan to work hard in my statistics course. (2) I plan to study hard for every statistics test. (14) I plan to attend every statistics class session. (27)	New component

Scores for each of the six attitude components of SATS-36© were obtained by first reverse coding negatively worded items (items 4, 5, 7, 8, 11, 13, 15, 16, 18, 21, 24, 25, 26, 28, 30, 33, 34, 35, and 36). Item responses within each component were summed and then divided by the number of items within each component to obtain a mean component score for each student. Higher component scores corresponded to more positive attitudes with respect to each attitude component (Schau, http://www.evaluationandstatistics.com/Final36scoring.pdf.) For example, a higher score on the Difficulty component indicates that students believe statistics is easy. Pre-course and post-course component scores were calculated for each of the six attitude components, and these scores were matched by student to determine change in attitude scores.

Other Independent Variables

For Research Question 1, several other independent variables besides attitudes were obtained, including gender, aptitude, and background attributes. Aptitude consisted of student reported Math SAT, Verbal SAT, and Total SAT scores as well as student reported grade point ratio and cumulative grade point ratio obtained from the University's data base. According to Gal (2004), literacy skills are needed for statistical literacy and can affect statistical literacy. Due to this relationship between general literacy and statistical literacy, Verbal SAT scores were collected. Background attributes obtained from SATS-36© included the total number of high school and college mathematics courses and the total number of high school and college statistics courses that were taken. Since some students did not have grade point ratios from the University because they were transfer students or first semester freshmen, cumulative attempted credits, cumulative earned credits and cumulative points were obtained from the University's data base.

Assessing the Level of Awareness of Statistical Literacy Components

The level of awareness of statistical literacy components prior to and after SIEL was assessed from students' challenges to two advertisements and an article. For the advertisements and article, students were asked to provide questions concerning the information or claims made and explain why each question was important to ask. All students' responses were transcribed into Microsoft WORD® documents for further analysis.

Advertisements were used to assess statistical literacy not only because they represent real-world examples of the use of statistics but because there is little text which should minimize the effect that a lower level of general literacy might have on a measure of statistical literacy (Gal, 2004). The advertisements were chosen to illustrate both a categorical measure (LiveActive advertisement) and a numerical measure (Allstate advertisement). Although the article pertaining to depression required general literacy skills, it was included to provide a range of media examples, and the general literacy skills required for this article would not exceed that which could be reasonably expected of college students.

Students' challenges were evaluated using the Statistical Literacy Components Rubric (SLCR) to assess the level of students' awareness to eight components of statistical literacy. These components of statistical literacy were identified based on Utts (2002), past experiences from teaching the course and consultation with Dr. William C. Bridges, the co-developer of the course. These statistical literacy components concerned Method, Bias, Reported statistic, Definitions, Variation, Generalize, Lurking variable, and Causality.

The statistical literacy component Method involved how the study was conducted, how the information in the media article was obtained, what sampling method was used, what type of study (observational versus experiment) was performed, what sample size was used, and what the response rate was. The Bias component referred to who conducted the study, the agenda of those who conducted the study, the manner in which responses were obtained (e.g. voluntary response), the wording of question(s) that were asked of respondents, and issues with coverage error. Questions concerning Reported statistic included how the statistic was obtained, whether the value in the article was a statistic or parameter, and if the statistic was appropriate for the situation. The statistical literacy component Definition involved how variables were defined, if there were alternate definitions, if what was defined could actually be measured, and if the definition gave reliable results. Questions concerning Variation pertained to the margin of error or confidence interval, variation associated with statistics, and to the variability of statistics from sample to sample. The Generalize component included whether information was generalized to a population and to what population the generalization would apply. The Lurking variable component pertained to concerns about information not included and additional factors that might affect results. Questions about Causality included concerns about one factor causing another and whether causation was implied but not necessarily justified.

Using SLCR, students' levels of awareness of each statistical literacy component were assessed using a 4-point scale. A score of 0 indicated no awareness of the statistical literacy component, a score of 1 represented basic or minimal awareness, a score of 2 corresponded to moderate awareness, and a score of 3 indicated advanced awareness.

An example of scoring for the statistical literacy component "questions concerning the method" is provided. If a student did not question how the study was conducted, a score of 0 was assigned to this component. If a student asked "How was the study conducted?" but asked no further questions concerning the method, a score of 1 was given for this component. If a student asked "How was the study conducted?" and then inquired about the sampling method used, a score of 2 was assigned to this component. If a student asked "How was the study conducted?" and then wondered if the selection of units was random because that would affect the results, then a score of 3 was given for this component. Each response to the articles was scored in this manner, and the mean score was determined for each component across articles to produce an average pre-course, average post-course and average change for each statistical literacy component. More detailed examples for each article are included in Chapter 4.

Statistical Analyses for Research Question 1

Descriptive statistics were calculated for all variables using SAS version 9.2 (Cary, NC). Cronbach's alpha was determined for each SATS-36© attitude component score. Correlation analysis was performed on dependent and independent variables to determine if there were significant linear relationships. The following are the hypotheses that were tested:

$$H_0 : \rho = 0$$
$$H_a : \rho \neq 0$$

where ρ represented the true strength of a linear relationship between the two variables.

For correlation analyses, a significance level of 0.05 was used but because many correlation hypothesis tests were performed the Bonferroni Correction was used to ensure an overall probability of a Type I error to be close to 0.05. Statistical significance using the Bonferroni correction was α/n where α was set to 0.05 and n represented the number of correlation hypothesis tests that were performed (Weisstein, http://mathworld.wolfram.com/BonferroniCorrection.html).

Inter-rater reliability, measured by the Kappa statistic, was determined for SLCR by first randomly selecting 25 of the 144 students in the study. Determination of the sample size was based on procedures described in Cantor (1996). Using SLCR, two raters independently assessed these students' pre-course and post-course responses to the two advertisements and article. In total, 150 documents were evaluated by each rater. Prevalence index, bias index, the simple Kappa statistics, and 95% confidence intervals for Kappa were determined based on the methods described in Sim and Wright (2005) for responses that were categorized as presence or absence of each SLCR component pre-course and post-course. Using SAS version 9.2 (Cary, NC), weighted Kappa statistics and 95% confidence intervals were determined for evaluations of pre-course and post-course responses using the 4-point scale in SLCR.

Regression analysis was performed using SAS version 9.2 in order to understand the relationship between average statistical literacy component scores and gender, aptitude, attitudes toward statistics, and background attributes. The forward selection procedure in SAS version 9.2 was used to determine the best model for predicting average pre-course SLCR component scores, average post-course SLCR component scores and changes in average SLCR component scores. Normality of residuals, an assumption for regression analysis, was tested. Independent variables that had a significant impact on the dependent variables were determined at the 10% level of significance. Since this aspect of the study was exploratory, a significance level of 10% instead of 5% was used because of concerns about possible Type II errors.

Because statistical literacy component scores may not be independent from each other, a multivariate model was created using all significant independent variables from the forward selection procedure. The Wilk's lambda test was performed to determine the significance of each independent variable. Statistical significance was determined at the 10% level.

Data Collection for Research Question 2

Research Question 2 focused on determining what challenges students made concerning statistical claims made in the two advertisements and article. Students' responses were grouped based on similar topics using QSR Nvivo 8, and matrix coding resulted in an Excel file that included each student's name and the presence or absence of a response in each topic category indicated by a 1 for presence or a 0 for absence. Students' pre-course and post-course responses were coded in this manner and matched by student. Figure 3.1 provides an example that illustrates how data were summarized for topic category, "Generalize."

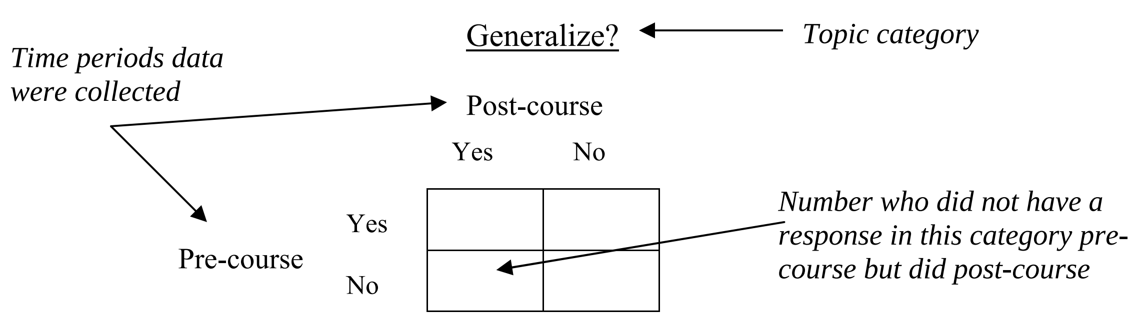


Figure 3.1: Data structure for Research Question 2

In addition, differences in topic categories for males and females were investigated. For this research question, the dependent variables were measured as the presence or absence of a student's response in each topic category prior to and after SIEL, and the independent variable was sex.

Statistical Analyses for Research Question 2

Research Question 2 focused on determining areas of concern that were raised for statistical claims made in the media, and several hypothesis tests were conducted. For all hypothesis tests, SAS version 9.2 was used, and statistical significance was determined at the 5% level.

For each media article, McNemar's tests were conducted to determine if the pattern of responses for topic categories had changed. The following hypotheses were tested:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

where p_{01} represented the proportion of students who did not have a response in the topic category pre-course and had a response in the topic category post-course, and p_{10} represented the proportion of students who had a response in the topic category pre-course and did not have a response in the topic category post-course (Levin & Serlin, 2000). McNemar's test was an appropriate test to use since the presence/absence of post-course topic categories was not independent of the presence/absence of pre-course topic categories.

In addition, hypothesis tests were also performed to determine if there was a difference in the proportions of males and females who had concerns about a topic category. This hypothesis test was performed for each topic category prior to and after

SIEL. McNemar's tests were also performed to compare differences in response patterns for males and females.

Data Collection for Research Question 3

Research Question 3 focused on determining what, if any, trend(s) related to the experience of taking a course focused on statistical literacy emerged from responses to a reflection paper. The purpose of this reflection paper was to understand students' experiences in a course focused on statistical literacy and past experiences with mathematics and statistics courses.

Students' responses were grouped based on gender and Effort levels. Effort was determined based on the percentage of class activities completed during the semester. The Effort levels were determined as follows: "low effort" (less than 70% activities completed), "moderate effort" (70% or more but less than 90% activities completed), and "high effort" (at least 90% activities completed).

Statistical Analyses for Research Question 3

The responses from randomly selected reflection papers were analyzed using QSR Nvivo 8. Sixty of the 144 reflection papers were randomly selected using proportional allocation to reflect the following percentages based on sex and effort levels for all students in the study: 36.1% female/high effort, 16.7% female/moderate effort, 5.6% female/low effort, 22.2% male/high effort, 15.3% male/moderate effort, and 4.2% male/low effort. The sample size of 60 was determined using the method defined by Cochran (1977, p. 110-111). The formulas used and the values used in the formulas were

$$n_0 = \frac{(pq)(z_\alpha)^2}{MOE^2} = \frac{(0.5 * 0.5)(2)^2}{(0.1)^2} = 100$$

$$n = \frac{n_0}{1 + \left(n_0 / N \right)} = \frac{100}{1 + \left(100 / 144 \right)} \approx 60$$

where n_0 represented an approximation of the sub-sample size ignoring the finite population correction, n represented the sample size when the finite population correction was considered, N represented the total sample size from which sub-sampling occurred, MOE represented the margin of error, p represented the probability a person belonged in a particular sex/effort group, and q represented the probability a person was not in the sex/effort group, (1-p). The values used in this formula were chosen to produce a conservative estimate of the sample size (i.e. to ensure that the sample size, n, was large enough, with the possibility of being slightly larger than needed). A value of p equal to 0.5 which resulted in a q of 0.5 was selected; Z_α was rounded from 1.96 to 2 to achieve a conservative sample size. The MOE was chosen as 0.1 to ensure our estimate was most likely within 10% of the true population value. Procedure *surveyselect* in SAS version 9.2 was used to determine which 60 students were selected.

Differences in experiences based on sex and Effort classifications were determined using Chi-square tests or Fishers exact tests. For Research Question 3, significance was determined at the 5% level. The hypotheses and statistical analyses performed for each question from the reflection paper are discussed in the paragraphs below. Fisher's Exact test was performed instead of the Chi-square test when more than

20% of expected counts were fewer than 5 and/or when all individual expected counts were not at least one (Baldi & Moore, 2009).

Students were instructed to "Please describe past experiences in other mathematics or statistics courses that you have taken." Responses were classified into four categories based on interpretation of students' reported view of past experiences. These categories of past experiences included: positive view, negative view, mixed view, or indifferent view. A Chi-square test was performed to determine if at least one of the proportions of responses were different for the categories of view of past experiences. Statistical significance was reported at the 5% level.

To determine if there was an association between sex and view of past experience, Fisher's Exact test was performed. Similarly, Fisher's Exact test was performed to determine if there was an association between effort level and view of past experience.

Students were asked to "Explain how your experiences in SIEL were similar to your past experiences in mathematics or statistics courses." Responses were classified into five categories: anxiety, course content, course format, problem solving nature, and no similarities. A Chi-square test was performed to test if at least one of the proportions of responses was different for the category of similarities with other courses.

Fisher's Exact test was performed to determine if there was an association between sex and categories of similarity with other courses and SIEL. Similarly, Fisher's Exact test was performed to determine if there was an association between effort level and similarities with other courses.

Students were also asked to "Explain how your experiences in the course were different from your past experiences with other mathematics or statistics courses." For students who had taken a previous statistics courses, differences were classified into three categories: analysis, application to everyday life, and course format. The fourth category represented a student's first statistic course. A Chi-square test was performed to determine if at least one of the proportions of responses was different for the category of differences with other courses. To determine if there was an association between sex and differences between this course and other courses, a Fisher's exact test was performed. In addition, to determine if there was an association between effort levels and differences between SIEL and other courses a Fisher's Exact test was performed.

Students were asked to explain what "statistical literacy" meant to them. Based on their explanations, their responses were classified into two categories: whether they had a basic understanding of the concept of statistical literacy or not. A response was put into the category for basic understanding of statistical literacy if the student said that statistical literacy was more than just computing statistics or producing graphs but related to understanding, questioning, and/or interpretation of statistics encountered in everyday life. Since this was an open-ended question and responses to this question lacked depth, only two categories of responses were used. A Chi-square test for equal proportions was performed to determine if the proportion that had a basic understanding of statistically literacy was different from the proportion that did not have a basic understanding of statistically literacy. Ninety-five percent confidence intervals were also constructed.

A Chi-square test was performed to determine if there was an association between sex and understanding statistical literacy. Fisher's Exact test was performed to determine if there was an association between effort level and understanding statistical literacy.

Students were asked if they felt that they were "statistically literate." Responses were categorized as "yes" or "no." A Chi-square test was performed to determine if the proportion who believed they were statistically literate was different from the proportion who did not believe they were statistically literate. Ninety-five percent confidence intervals were also constructed. Fisher's Exact tests were performed to determine if there was an association between sex and being statistically literate, and if there was an association between effort level and being statistically literate.

Students were asked if the course affected the way they looked at statistics they encountered in everyday life. Responses were classified as "yes" or "no." A Chi-square test was performed to determine if the proportion who believed the course affected the way in which they looked at statistics encountered in everyday life was different from the proportion who did not believe the course affected the way in which they looked at statistics encountered in everyday life. Ninety-five percent confidence intervals were also constructed. Fisher's Exact tests were performed to determine if there was an association between sex and whether the course changed the way students looked at statistics in everyday life and if there was an association between effort level and whether the course changed the way students looked at statistics in everyday life.

Students were asked to identify attributes they believed were necessary to be successful in the course. Overall percentages were computed for each attribute. Attributes were then classified into four attribute categories: Abilities, Class Behavior, Student Quality and Thinking. The “Abilities” attribute included “accept past mistakes,” “apply definitions,” “do basic math or math skills,” “learn definitions,” “learn or use formulas,” “memorize,” “understand and apply concepts,” “understand basics,” “understand big picture,” “understand graphs,” “understand marketing and advertising,” “understand statistics before course,” and “understand theories.” The “Class Behavior” attribute included “ask questions in class,” “attend class,” “do homework or practice,” “participate in activities or class,” “pay attention in class,” “print notes for class,” “study,” “ take good notes,” and “work well in groups.” The “Student Quality” attribute included “apply yourself,” “appreciation for statistics,” “excitement about course,” “be prepared or willing to learn new things,” “conscientiousness,” “creativity,” “dedication,” “desire,” “determination or drive,” “do not underestimate course,” “do not need to be good or interested in statistics,” “effort,” “good listeners,” “intelligence,” “interest,” “like math,” “motivation,” “observant,” “open minded,” “organized,” “patience,” “perceptive,” “responsibility,” “self-control,” “serious attitude,” “visual learner,” “want to learn,” and “work hard or work ethic.” The “Think” attribute included “analytical,” “common sense,” “comprehend word problems,” “different way of thinking,” “do not think analytically,” “good memory,” “problem solving or logical thinking skills,” “question statistics,” and “think on your feet.”

For each attribute category, a Chi-square test was performed to determine if the proportion who said attribute category was necessary for course success was different from the proportion who did not say attribute category was necessary for course success.

Ninety-five percent confidence intervals were also constructed. Fisher's Exact tests were performed for each attribute category to determine if there was an association between sex and whether students thought the attribute category was necessary for course success. In addition, Fisher's Exact tests were performed for each attribute category to determine if there was an association between effort level and whether students thought the attribute category was necessary for course success.

Students were asked if the course challenged their beliefs. Responses were categorized as "yes" or "no." A Chi-square test was performed to test if the proportion who said their beliefs were challenged by the course was different from the proportion who said their beliefs were not challenged by the course. Fisher's Exact tests were performed to determine if there was an association between sex and whether students' beliefs were challenged during the course and between effort level and whether students' beliefs were challenged during the course.

Students were asked to give three adjectives which they would use to describe "statistics." Percentages were determined overall and by sex and effort level. The highest percentages within each group were observed.

Students were asked to describe their attitude toward statistics as "good," "moderate," or "poor." A Chi-square test was performed to determine if at least one of the proportions of responses was different for each attitude classification. Fisher's Exact tests were performed to determine if there was an association between sex and student attitude classification and effort level and student attitude classification.

Students were asked if they felt their attitudes toward statistics had changed as a result of this course. Responses were classified as "yes" or "no." A Chi-square test was performed to determine if the proportion of students who said that their attitudes changed because of the course was different from the proportion of students who said that their attitudes did not change because of the course. A Chi-square test was performed to determine if there was an association between sex and whether attitudes had been changed because of the course. Fisher's Exact test was conducted to determine if there is an association between effort level and whether attitudes had been changed because of the course.

For those students who said that their attitude had not changed as a result of the course, their attitude was classified as "negative," "positive," or "neutral." A Chi-square test was performed to determine if at least one of the proportions of responses was different for the attitude classifications. Fisher's Exact test was performed to determine if there was an association between sex and attitude classifications for those who said their attitudes had not changed as a result of the course. Also, a Fisher's Exact test was performed to determine if there was an association between effort level and attitude classifications for those who said their attitudes had not changed as a result of the course.

For those students who said that their attitude had changed as a result of the course, their change in attitudes was categorized as "negative to positive" and "neutral to positive." A Chi-square test was performed to determine if the proportion with a change in attitude from negative to positive was different from the proportion with a change in attitude from neutral to positive. Fisher's Exact tests were performed to determine if there was an association between sex and change in attitude as a result of the course and to determine if there was an association between effort level and change in attitude as a result of the course.

In the next chapter, the results for each research question are provided. Discussions of results and summaries are also included.

CHAPTER FOUR

RESULTS AND DISCUSSION

To provide context to the discussion of the research results that follow, course demographics and their correlations are presented first. Following this, results from the statistical analyses of each of the three research questions are presented and discussed. Correlation analyses pertaining to each research question are also included. After discussion of results for each research question, a summary of major points is presented.

Prior to discussion of research questions, course performance statistics are presented. In addition, results from correlation analyses of demographic variables are discussed.

Course Performance Demographics

Two measures of course performance, the average final course grade and course effort, were determined at the end of the semester. Based on a 4-point scale, the average final course grade was 3.09 with a standard deviation of 0.89. Course effort was determined based on the percentage of completed activities during the semester. Average course effort was 87.91%, and the standard deviation was 12.7%. In the sample, 58.33% of students were classified in the high effort category, 31.94% in the moderate effort category, and 9.72% in the low effort category.

Correlation Analyses involving Demographic Variables

In order to determine if there were significant linear relationships between demographic variables found in Table 3.2, correlation coefficients were determined, and hypothesis tests were conducted using the Bonferroni correction as described in Chapter 3. Table 1 Appendix D lists results from the correlation analyses for all demographic variables including the names of variables that were correlated, correlation coefficients, and p-values for the hypothesis tests. With 105 combinations of variables that were correlated and an alpha of 0.05, the significance level for comparisons in Table 1 Appendix D was determined as 0.0005 (0.05/105). Table 4.1 provides a summary of the statistically significant comparisons. The correlations listed in Table 4.1 were all positive.

Table 4.1: Significant* correlations for demographic variables

	Math SAT	Verbal SAT	Total SAT	# HS math courses	# HS statistics courses	# college math courses	# college statistics courses	Cumulative attempted credits	Cumulative earned credits	Cumulative points	Cumulative GPA	Course Grade	Course Effort	Reported GPA
Math SAT			0.42		0.41						0.35	0.34		
Verbal SAT			0.57											
Total SAT														
# HS math courses					0.40		0.42							
# HS statistics courses							0.31							
# college math courses							0.43	0.47	0.53	0.46				
# college statistics courses								0.35	0.45	0.33				
Cumulative attempted credits									0.84	0.96				
Cumulative earned credits										0.84				
Cumulative points														
Cumulative GPA												0.43	0.20	0.69
Course Grade													0.36	0.54
Course Effort														
Reported GPA														

*Significant at the $p < 0.0005$ level.

Reported Math SAT scores and reported Verbal SAT were both correlated with Total SAT score. This result would be expected since the Total SAT score is dependent on Math and Verbal SAT scores.

Reported Math SAT score and the number of high school math courses taken were correlated with a higher number of high school math courses associated with an increase in the reported Math SAT score. Reported Math SAT score was also correlated with cumulative GPA and course grade. Since statistics involves quantitative skills this might explain the relationship between reported Math SAT score and course grade. A similar relationship between Math SAT score and mathematics course grade was observed in Pugh and Lowther (2004) and between Math SAT score and grade in a statistics course in Stephens (1982).

The number of high school math courses taken was correlated with the number of high school statistics courses taken as well as the number of college statistics courses taken. This relationship is expected since those who are more inclined to take math courses may be more inclined to take statistics courses. The number of high school statistics courses taken was correlated with the number of college statistics courses taken. An explanation for this relationship may be that once in college those taking statistics in high school may be more likely to be in majors that require statistics.

The number of college math courses taken was correlated with the number of college statistics courses taken, cumulative attempted credits, cumulative earned credits, and cumulative points. Similarly, the number of college statistics courses taken was significantly correlated with cumulative attempted credits, cumulative earned credits, and cumulative points. Those who have taken more college courses are further along in their college careers, and this may explain these relationships. The correlation results involving cumulative attempted credits, cumulative earned credits, and cumulative points would be expected since these variables are functions of each other.

Cumulative GPA was correlated with course grade and reported GPA. The relationship between cumulative GPA and course grade is understandable since those who have higher cumulative GPA would be likely to have higher course grades. This relationship was also observed in Johnson and Kuennen (2006).

Course grade was correlated with course effort and reported GPA. Since course effort was a component of the final course grade, this may explain this relationship. In Bude′ et al. (2007), similar results concerning the relationship between effort and course achievement were observed.

Results from correlation analyses of demographic variables are important for understanding the relationship between these variables. Understanding these relationships is especially valuable when developing models as described in Research Question 1.

Research Question 1

Overview

The overall purpose of Research Question 1 was to assess the level of awareness of statistical literacy components of students enrolled in a course focused on statistical literacy. This research question was divided into three more specific questions, Research Question 1a, Research Question 1b, and Research Question 1c. The focus of Research

Question 1a was to assess the level of awareness of statistical literacy components before taking Statistics in Everyday Life (SIEL). For Research Question 1b, the level of awareness of statistical literacy components was assessed after taking SIEL. Assessing changes in levels awareness of statistical literacy components by comparing pre-course and post-course statistical literacy awareness levels was the focus of Research Question 1c. Effects of gender, attitude, aptitude, and background on the level of statistical literacy were also investigated.

For Research Question 1, the level of awareness of statistical literacy components was assessed using SLCR. As discussed in Chapter 3, eight components of statistical literacy were assessed with SLCR. Statistical literacy was determined based on the level of awareness of each statistical literacy component in SLCR from students' challenges to the two advertisements and article. In the following sections, typical responses for each statistical literacy component for each media article and examples of responses for each SLCR score level are presented. In addition, results from analysis of inter-rater reliability are also discussed.

SLCR Responses

SLCR was developed to assess the level of statistical literacy for eight components of statistical literacy. These components included Bias, Causality, Definitions, Generalize, Lurking variable, Method, Reported statistic, and Variation. In this section, examples of questions associated with each statistical literacy component are presented.

The questions that students raised concerning Bias included who conducted the study, the agenda of those conducting the study, if the study involved voluntary response, the wording of questions used in the study, and coverage error concerns. For the Allstate advertisement, students had questions concerning the source of “6 million” and whether potential errors in measurement could have occurred. For the LiveActive advertisement, students were concerned about possible measurement error, the motivation for the ad, and the effect of the sensitive nature of the ad’s topic on responses. For the depression article, students were concerned about the questions asked of the respondents, motivation for the study, and how the sensitive nature of depression could affect responses.

Questions concerning Causality included concerns about one factor causing another and whether causation or correlation was implied. For the Allstate advertisement, students questioned whether there was a cause and effect relationship between having insurance and reducing accidents. For the LiveActive advertisement, questions focused on whether eating this brand of cottage cheese would cause a person to be regular. In the article on depression, students questioned whether the author was implying that a person’s sex was a cause of depression.

Questions concerning Definitions included how variables were defined, if other definitions were used, if what was defined could actually be measured, and if responses from the defined variable were reliable. For the Allstate advertisement, students had concerns about how an “accident” was defined and if the severity of the accident (fender bender versus car totaled) was considered in this definition. Students expressed concerns about definitions of “America,” “car,” and “too many.” Questions were raised concerning how an accident was counted. Was the number of accidents determined based on the number of cars in an accident or the number of accidents without regard to the number of cars involved in an accident? In addition, students questioned whether an accident was counted if it was not the driver's fault. For the LiveActive advertisement, students had concerns about the definition or type of irregularity, definition of "occasional," definition of "occasional irregularity," and what was meant by "be that other woman." For the depression article, concerns were raised about the definition of "depression," "adolescence," "anxiety," "major," and "substantial."

Questions concerning Generalize focused on whether the information in the advertisement or article could be generalized to a population and to what population the generalization would be appropriate. For the Allstate advertisement and depression article, students were concerned about the population size. For the LiveActive advertisement, students were concerned about the exclusion of men. For the depression article, students were concerned that non-teen years were included when making statements about teens. For all three media articles, students questioned whether the sample was representative.

Questions concerning Lurking variable included concerns about information that might have been excluded and additional factors that may have affected results. For the Allstate advertisement, students raised concerns about factors that could have affected the number of accidents including unreported accidents, location of accidents (rural versus urban areas), years for which information was obtained, and causes of accidents (e.g., due to inclement weather or drunk driving). For the LiveActive advertisement, students were concerned about factors that could affect irregularity such as when the study was conducted, geographic location of respondents, length of study, respondents' diets, and ages of respondents. For the depression article, concerns were raised about how depression could be affected by the geographic location of the study, when the study was conducted, the inclination for subjects to admit depression, and the effect of social pressures during teen years.

Questions concerning Method included how the study was conducted, how the information was obtained, what sampling method was used, the type of study that was conducted (observational versus experiment), the choice of sample size, and the response rate. For the Allstate advertisement, students questioned if all accidents could have actually been counted. Students were concerned about the non-response rate due to the sensitive nature of irregularity addressed in the LiveActive advertisement. For the article concerning depression, students questioned what sample size was used for each gender and the length of the study.

Questions concerning Reported statistic included such topic categories as whether the value in the media articles was a statistic or parameter and if the appropriate statistic was used. For the Allstate advertisement, students questioned if "6 million" represented an average and whether other statistics would be more appropriate to use. In addition, students had concerns about the accuracy of "6 million." For both the Allstate and LiveActive advertisements, students questioned if the value in the advertisement was an estimate as well as whether this value was relevant. For the article on depression, students asked whether the values in the article were exact numbers.

Questions concerning Variation pertained to the margin of error and confidence interval. Students also raised concerns about the variation of statistics from sample to sample. For the Allstate advertisement, students commented on the variation in the number of accidents and the potential for outliers. For the LiveActive advertisement, students had concerns if "4 out of 5" was a constant proportion. Students were concerned that neither a margin of error or confidence interval was given in the depression article.

Although the three media articles have different contexts, the questions associated with each statistical literacy component described in this section are similar in nature. In the next section, the method of rating students' responses using SLCR is described and examples of each SLCR score are presented.

Examples of Student Responses for Levels of SLCR Statistical Literacy Scores

For each component of statistical literacy, students' responses were assigned a score from 0 to 3 based on the level of awareness of each component. As described in Chapter 3, a score of 0 indicated that students did not raise any questions for the statistical literacy component. To be assigned a score level of 1, responses included questions pertaining to the statistical literacy component but did not include concerns related to their questions or why the concern was important to question. A common attribute of responses scored at a level of 2 was to provide questions related to the statistical literacy component and either concerns about this question or why the concern was important to consider. For a score of 3, students not only had questions concerning the statistical literacy component but also included how this concern might have specific effects on results or showed a deeper awareness of the component.

Tables 4.2, 4.3, and 4.4 provide specific examples of students' comments for each observed SLCR score level for the two advertisements and article. In some cases, not all score levels were observed.

Table 4.2: Examples of students' responses and SLCR scoring for the Allstate advertisement

	Examples of student comments based on SLCR scores		
Component	**1**	**2**	**3**
Bias	"Where did they get this number? Is this the number of claims each year – or is it an official number from a gov. agency (i.e. DMV)?"	"Who conducted the study to conclude that 6 million cars were in accidents every year? Was it conducted by Allstate who may have calculated bias into their argument?"	"Who took this survey? Was it an insurance agency or the government? Depending on who took it could skew the numbers, because they want them to look better in their favor or point of view."
Causality	"How is buying car insurance going to reduce the amount of accidents in America?"	"They imply that having Allstate you become a safer or better driver. But you are not able to make the connection between the cause (the way you drive) and the effect (being a better/safer) driver."	
Definition	"What describes a car accident? Are they including "fender bender" in this data?"	"What type of accident are they referring to? This is important to ask because their description of	"Do car accidents have to involve multiple cars or can it just be backing into a tree? The inclusion of all these

		accident can be very broad or it could be a more narrowed down approach and you should know what exactly they have in mind when they refer to an 'accident.'"	sorts of accidents would raise the statistic and create the large need for auto insurance that Allstate is hoping for."
Generalize	"What was the population?"	"Who is the sample and is it representative of America as a whole?"	"If the sample they used was representational of the entire population – because their information could be incorrect if their sample was not representational of the entire population."
Lurking variable	"Does that include reported, unreported or both?"	"It doesn't say where the accidents occur. Do they occur on highways or back roads etc.? Location can make all the difference in the claim."	"How many cars travel on the road each year (6 million could be a small % if 100 billion drive each year)? What were the ages of people in the accidents? Younger less experienced drivers get into more accidents which could increase accidents/yr."
Method	"It's also important to ask how they came up with this number. The reader doesn't know if it was a poll, an observational study or any other method."	"How was this study conducted? Was it observational? This is important because it is imperative that you choose the most appropriate method of study to conduct in order to arrive at	"What kind of experiment or surveys or observational methods did they perform to come up with this conclusion? This question is

		the best conclusion."	important because if they didn't go through the correct procedures as to how they came up with this number then, this advertisement is giving skewed or false numbers."
Reported statistic	"How did Allstate come up with the total number of car accidents being 6 million?"	"A third concern may be the number '6 million.' Did they find every accident recorded or did they run some sort of formula for a specific sample size. Knowing how an advertisement got its number from is a very big chance of changing information."	"Is this an average between many years? If it is an average between years, then there might have been a year that was really bad and full of car accidents which would raise the average. If it is not an average then maybe last year was an outlier and not normal and won't happen again."
Variation	"The car accident rate could have gone up or down."	"Is it a guaranteed fact that there will be 6 million car accidents every year? Statistics vary from year to year. The amount of car accidents could go below 6 million one year."	"Was a census conducted in order to accurately represent a parameter? Or an observational study, if so where is the MOE? It's unlikely that there are exactly 6 million accidents per year. A MOE would be more likely to encompass the true parameter."

Table 4.3: Examples of students' responses and SLCR scoring for the LiveActive advertisement

	Examples of student comments based on SLCR scores		
Component	**1**	**2**	**3**
Bias	"The question I would ask concerning LiveActive's claim would be who collected the information for this poll? Were all women properly represented or was there a bias on who took the poll?"	"Was the sample of women tested biased in any way? A biased sample can lead to incorrect results."	
Causality	"And does the substances contained in 'cottage cheese' cause 'women' to be 'regular' exclusively? This study (or any study) is NOT able to prove their claim.		
Definition	"In this claim, we do not know what LiveActive means by 'occasional irregularity.' How messed up does their cycle have to be in order for them to be included in this category?"	"Also the definition of irregularity? An upset stomach or major intestinal complications? This definition as well as the definition of occasional could have a major impact on the responses of the participants."	"Firstly, what is the definition of irregularity? If this is a broad definition it could lead to false positives and make irregularity seem more common than it is, in order to sell LiveActive."
Generalize	"Additionally who were these women? Does it apply only to Americans or is it on a global scale."	"Also, when they say '4 out of 5 women' are they concluding women all over the world or just in the U.S. or in Costa Rica? If there is no clear idea of what group of women they're	"Once again, non-random sampling creates bias and is not representative of the population. We have to be sure this sample was accurate otherwise it cannot be applied to all women. We

		talking about, then the whole ad could be false or mistaken to be false."	do not know if this can be applied to everyone."
Lurking variable	"What were the age/race/height/weight of the women? Each factor can have an unwanted effect."	"What age were the women who were asked? This could majorly affect your results because maybe older women struggle more with this than younger women. What was the current health condition of the women asked because there could be a lurking variable. Were the women from different environmental surroundings or all from the same place? This could majorly affect your results because people who live in similar conditions often have similar struggles and such."	"I wonder where the sample of women was acquired from, and if the dietary traditions of the region lead to an increase in the proportion of women who experience irregularity. The diet may be a lurking variable in this instance."
Method	"How was this study conducted? This is important because this is a topic that could be sensitive with a lot of women."	"Did they construct a random and representative survey? If they surveyed people in a doctor's office they might get different results than they would in a park."	"How did LiveActive choose their women to survey? I feel that this should be identified because it is important to know if this was a random sample or not. What if LiveActive got their info from

			doctors about patients? I think that would increase the proportion of women who reported suffering from irregularity because who would visit their doctor if they were completely healthy and 'regular'? Not as many healthy, regular people visit the doctor, or at least I don't believe so. How did LiveActive obtain this information? Like I said in reason #2, if they just collected info from doctors about female patients that would probably affect the proportion suffering from irregularity."
Reported statistic	"Another question is how did they come up with these numbers '4 out of 5'"	"How did they reach this claim? What type of research did they conduct to get this statistic? Depending on how they got their answers affects how the participants responded."	"And statistics do not equal parameters."
Variation	"Because results can vary between each woman that decides to eat LiveActive cottage	"First of all, 'How big is the sample?' If they only polled 10 people, their	"What was the sample size? Sample size is important

	cheese."	standard error would be very high. The more people that they ask, the better."	considering that their results could have a large margin of error if the sample size is small. This margin of error could create an entirely different claim that could prove the claim false."

Table 4.4: Examples of students' responses and SLCR scoring for the depression article

	Examples of student comments based on SLCR scores		
Component	**1**	**2**	**3**
Bias	"How were the questions worded? May influence the girls' responses. Why was this study conducted? The researcher may be wanting a specific outcome in her results."	"What was the question the researchers asked that led to the teens admitting they have been depressed once in their life? The wording of the question can have a huge impact on the answer of the participant."	
Causality			"How did these researchers conclude a cause-effect relationship from an observational study? There were no treatments in this study, it was not a designed experiment if they simply asked the participants whether or not they suffered from depression. Therefore there is

			no possible way to conclude the subtitle – 'Teen girls twice as likely to suffer depression than boys' The results of this study can show correlation between gender and depression, but not causation."
Definition	"What is the definition of 'depression'? Many people think that just a sad moment or when you cry means depression."	"What classifies a person as being depressed? This is important to ask because the definition for one person might not be the same definition as another person thinks when they hear the word depressed which would make the results not as accurate."	"I would first like to ask what the specific detail of 'depression' is. This is an important question because we don't know how broad or narrow these researchers classified 'depression' as. If the definition was too broad then more girls than expected could have been depressed at least once in their life. Depending on what the definition means any girl could be depressed."
Generalize	"What was the population? All teens in Canada? All teens in the world? This is important to ask because we need to know who the target group of the study was."	"How did they extrapolate to say that 1 in 10 teen girls (implying all teen girls, not just Canadians) suffer from depression each year? The article only reported results for Canadian girls, yet the title and opening line of	"Is it just Canadians that have these depression rates or do these statistics apply to teens world wide? Furthermore, if the researchers are just surveying a sample of Canadian teens, their conclusion could be inaccurate and

		the article imply that all teen girls are more likely to suffer from depression. This is important to ask because you must know if the study's sample was representative of all teen girls in order to evaluate their conclusion about all teen girls."	might not apply to every teen."
Lurking variable	"What type of environment/ surroundings are these teens growing up with? A teen's home life may have a lot to do with their emotions. If a teen is growing up with an alcoholic parent or parents who are never around, they may not be as happy. If the teen is living in a family with money restrictions, going through a divorce, or loss of loved ones, this could also have a 'temporary' impact on the teen for depression."	"Did the subjects have previous depression episodes? This is important because your results could be biased if the subjects you tested had previous problems with depression and therefore the study would not be accurate. What type of social life do the teen girls have? This is very important because if a girl does not have a lot of friends and doesn't socialize then they will be more lonely and more susceptible to depression. What type of home life do the girl subjects have? This is also important because it could skew the results. If a teen girl	"How were the lifestyles of the teens studied like? Lurking variables can play a huge role in the misleading information given. What were the demographics of the teens? A certain category could show a higher percentage of suffering from depression than others."

		has an abusive dad or had a tragic event happen it could cause an inclination towards depression."	
Method	"How did the study select the boys and girls surveyed? No info is given about the possible similarities of the subjects. I think it is important to know whether or not the numbers of participants were evenly distributed between 12-19 years of age since age is a possible affecting factor of depression occurrence."	"What were the teenagers taken from? A good, random, representative sample must be true for this study to conclude this. What kind of questions did they ask to receive their results? Not using correct methods of surveying evokes different responses from people, altering results. What sampling method was used to get data from the 1,322 boys and girls? The sampling method directly influences the outcome. Not having a good sampling method can make the sample either not be random or representative like it should."	"An important question to ask is how they got this sample. We are unaware if the boys and girls were randomly selected or non-random selection. This is an important issue to identify because non-random sampling can be biased and unrepresentative of the population. Another issue is our sample size. We are told there are 1,322 subjects in this study, but we do not know how many were boys and how many are girls. It's important to have a large sample because a statistic is not a parameter. If we have a larger sample, we can be sure the statistic is getting closer to the true proportion or mean. Another risk is the time frame. Even four years later, the women were not all the

			same age. The variables are not controlled well in this observational study.”
Reported statistic	“What were the actual rates of depression? The exact proportions are not given, only ‘about one in 10’ and ‘about twice the rate.’ It would be good to know the true proportions.”	“In Dr Galambo’s statement, what number is ‘substantial’? This could provide insight to the accuracy of the statistics in the article.”	“I also would ask what they mean by ‘at the start of the study teens ranged from 12 to 19.’ I would ask that because if they just started with those ages but also included older people it would give a false statistic.”
Variation	“What level of confidence do we have that we can make generalizations about all of Canada from this sample?”	“What is her margin of error?”	“What is the level of confidence and confidence interval? We want to know how accurate the statistic is.”

Although responses observed from open-ended questions as used in this study were to simulate real life, in some cases assigning scores to students’ responses was difficult. It was more difficult to discriminate between responses that were assigned a score of 2 versus 3, and in retrospect, using a 3-point rather than 4-point scale may have made scoring easier and possibly more reliable. The use of a 4-point scale may be more applicable if students were directed to comment on a particular statistical literacy component.

In addition, results could have been scored with either a 0 or 1 indicating the absence or presence of questioning a particular statistical literacy component. This approach was not used because it would not provide information concerning the extent to which respondents were aware of the statistical literacy component but would only indicate that students were aware of the statistical literacy component.

Inter-rater Reliability

The reliability of scoring of SLCR components was evaluated, and results are discussed in this section. Inter-rater reliability is a measure of the agreement between ratings made by two or more people. The Kappa statistic is a measure of inter-rater reliability and is defined as the proportion of agreement beyond that expected by chance (Sim & Wright, 2005). The formula for Kappa is:

$$\text{Kappa} = \frac{\text{observed agreement} - \text{chance agreement}}{1 - \text{chance agreement}}$$

To determine the inter-rater reliability of SLCR scoring, two instructors who taught SIEL since inception independently evaluated 25 randomly selected students' responses to the two advertisements and article. Determining this sample size was determined using the procedure described in Cantor (1996).

For each SLCR component, inter-rater reliability of scoring was evaluated based on two methods of characterizing responses: (1) presence/absence of response for a component and (2) on a scale from 0 to 3 based on a student's level of awareness of the component. For the first method of categorizing, the Kappa statistic, asymptotic standard error, 95% confidence interval for Kappa, prevalence index, bias index, percentage of positive and negative agreements, and Kappa_{max} were computed for each statistical literacy component (Table 4.5). Table 4.6 provides a comparison between the Kappa statistics and $Kappa_{max}$ for each statistical literacy component. Weighted Kappa statistics and 95% confidence intervals for Kappa were determined for the second method of categorizing responses. Weighted Kappa statistics were computed since the 4-point scale represents ordinal categories (Table 4.7).

Table 4.5: Kappa statistics, 95% confidence interval for Kappa, prevalence index, bias index and $Kappa_{max}$ for components of SLCR for presence/absence of each component

Component	Kappa[1]	ASE[2]	95% Confidence interval for Kappa	Prevalence index[3]	Bias Index[4]	% positive agreement	% negative agreement	$Kappa_{max}$[5]
Bias	0.3998	0.0715	(0.2596, 0.5399)	0.42	0.2067	16%	58%	0.5229
Causality	0.4292	0.2088	(0.02, 0.8385)	0.94	0.02	1.33%	95.33%	0.6575
Definition	0.7950	0.0542	(0.6888, 0.9011)	0.3933	0.0067	65.33%	26%	0.9841
Generalize	0.2809	0.0989	(0.0871, 0.4747)	0.6667	0.0267	6.67%	73.33%	0.904
Lurking Variable	0.4402	0.0684	(0.3061, 0.5743)	0.06	0.1667	38.67%	32.67%	0.6745
Method	0.5910	0.0634	(0.4668, 0.7153)	0.0733	0.1267	36%	43.33%	0.7493
Reported Statistic	0.3173	0.0886	(0.1436, 0.4911)	0.5267	0.0067	11.33%	64%	0.9815
Variation	0.5588	0.1047	(0.3535, 0.7641)	0.76	0.0267	7.33%	83.33%	0.8738

[1]Proportion of agreement beyond that expected by chance
[2]Asymptotic standard error
[3]Extent to which proportion of agreements on positive classification differs from proportion of agreements on the negative classification
[4]Extent to which raters disagree on the proportion of positive cases
[5]Maximum attainable Kappa

Table 4.6: Relationship between Kappa statistic and $Kappa_{max}$ for each component of statistical literacy in SLCR

Component	Difference between $Kappa_{max}$ and kappa statistic	Percentage of Kappa statistic to $Kappa_{max}$
Bias	0.1231	76.46%
Causality	0.2283	65.28%
Definition	0.1891	80.78%
Generalize	0.6231	31.07%
Lurking Variable	0.2343	65.26%
Method	0.1583	78.87%
Reported Statistic	0.6642	32.33%
Variation	0.315	63.95%

Table 4.7: Weighted Kappa statistics and 95% confidence interval for Kappa for components of SLCR based on a 4-point scale

Component	Weighted Kappa	95% Confidence interval for weighted Kappa
Bias	0.3607	(0.2404, 0.4809)
Causality	0.3190	(0.0294, 0.6085)
Definition	0.5260	(0.4221, 0.63)
Generalize	0.1839	(0.0564, 0.3115)
Lurking Variable	0.3680	(0.2623, 0.4738)
Method	0.4620	(0.3598, 0.5642)

Reported Statistic	0.3114	(0.1504, 0.4724)
Variation	0.3555	(0.2179, 0.4931)

Landis and Koch (1977) provided criteria for evaluating the strength of inter-rater reliability using the Kappa statistic. Using results from Table 4.5 and the criteria from Landis and Koch (1977), inter-rater reliabilities for statistical literacy components Bias, Generalize, and Reported Statistic were considered "fair" since these Kappa values fell in the range from 0.21 to 0.4. "Moderate" agreement was observed for components Causality, Lurking Variable, Method, and Variation because the Kappa statistics for these components were from 0.41 to 0.6. The strength of agreement for the Definition component was considered "substantial" because the Kappa statistic was between 0.61 and 0.8. Definition and Method had the highest values of Kappa, while the lowest values of Kappa were observed for Generalize and Reported statistic.

Students' responses concerning Definition were more obvious to detect making scoring more consistent. Components, Bias, Generalize, and Reported statistic, were sometimes difficult to determine from students' open-ended responses and this made scoring these components challenging.

The asymptotic standard error (ASE) for each statistical literacy component and 95% confidence intervals for Kappa were determined using SAS version 9.2 (Cary, NC). The highest values of ASE were observed for components Causality and Variation indicating the largest degree of variation of ratings. The lower limits of all 95% confidence intervals were above 0 which indicated the strength of agreement of ratings for each statistical literacy component was more than what would be expected by chance. Even so, the lower limits of the confidence intervals for Causality and Generalize were barely above 0.

Evaluating inter-rater reliability should not exclusively be determined based on the value of the Kappa statistic. According to Hoehler (1999) and Sim and Wright (2005), values of Kappa statistics are influenced by prevalence and bias and should be taken into consideration when interpreting Kappa statistics. The prevalence index measures the extent to which the proportion of agreements on positive classifications differs from the proportion of agreements on negative classifications. If the prevalence index is high then chance agreement is also high (Sim & Wright, 2005). When the percentage of positive agreements is very different from 50%, prevalence has an effect on the Kappa statistic (Hoehler, 1999). The Kappa statistic value is lower when the prevalence index is low or zero.

The highest prevalence index was observed for the Causality component and lowest prevalence indices for Lurking Variable and Method components (Sim & Wright, 2005). These results can be explained by looking at the percentages of positive and negative agreements. The greatest discrepancy between the percentages of positive and negative agreements resulted in a large prevalence index while percentages of positive and negative agreements that were more similar resulted in a prevalence index that was smaller. Since the prevalence indices for Lurking variable and Method were low, this may have reduced the value of the Kappa statistics for these statistical literacy components.

Another factor to consider when evaluating inter-rater reliability is bias. Bias is the degree to which raters disagree on the proportion of positive cases (Sim & Wright, 2005). When two raters differ on the proportion of positive results, bias can affect the value of the Kappa statistic. High values of bias tend to reduce Kappa (Hoehler, 1999). The highest values of bias were observed for components Bias, Lurking Variable, and Method. Although this result may indicate that there was a need for more training for rating these components, high bias indices may be the result of low prevalence indices. According to Hoehler (1999), increased bias was associated with low prevalence. This relationship is shown in Table 4.5 for Lurking variable and Method but not for Bias.

Another consideration when interpreting Kappa statistics is to consider the value of $Kappa_{max}$. The greatest attainable value of a Kappa statistic is defined as $Kappa_{max}$. It is a "reference value for Kappa that preserves the individual clinician's overall propensity to diagnose a condition or select a rating within the restraints imposed by marginal totals" (Sim & Wright, 2005, p. 264). Kappa statistics should be considered in relation to $Kappa_{max}$ by determining the difference between $Kappa_{max}$ and Kappa statistic. This difference represents the "unachieved agreement beyond chance within the constraints of the marginal totals" (Sim & Wright, 2005, p. 264).

For each statistical literacy component, the difference between $Kappa_{max}$ and Kappa statistic can be found in Table 4.6. Also, the percentage of each Kappa statistic to its respective $Kappa_{max}$ was determined. This would provide additional information concerning the relationship between the Kappa statistic and.$Kappa_{max.}$

The largest difference between $Kappa_{max}$ and Kappa statistic was observed for Generalize and Reported Statistics. These components also had the smallest percentage of Kappa statistic to $Kappa_{max}$ (Table 4.6) and the smallest Kappa statistics of the eight statistical literacy components (Table 4.5). These results suggest that components Generalize and Reported Statistics were most difficult to identify and that more information in SLCR or training should be given to help raters make identifications of these components.

Table 4.7 provides weighted Kappa statistics and 95% confidence intervals for Kappa for each component of statistical literacy when students' responses were scored using the SLCR 4-point scale. Highest weighted Kappa statistics were observed for Definition and Method, while the lowest weighted Kappa statistics were found for Generalize and Reported statistic. Similar results were observed in Table 4.5 for the first method of categorizing responses. All lower limits of the 95% confidence intervals were above 0 suggesting that rater agreement using the 4-point scale was more than what would be expected by chance. Again, the lower limits of confidence intervals for Causality and Generalize were slightly above 0 using the 4-point scale.

When evaluating inter-rater reliability, Kappa statistics should not be the only measure that is used since this statistic can be influenced by other factors as described in this section (Sim & Wright, 2005). Prevalence, bias, percentage of positive and negative agreements, $Kappa_{max}$, and 95% confidence intervals for Kappa should also be taken into consideration.

Since the Kappa statistics for Generalize and Reported statistic were in the "fair" category, and they had the lowest percentages of Kappa statistic to $Kappa_{max}$, highest differences between $Kappa_{max}$ and Kappa statistic, lowest weighted Kappa statistics, and lower limits for the 95% confidence intervals for Kappa and weighted Kappa were near 0, there is concern about the inter-rater reliabilities for Generalize and Reported statistic. Similar to Generalize and Reported statistic, the lower limits of the 95% confidence intervals in Table 4.5 and 4.7 for Causality were near 0. Unlike Generalize and Reported statistic, the difference between $Kappa_{max}$ and Kappa statistic for Causality was lower, the percentage of Kappa statistic to $Kappa_{max}$ was higher, and the percentage of negative agreements was much higher. Based on these results, concern about the inter-rater reliability of Causality was not at the same level as for Generalize and Reported statistic.

Identification of responses in the Generalize and Reported statistic components was more difficult than the other components. Therefore, caution should be taken when interpreting results for the Generalize and Reported statistic components. In the future, more guidance in identification of these components, perhaps by giving specific examples of responses in each scoring category, should be provided.

Research Question 1a

Overview.

The purpose of Research Question 1a was to assess the level of statistical literacy for college students prior to taking a course focused on statistical literacy. The relationship between gender, attitude, aptitude, and background upon this level was investigated. The results for this research question are discussed in this section.

Results for pre-course SATS-36© attitude scores.

In this section, pre-course results from SATS-36© are discussed. Internal consistency was determined for each pre-course average attitude component score and results are presented in this section. Results from correlation analyses of pre-course SATS-36© attitude components are also included.

On the first day of classes, students enrolled in SIEL were given the opportunity to take SATS-36©. Table 4.8 provides a summary of average SATS-36© attitude component scores, standard deviations, and standard errors for participating students. The averages were based on a 7-point Likert scale. Based on Tempelaar, Van Der Loeff, and Gijselaers (2007), an average above a neutral value of four indicated a positive attitude toward the attitude component. All pre-course averages for SATS-36© attitude components were above four except for Difficulty which was approximately four. These results suggested that students in the study expressed positive attitudes toward statistics for Affect, Cognitive Competence, Value, Interest, and Effort and a neutral attitude toward Difficulty. A positive attitude toward Affect indicated that prior to taking SIEL students had positive feelings toward statistics. A positive attitude toward Cognitive Competence suggested that students had positive attitudes about their intellectual knowledge and skills concerning statistics. A positive attitude toward Value was related to a positive attitude about the worth and usefulness of statistics in everyday life. A positive attitude toward Interest suggested that students had an interest in statistics, and a positive attitude toward Effort indicated that students felt that they did not have to do a lot of work to learn statistics. A neutral attitude toward Difficulty indicated neither positive nor negative attitudes about students' perception of the difficulty of learning statistics.

Effort had the highest average attitude component score, and Difficulty had the lowest average attitude component score. These results were similar to results obtained in Schau (2003) and Tempelaar et al. (2007).

Table 4.8: Averages, standard deviations, standard errors, and Cronbach's α for pre-course attitude component scores from SATS-36©

Attitude Component	Average	Standard deviation	Standard error	Cronbach's α
Affect	4.49	1.13	0.10	0.80
Cognitive Competence	5.12	1.06	0.09	0.80
Difficulty	3.96	0.69	0.06	0.81
Effort	6.32	0.72	0.06	0.86
Interest	4.78	1.15	0.10	0.81
Value	5.20	0.82	0.07	0.81

Internal consistency results for pre-course SATS-36© attitude components.

Cronbach's α , a measure of internal consistency, was determined for each attitude component in SATS-36© (Table 4.8). These values ranged from 0.8 to 0.86 and were within the range of values from Schau (2003) and Tempelaar et al. (2007) for components Affect, Cognitive Competence, Value, and Difficulty. Since attitude components, Interest and Effort, were added since Schau (2003), comparisons for these components were made with Tempelaar et al. (2007). In Tempelaar et al. (2007), a comparable Cronbach's α value was observed for Interest (0.8) but lower values for Effort (0.76) and Difficulty (0.68) as compared to results in Table 4.8. Based on criteria for interpreting values of Cronbach's α from George and Mallory (2003), α-values at least 0.8 but less than 0.9 are considered to have "good" internal consistency. Therefore, all pre-course attitude components were categorized as having "good" internal consistency.

Correlation analyses involving pre-course SATS-36© attitude component scores and demographic variables..

In order to determine if there was a significant linear relationship between demographic variables and pre-course attitude component scores, Pearson correlation coefficients were determined, and hypothesis tests were conducted using the Bonferroni correction as described in the Chapter 3. The results from all correlation analyses of demographic variables and pre-course attitude component scores from SATS-36© can be found in Table 2 Appendix D. Since 90 comparisons were made and an alpha of 0.05 was used, the significance level for Table 2 Appendix D was determined to be 0.0006. Table 4.9 provides a summary of all statistically significant variables, and these correlations were positive.

Table 4.9: Significant* correlations between demographic variables and pre-course SATS-36© attitude components

	Reported Math SAT score	# college math courses taken	Pre-course Affect	Pre-course Cognitive Competence	Pre-course Difficulty
Reported Math SAT score			0.39589	0.37077	
# college math courses taken			0.3468	0.34418	0.29572
Pre-course Affect					
Pre-course Cognitive Competence					
Pre-course Difficulty					

*Significant at the $p < 0.0006$ level.

Reported Math SAT scores were correlated with pre-course Affect and pre-course Cognitive Competence. A higher reported Math SAT score may be associated with more success in the math courses taken and that might have influenced Affect and Cognitive Competence. Similar results were observed with Cognitive Competence in Dempster and McCorry (2009). The number of college math courses taken was correlated with pre-course Affect, pre-course Cognitive Competence, and pre-course Difficulty. Those taking more college math courses may be enrolled in majors that require quantitative skills and have had more positive experiences with mathematics. This may provide an explanation for these relationships.

In this study, pre-course Affect was not significantly correlated with achievement (course grade), but there was a significant correlation between these variables in Bude′ et al. (2007). Effort, as determined by tutors’ perceptions, was correlated with Affect in Bude′ et al. (2007), but this relationship was not observed in this study.

Pre-course SLCR scores.

At the beginning of the semester, students were asked to provide questions or concerns about information in two advertisements and an article. These responses were evaluated based on the level of awareness of eight statistical literacy components in SLCR.

The distribution of SLCR pre-course scores for each statistical literacy component observed for the Allstate advertisement, LiveActive advertisement, and depression article, respectively, are presented in Tables 4.10, 4.11, and 4.12, respectively. For the Allstate advertisement, the mode pre-course scores were 0 for all statistical literacy components except for Definition (Table 4.10). The mode of pre-course scores for Definition was 1. Results indicated that students had virtually no awareness of most statistical literacy components and a minimal awareness of definition issues with the Allstate advertisement prior to taking SIEL.

Table 4.10: Summary of SLCR scores pre-course for Allstate advertisement (number of students and percentage)

	Number of SLCR scores (%)			
Component	**0**	**1**	**2**	**3**
Bias	97 (67.36)	44 (30.56)	1 (0.69)	2 (1.39)
Causality	132 (91.67)	11 (7.64)	1 (0.69)	0 (0)
Definition	51 (35.42)	67 (46.53)	18 (12.5)	8 (5.56)
Generalize	140 (97.22)	1 (0.69)	2 (1.39)	1 (0.69)
Lurking variable	68 (47.22)	59 (40.97)	13 (9.03)	4 (2.78)
Method	119 (82.64)	17 (11.81)	5 (3.47)	3 (2.08)
Reported statistic	91 (63.19)	30 (20.83)	21 (14.58)	2 (1.39)
Variation	130 (90.28)	2 (1.39)	12 (8.33)	0 (0)

For the LiveActive advertisement, a score of 0 was the mode of scores for all pre-course statistical literacy components in SLCR except for Definition and Method (Table 4.11). For these components, a score of 1 was the mode. These results suggested that for the LiveActive advertisement students had the highest level of awareness for Definition and Method before taking SIEL.

Table 4.11: Summary of SLCR scores pre-course for LiveActive advertisement (number for each score and percentage)

	Number of SLCR scores (%)			
Component	**0**	**1**	**2**	**3**
Bias	100 (69.44)	41 (28.47)	3 (2.08)	0 (0)
Causality	128 (88.89)	16 (11.11)	0 (0)	0 (0)
Definition	58 (40.28)	66 (45.83)	18 (12.5)	2 (1.39)
Generalize	120 (83.33)	19 (13.19)	5 (3.47)	0 (0)
Lurking variable	69 (47.92)	61 (42.36)	11 (7.64)	3 (2.08)
Method	60 (41.67)	68 (47.22)	14 (9.72)	2 (1.39)
Reported statistic	127 (88.19)	10 (6.94)	7 (4.86)	0 (0)
Variation	140 (97.22)	2 (1.39)	1 (0.69)	1 (0.69)

For the depression article, the mode of pre-course scores for all statistical literacy components in SLCR was 0 except for components, Lurking variable and Method (Table 4.12). The mode of scores for these components was 1. For Bias, the percentage of scores of 0 and 1 were the same. Prior to taking SIEL, the highest level of awareness for the depression article was observed for Lurking variable and Method.

Table 4.12: Summary of SLCR scores pre-course for depression article (number for each score and percentage)

	Number of SLCR scores (%)			
Component	**0**	**1**	**2**	**3**
Bias	108 (75)	25 (17.36)	11 (7.64)	0 (0)
Causality	144 (100)	0 (0)	0 (0)	0 (0)
Definition	83 (57.64)	42 (29.17)	18 (12.5)	1 (0.69)
Generalize	113 (78.47)	9 (6.25)	17 (11.81)	5 (3.47)
Lurking variable	52 (36.11)	73 (50.69)	14 (9.72)	5 (3.47)
Method	49 (34.03)	68 (47.22)	27 (18.75)	0 (0)
Reported statistic	135 (93.75)	4 (2.78)	4 (2.78)	1 (0.69)
Variation	141 (97.92)	0 (0)	2 (1.39)	1 (0.69)

Despite the different contexts of the media articles, Definition and Method had the highest level of awareness prior to taking SIEL for two of the three media articles. These results suggested that there was some awareness of Definition and Method but virtually no awareness of the other statistical literacy components before taking SIEL.

Based on the information in Tables 4.10, 4.11, and 4.12, selection of media articles is important to consider when assessing statistical literacy because the focus of the media article may have an effect on responses. The topic of the article may influence what statistical literacy components are more readily questioned. Watson and Moritz (2000) also found similar results of the effect of context on students' responses. A broad range in articles' foci is important to use when assessing statistical literacy so as to mimic different situations encountered in everyday life. Another explanation for results may be that students have more familiarity with statistical literacy components, Definition and Method and possibly Lurking variable, prior to taking a course on statistical literacy. It is important to note that average scores for these components reflected only minimal awareness.

Like attitudes toward statistics, a total statistical literacy score was not computed. Instead, individual statistical literacy component scores were determined for each media article and then averaged to give an average statistical literacy component score for each component. The average and median pre-course SLCR component scores along with standard deviations and standard errors can be found in Table 4.13. These averages were based on a scale from 0 to 3 in which a higher average component score indicated an increase in awareness of the statistical literacy component. Medians were included since the distribution of SLCR scores in Tables 4.10, 4.11, and 4.12 were skewed.

The highest average and median pre-course SLCR component scores were observed for Definition, Lurking Variable, and Method. Although these components had the highest averages and medians, these values were less than 1 indicating at most a basic awareness of each statistical literacy component. The lowest average pre-course SLCR scores were observed for Causality and Variation which were close to 0, indicating essentially no awareness of these statistical literacy components. The medians for these components were 0.

Table 4.13: Averages, medians, standard deviations, and standard errors for average pre-course SLCR component scores

Component	Average	Median	Standard deviation	Standard error
Bias	0.33796	0.3333	0.3698	0.0308
Causality	0.06713	0	0.1653	0.0138
Definition	0.73148	0.6667	0.5037	0.420
Generalize	0.21991	0	0.3785	0.0315
Lurking variable	0.70602	0.6667	0.4994	0.0416

Method	0.60185	0.6667	0.4395	0.0366
Reported statistic	0.26389	0	0.3434	0.0286
Variation	0.09259	0	0.2628	0.0219

Table 4.14 provides the counts and percentages of students' average pre-course SLCR scores. The range is one measure of variation with a higher value indicating more variation. The largest pre-course range of average SLCR scores was observed for Method (2.6667) while Definition and Lurking variable had the second largest range of average SLCR scores (2.3333). The lowest pre-course range of average SLCR scores was observed for Causality (1). Results concerning variation were also observed in Table 4.13 with Causality scores having the lowest standard deviation and Definition, Lurking variable, and Method having the highest standard deviations. The pre-course SLCR average with the highest percentage for Method and Lurking variable was 0.6667 while for Bias, Causality, Generalize, Reported statistic, and Variation it was 0.

Table 4.14: Summary of average pre-course and post-course SLCR scores (number for each score and percentage)

	Number of SLCR scores and (Percentage of Scores)									
Component	**0**	**0.3333**	**0.6667**	**1**	**1.3333**	**1.6667**	**2**	**2.3333**	**2.6667**	**3**
Bias pre-course	63 (43.75)	35 (24.31)	31 (21.53)	12 (8.33)	2 (1.39)	1 (0.69)	0	0	0	0
Bias post-course	25 (17.36)	40 (27.78)	35 (24.31)	26 (18.06)	13 (9.03)	3 (2.08)	2 (1.39)	0	0	0
Causality pre-course	120 (83.33)	20 (13.89)	3 (2.08)	1 (0.69)	0	0	0	0	0	0
Causality post-course	121 (84.03)	18 (12.5)	1 (0.69)	4 (2.78)	0	0	0	0	0	0
Definition pre-course	21 (14.58)	33 (22.92)	30 (20.83)	26 (18.06)	28 (19.44)	4 (2.78)	1 (0.69)	1 (0.69)	0	0
Definition post-course	1 (0.69)	3 (2.08)	21 (14.58)	48 (33.33)	31 (21.53)	18 (12.5)	12 (8.33)	8 (5.56)	2 (1.39)	0
Generalize pre-course	97 (67.36)	17 (11.81)	19 (13.19)	6 (4.17)	4 (2.78)	0	1 (0.69)	0	0	0
Generalize	78	16	24	11	9	4	2	0	0	0

post-course	(54.17)	(11.11)	(16.67)	(7.64)	(6.25)	(2.78)	(1.39)			
Lurking variable pre-course	22 (15.28)	33 (22.92)	33 (22.92)	30 (20.83)	17 (11.81)	7 (4.86)	1 (0.69)	1 (0.69)	0	0
Lurking variable post-course	9 (6.25)	24 (16.67)	37 (25.69)	26 (18.06)	28 (19.44)	9 (6.25)	10 (6.94)	1 (0.69)	0	0
Method pre-course	22 (15.28)	40 (27.78)	45 (31.25)	26 (18.06)	8 (5.56)	0	2 (1.39)	0	1 (0.69)	0
Method post-course	4 (2.78)	18 (12.5)	30 (20.83)	38 (26.39)	30 (20.83)	15 (10.42)	4 (2.78)	4 (2.78)	1 (0.69)	0
Reported statistic pre-course	80 (55.56)	26 (18.06)	28 (19.44)	8 (5.56)	2 (1.39)	0	0	0	0	0
Reported statistic post-course	46 (31.94)	38 (26.39)	39 (27.08)	12 (8.33)	7 (4.86)	1 (0.69)	0	1 (0.69)	0	0
Variation pre-course	124 (86.11)	4 (2.78)	15 (10.42)	0	0	0	1 (0.69)	0	0	0
Variation post-course	98 (68.06)	7 (4.86)	22 (15.28)	9 (6.25)	1 (0.69)	3 (2.08)	2 (1.39)	0	1 (0.69)	1 (0.69)

Correlation analyses for pre-course average SLCR component scores with demographic variables and SATS-36© attitude component scores.

Correlation analyses were performed between pre-course average SLCR statistical literacy component scores and demographic variables. In addition, pre-course average SLCR statistical literacy component scores and attitude component scores from SATS-36© were correlated.

A complete list of results from correlation analyses between demographic variables and pre-course average SLCR statistical literacy component scores can be found in Table 3 Appendix D. Since 120 comparisons were made and an alpha of 0.05 was used, the significance level using the Bonferroni correction was 0.0004. The only variables which were significantly correlated were average pre-course Variation score and the number of college statistics courses taken. The correlation coefficient for these variables was 0.29808 (p-value = 0.0003). Since "variation" is a fundamental concept in statistics courses, this may explain this relationship. Since a significant correlation existed between the average Variation score and the number of college statistics courses taken but not the number of high school statistics courses taken, this may suggest that "variation" is a concept which takes either more statistics classes to understand or a higher level of instruction.

A complete list of results from correlation analyses between pre-course average SLCR statistical literacy component scores and pre-course SATS-36© attitude component scores can be found in Table 4 Appendix D. Using the Bonferroni correction, the significance level was 0.001 (alpha of 0.05 and 48 comparisons). No significant correlations were observed.

Pre-course average SLCR statistical literacy component scores and post-course SATS-36© attitude components were correlated. A complete list of results can be found in Table 5 Appendix D. Using the Bonferroni correction, the significance level was 0.001 (alpha of 0.05 and 48 comparisons). No significant correlations were observed.

Correlation results between pre-course average SLCR scores and pre-course and post-course SATS-36© attitude component scores suggest that there is not a strong linear relationship between these variables. With a small significance level that was used, it is not surprising that these results were obtained. More research should be conducted to further study statistical literacy and attitude toward statistics.

Regression analyses for pre-course average SLCR component scores.

To determine if gender, attitude, aptitude, and backgrounds have a significant effect on statistical literacy levels, regression models were determined for each statistical literacy component as described in Chapter 3. The dependent variable of each model was the average pre-course statistical literacy component score. Independent variables were those variables listed in Table 3.2 as well as gender. As described in Chapter 3, forward selection was used to determine significant independent variables for each dependent variable to produce eight regression models. For each regression model, the normality assumption was tested using normal probability plots. These plots did not indicate any problems with the assumption of normality.

Statistical significance was determined at a level of 0.1. This significance level was changed from the traditional 0.05 level in order to reduce the likelihood of a Type II error. Significant independent variables for each statistical literacy component can be found in Table 4.15. The adjusted R^2 value, a measure of the amount of variation in the dependent variable that is explained by the model, is provided for each regression model in Table 4.15.

Table 4.15: Significant* independent variables resulting from forward selection for pre-course statistical literacy component scores

Statistical Literacy Component	Significant independent variable and (p-value)	Adjusted R^2 value
Bias	Effort Pre-course (0.0732)	0.0235
Causality	Cumulative earned credits (0.0329)	0.0374
Definition	Verbal SAT score (0.0034) Course grade (0.0614) Cumulative earned credits (0.0952)	0.1582
Generalize	Number of college statistics courses	0.1003

	taken (0.0198) Number of high school statistics courses taken (0.0315) Math SAT score (0.0994)	
Lurking variable	Math SAT score (0.0116) Difficulty Pre-course (0.051)	0.0842
Method	Course grade (0.0116) Difficulty Pre-course (0.0531)	0.0484
Reported statistic	None	0
Variation	Number of college statistics courses taken (0.0023) Cumulative college points (0.0263) Value Pre-course (0.0312)	0.1574

*Significant at the $p < 0.1$ level.

It is interesting to note that statistical literacy components Generalize and Variation, which are essential foundational concepts in statistics, are significantly predicted by the number of college statistics courses taken before SIEL. Verbal SAT score was a significant predictor of Definition component which may suggest that the level of awareness of the Definition component may be associated with verbal abilities. Math SAT score was a significant predictor of Generalize and Lurking variable which may imply that these components are associated with quantitative skills.

Several pre-course attitude components from SATS-36© were significant predictors of Bias, Lurking variable, Method, and Variation although why these results were observed is unclear. Future research should focus on investigating relationships between specific attitude components and statistical literacy.

Despite statistically significant predictors of pre-course statistical literacy components, the adjusted R^2 values for the regression models were low. This indicates that there are other predictors of the statistical literacy components not accounted for in this study.

The following are the regression model equations (numbered from 1 to 8) for each average pre-course statistical literacy component and significant independent variables listed in Table 4.15.

(1) Average Pre-course Bias score = 0.8412 – 0.08*Effort Pre-course

(2) Average Pre-course Causality score = 0.1299 – 0.0014*Cumulative earned credits

(3) Average Pre-course Definition score = -0.726 + 0.0015*Verbal SAT score + 0.0047*Cumulative earned credits + 0.1245*Course grade

(4) Average Pre-course Generalize score = -0.2181 + 0.0008*Math SAT score – 0.1559*Number of high school statistics courses + 0.1393*Number of college statistics courses

(5) Average Pre-course Lurking variable score = 1.3426 + 0.1546*Difficulty Pre-course – 0.0021*Math SAT score

(6) Average Pre-course Method score = 0.8526 – 0.1393*Difficulty Pre-course + 0.114*Course grade

(7) Average Pre-course Reported statistic score = 0.2674

(8) Average Pre-course Variation score = -0.2472 + 0.0757*Value Pre-course + 0.1614*Number of college statistics courses – 0.0009*Cumulative points

For regression models (1) and (2), there was an inverse relationship between average pre-course statistical literacy score and the independent variable. For every 1-point increase in pre-course Effort, the average pre-course average Bias score decreased by 0.08 point. For every 1-point increase in cumulative earned credits, the pre-course average Causality score decrease by 0.0014 points. An explanation for these results is unclear.

Model (3) is the only regression model in which the partial slopes for the independent variables in the model were all positive. Since the Definition component may involve verbal skills, it is not surprising that the partial slope for Verbal SAT score was positive. For models (4) and (8), the partial slope for the number of college statistics courses taken is positive. Since the concept of "generalization" and "variation" are fundamental to statistics, this might explain this result.

Multivariate multiple regression model analyses for pre-course average SLCR component scores.

Since pre-course statistical literacy components scores may be dependent, a multivariate multiple regression model for the vector of pre-course average statistical literacy component scores was created using all significant independent variables listed in Table 4.15. The Wilk's lambda test was performed to determine the significance of the independent variables in Table 4.15 on the vector of pre-course statistical literacy component scores. Results from the Wilk's lambda test can be found in Table 4.16. Significance was determined at a level of 0.1 instead of the traditional 0.05 to reduce the chance of a Type II error. For comparisons, Table 4.16 also includes the results of the Wilk's lambda test for post-course statistical literacy scores and changes in scores.

Table 4.16: Multivariate multiple regression results for significant* variables for pre-course, post-course and the change in post-course and pre-course statistical literacy scores from SLCR

	Wilk's λ (p-value)		
Variable	Pre-Course SLCR components	Post-Course SLCR components	Difference (post-course – pre-course) in SLCR components
Gender		0.84 (0.0666)*	0.8337 (0.0773)*
Affect Pre-course		0.8831 (0.2360)	0.8872 (0.3156)
Difficulty Pre-course	0.8418 (0.0632)*		
Effort Pre-course	0.926 (0.5799)	0.852 (0.0975)*	0.8964 (0.3843)
Interest Pre-course			0.8617 (0.17)
Value Pre-course	0.9416 (0.7389)		0.9479 (0.8415)
Difficulty Post-course			0.9257 (0.6448)
Interest Post-course		0.8392 (0.0651)*	0.8201 (0.0509)*
Math SAT score	0.9252 (0.5713)	0.8771 (0.2014)	0.9046 (0.4515)
Verbal SAT score	0.8682 (0.1450)		
Total SAT score		0.8239 (0.0388)*	0.8226 (0.0551)*
No. high school math courses taken	0.9186 (0.5045)		
No. high school statistics courses taken	0.8977 (0.3195)	0.8876 (0.2644)	0.8837 (0.2913)
No. college math courses taken			0.9109 (0.5074)
No. college statistics courses taken	0.8197 (0.0293)*	0.9083 (0.4265)	0.8582 (0.1551)
Course Effort			0.8551 (0.1427)
Course grade	0.8935 (0.2880)		
Cumulative attempted credits			0.9292 (0.6782)
Cumulative points	0.8722 (0.1627)	0.9437 (0.7716)	
Cumulative earned credits	0.8482 (0.0781)*		

*Significant at the $p < 0.1$ level.

Of the significant independent variables from Models 1 through 8, only pre-course Difficulty, number of college statistics courses taken, and cumulative earned college credits had a significant effect on the vector of pre-course average statistical literacy component scores. Background attributes, the number of college statistics courses taken and cumulative earned college credits suggest the importance of experience of coursework and specifically statistics coursework in college on pre-course statistical literacy. In addition, how difficult students perceive statistics to be prior to the course may influence statistical literacy.

Summary for Research Question 1a.

The highest average SATS-36© attitude component scores prior to SIEL were observed for Cognitive Competence, Value, and Effort. All average attitude component scores were positive values indicating more positive attitudes toward the components. Higher average scores suggest that students felt that they had the intellectual ability to succeed in SIEL, would not need to expend much work for success in SIEL, and saw the usefulness of statistics in their lives. "Good" internal consistency results were observed for all attitude components based on Cronbach's α .

SLCR scores were determined for each media article (Tables 4.10, 4.11, and 4.12). The mode SLCR score for Definition was 1 for the Allstate and LiveActive advertisement, but a mode of 0 was observed for the depression article. For Method, a mode of 1 was observed for the LiveActive advertisement and depression article, but a mode of 0 was observed for the Allstate advertisement. A mode of 1 was observed for Lurking variable for the depression article, but for the Allstate and LiveActive advertisements the mode was 0. For all other statistical literacy components and media articles, the mode was 0. These results indicated minimal to no awareness of any of the statistical literacy components prior to SIEL. The level of statistical literacy may have been affected by the context of the media article as observed in Watson and Mortiz (2000) although for all components, awareness was low.

The average pre-course SLCR scores were determined across media articles, and the averages ranged from 0.09 to 0.73. These average scores indicated a minimal awareness of statistical literacy components prior to SIEL. Despite low scores for all components, the highest average scores were observed for Definition, Method and Lurking variable.

Reported Math SAT score and the number of college mathematics courses taken were both positively significantly correlated with pre-course Affect and Cognitive Competence. These results may indicate that those with better quantitative skills as suggested by higher Math SAT scores and more college mathematics courses taken liked statistics and felt they had the intellectual knowledge to succeed in SIEL prior to taking it.

The number of college mathematics courses was also positively significantly correlated with pre-course Difficulty. Those taking more college mathematics courses may have believed that SIEL would not require much work to succeed in it prior to taking it.

No significant correlations were observed between pre-course average SLCR scores and either pre-course or post-course SATS-36© attitude component scores. The use of the Bonferroni correction to determine statistical significance made the significance level small and this affected results. The only significant correlation observed for average pre-course SLCR scores and demographic variables was between average pre-course Variation score and the number of college statistics courses taken. Since "variation" is an important concept in statistics, students would have been exposed to this concept at a higher level, and this may provide an explanation for this relationship.

Regression analyses were performed using pre-course average SLCR scores as dependent variables and demographic variables as well as pre-course SATS-36© attitude component scores as potential independent variables. Although significant results were obtained, the regression coefficients for the models were small and not of practical significance. For the multivariate multiple regression model, pre-course Difficulty, number of college statistics courses taken, and cumulative earned college credits had a significant effect on the vector of pre-course average statistical literacy component scores. It is important to remember that the average pre-course SLCR scores indicated minimal to essentially no awareness of the statistical literacy components. Therefore, results from the regression and multivariate multiple regression models for pre-course SLCR scores are not of practical importance.

Research Question 1b

Overview.

The focus of Research Question 1b was to assess the level of statistical literacy for college students after taking a course on statistical literacy. The effect of gender, attitude, aptitude, and background upon this level of statistical literacy was investigated. The results for this research question are discussed in this section.

Results for post-course SATS-36© attitude scores.

Table 4.17 provides a summary of average post-course SATS-36© attitude component scores, standard deviations, and standard errors for participants. Based on Tempelaar et al. (2007) in which an average attitude component score above four was associated with a positive attitude, all average component values indicated positive attitudes for all SATS-36© components because the averages were above four. The highest average attitude component score was for Effort, and the lowest average was for Difficulty. This indicated that after taking SIEL students felt statistics was difficult to learn but did not require much work to learn. Although the results for Effort and Difficulty may seem contradictory, SIEL focused on statistical concepts which may be more difficult to learn, but concepts were presented using activities to make them more understandable to students requiring less effort to learn.

Table 4.17: Averages, standard deviations, standard errors, and Cronbach's α for post-course attitude component scores from SATS-36©

Attitude Component	Average	Standard deviation	Standard error	Cronbach's α
Affect	4.81	1.31	0.11	0.80
Cognitive Competence	5.38	1.21	0.10	0.80
Difficulty	4.35	0.89	0.07	0.82
Effort	5.81	0.97	0.08	0.86
Interest	4.65	1.12	0.09	0.81
Value	5.26	0.85	0.07	0.81

Internal consistency results for post-course SATS-36© attitude components.

Cronbach's α values were determined for each post-course attitude component of SATS-36© (Table 4.17) and ranged from 0.8 to 0.86. These values were similar to those obtained in Schau (2003) and Tempelaar et al. (2007). Based on the criteria established by George and Mallory (2003), post-course internal consistency for SATS-36© attitude components would meet the criteria of "good" internal consistency since alpha values were at least 0.8 but less than 0.9.

Correlation analyses involving post-course SATS-36© attitude component scores and demographic variables.

Demographic variables were correlated with post-course attitude component scores from SATS-36©. A complete list of these results can be found in Table 6 Appendix D. As with Table 2 Appendix E, the Bonferroni correction was used to produce a significance level of 0.006. Table 4.18 provides a summary of the significant correlations from Table 6 Appendix D.

Table 4.18: Significant* correlations between demographic variables and post-course SATS-36© attitude components

	Reported Math SAT score	Course grade	Course effort	Post-course Affect	Post-course Cognitive Competence	Post-course Difficulty	Post-course Effort	Post-course Value
Reported Math SAT score					0.39806			
Course grade				0.43058	0.51016	0.37855		0.35759
Course effort							0.38672	
Post-course Affect								
Post-course Cognitive Competence								
Post-course Difficulty								
Post-course Effort								

Post-course Value								

*Significant at the $p < 0.0006$ level.

Reported Math SAT score was correlated with post-course Cognitive Competence. Those with more mathematical skills which may be reflected by a higher Math SAT score tend to have a more positive attitude toward the difficulty of the course (e.g., they did not think the course was difficult).

Course grade was positively correlated with post-course SATS-36© attitude components: Affect, Cognitive Competence, Value, and Difficulty. Higher course grades were associated with having a more positive attitude toward statistics, toward their intellectual knowledge and skills needed for statistics, toward the worth and usefulness of statistics, and toward the difficultness of statistics. Similar results with Cognitive Competence and grades were found in Hilton et al. (2004). Unlike results from Dempster & McCorry (2009), post-course Affect and course grade were correlated.

It is interesting to note that course grade was not significantly correlated with attitude components: Interest and Effort. SIEL was developed to show the applicability of statistics in everyday life which the course developers believed would influence students' interest.

Course effort, defined as the percentage of activities completed during the semester, was correlated with post-course Effort. Since there was concern about students overestimating the amount of effort they may expend in a course, this relationship between post-course Effort and Course effort suggested that students were realistic in their evaluation of the amount of effort they would expend during the semester. Unlike results from Lalonde and Gardner (1993) but similar to results from Bude′ et al. (2007), Course effort and course grade were not significantly correlated.

Post-course SLCR scores.

Post-course student responses were evaluated based on awareness of each of the eight statistical literacy components in SLCR. Possible score values ranged from 0 to 3 with a higher score indicating a higher level of awareness of the component.

Counts and percentages of post-course statistical literacy scores for each statistical literacy component can be found in Tables 4.19, 4.20, and 4.21. For the Allstate advertisement, the mode post-course scores for each component was 0 except for Definition and Lurking Variable where the mode for these components was 1 (Table 4.19).

Table 4.19: Summary of SLCR scores post-course for Allstate advertisement (number for each score and percentage)

	Number of SLCR scores (%)			
Component	0	1	2	3
Bias	72 (50)	54 (37.5)	15 (10.42)	3 (2.08)
Causality	133 (92.36)	10 (6.94)	1 (0.69)	0 (0)
Definition	14 (9.72)	69 (47.92)	43 (29.86)	18 (12.5)
Generalize	134 (93.06)	6 (4.17)	3 (2.08)	1 (0.69)
Lurking variable	49 (34.03)	52 (36.11)	25 (17.36)	18 (12.5)
Method	78 (54.17)	48 (33.33)	12 (8.33)	6 (4.17)
Reported statistic	60 (41.67)	52 (36.11)	31 (21.53)	1 (0.69)
Variation	111 (77.08)	7 (4.86)	20 (13.89)	6 (4.17)

Table 4.20: Summary of SLCR scores pre-course and post-course for LiveActive advertisement (number for each score and percentage)

	Number of SLCR scores (%)			
Component	0	1	2	3
Bias	82 (56.94)	52 (36.11)	10 (6.94)	0 (0)
Causality	133 (92.36)	11 (7.64)	0 (0)	0 (0)
Definition	10 (6.94)	97 (67.36)	32 (22.22)	5 (3.47)
Generalize	107 (74.31)	21 (14.58)	12 (8.33)	4 (2.78)
Lurking variable	66 (45.83)	53 (36.81)	22 (15.28)	3 (2.08)
Method	24 (16.67)	69 (47.92)	37 (25.69)	14 (9.72)
Reported statistic	108 (75)	26 (18.06)	9 (6.25)	1 (0.69)
Variation	125 (86.81)	4 (2.78)	11 (7.64)	4 (2.78)

Table 4.21: Summary of SLCR scores pre-course and post-course for depression article (number for each score and percentage)

	Number of SLCR scores (%)			
Component	0	1	2	3
Bias	62 (43.06)	62 (43.06)	20 (13.89)	0 (0)
Causality	141 (97.92)	0 (0)	0 (0)	3 (2.08)
Definition	21 (14.58)	83 (57.64)	38 (26.39)	2 (1.39)
Generalize	95 (65.97)	10 (6.94)	34 (23.61)	5 (3.47)
Lurking variable	39 (27.08)	74 (51.39)	28 (19.44)	3 (2.08)
Method	21 (14.58)	75 (52.08)	46 (31.94)	2 (1.39)
Reported statistic	120 (83.33)	15 (10.42)	7 (4.86)	2 (1.39)
Variation	134 (93.06)	1 (0.69)	5 (3.47)	4 (2.78)

For the LiveActive advertisement, the mode for Definition and Method was 1. For all other statistical literacy components, the mode was 0 (Table 4.20).

For the Depression article, the mode of post-course statistical literacy component scores was 0 for all components except Definition, Lurking Variable, and Method. For these components, the mode was 1 (Table 4.21).

Averages, medians, standard deviations, and standard errors for average post-course SLCR component scores averaged across the two advertisements and article are presented in Table 4.22. The highest average and median component scores were for Definition, Lurking variable, and Method; these averages were slightly above 1 which indicated some awareness of the statistical literacy component. The lowest average component score was for Causality; a score barely above 0, an indication of essentially no awareness of the statistical literacy component.

Table 4.22: Averages, medians, standard deviations, and standard errors for average post-course SLCR component scores

Component	Average	Median	Standard deviation	Standard error
Bias	0.61806	0.6667	0.4614	0.0384
Causality	0.07407	0	0.1989	0.0166
Definition	1.2662	1	0.5134	0.0428
Generalize	0.38194	0	0.5124	0.0428
Lurking variable	0.92593	1	0.5412	0.0451
Method	1.03472	1	0.5226	0.0435
Reported statistic	0.44444	0.3333	0.4246	0.0354
Variation	0.29167	0	0.5389	0.0449

Table 4.14 provides average post-course SLCR scores for the two advertisements and article. The Variation component had the largest range of average post-course SLCR scores (3). Definition and Method had the second largest range of average post-course SLCR scores (2.6667) followed by Lurking variable and Reported statistic (2.3333). Causality had the smallest range (1) and the lowest standard deviation.

Method and Definition had the highest percentage of average post-course SLCR scores at an average of 1. These results suggested that statistical literacy components

Method, Lurking variable, Reported statistic, and Definition, were components in which students had the highest level of awareness after taking SIEL. Despite completing a course which emphasized statistical literacy, student scores were still low after taking SIEL, suggesting a lack of awareness of components essential to statistical literacy. Due to issues with inter-rater reliability, caution should be taken when interpreting results from Generalize and Reported statistics.

***Correlation analyses for post-course SLCR component scores with demographic variables and SATS-36© attitude component scores.*.**

Correlation analyses were performed between demographic variables and post-course average SLCR statistical literacy component scores. A complete list of results can be found in Table 7 Appendix D. Using a significance level of 0.0004 determined using the Bonferroni correction (alpha of 0.05 and 120 comparisons), no significant correlations were found. These results may be due in part to the low level of statistical significance that was used.

Table 8 Appendix D lists the complete results from correlation analyses between post-course average SLCR component scores and pre-course attitude components scores from SATS-36©. Using the Bonferroni correction, a significance level of 0.001 was determined based on an alpha of 0.05 and 48 comparisons, no significant correlations were observed.

Correlation analyses were performed on post-course average SLCR statistical literacy component scores and post-course attitude component scores from SATS-36©. A complete list of results can be found in Table 9 Appendix D. As in Table 8 Appendix D, a significance level of 0.001 was used, but again no significant correlations were observed.

Regression analyses for post-course average SLCR component scores.

For post-course statistical literacy component scores, significant independent variables were determined as described in Chapter 3 using a significance level of 0.1 to reduce

the chance of a Type II error (Table 4.23). The normality assumption was checked using normal probability plots, and these plots did not indicate concerns with the assumption of normality. The adjusted R^2 value, a measure of the percentage of variation in the dependent variable that is explained by the model, is provided for each regression model and appears in parentheses after the model.

Table 4.23: Significant* independent variables resulting from forward selection for post-course statistical literacy component scores

Statistical Literacy Component	Significant independent variable and p-values	Adjusted R^2 value
Bias	None	0
Causality	None	0
Definition	Gender (0.0786) Math SAT score (0.0578)	0.0813
Generalize	Interest Post-course (0.0148) Number of college statistics courses taken (0.0599) Affect Pre-course (0.0783)	0.1026
Lurking variable	Math SAT score (0.031) Effort Pre-course (0.0783)	0.069
Method	Total SAT score (0.0043) Gender (0.0283)	0.1161
Reported statistic	Gender (0.0333) Cumulative college points (0.0393)	0.0735

Variation	Effort Pre-course (0.0658) Number of high school statistics courses taken (0.0755) Affect Pre-course (0.0582) Interest Post-course (0.0603)	0.1045

*Significant at the $p < 0.1$ level.

Gender was a significant predictor of post-course Definition, Method, and Reported statistic. Math SAT score was a significant predictor of post-course Definition and Lurking variable, and Total SAT score was a significant predictor of Method. Post-course Variation was significantly predicted by the number of high school statistics courses taken while the number of college statistics courses was a significant predictor of post-course Generalize.

Three statistical literacy component average post-course scores were predicted by SATS-36© attitude component scores. Pre-course Effort, pre-course Affect, and Interest Post-course were predictors of Variation. Post-course Lurking variable was significantly predicted by pre-course Effort. Pre-course Interest was a significant predictor of post-course Generalize. Further research should focus on investigating these relationships.

Despite statistically significant predictors of post-course statistical literacy components, the adjusted R^2 values for the regression models were low. This indicates that there are other predictors of the statistical literacy components not accounted for in this study.

The following are the regression model equations (numbered 9 through 16) for each average post-course average statistical literacy component and significant independent variables listed in Table 4.23.

(9) Average Post-course Bias score = 0.6377

(10) Average Post-course Causality score = 0.0725

(11) Average Post-course Definition score = -0.1073 + 0.2427*Gender
+ 0.0021*Math SAT score

(12) Average Post-course Generalize score = 0.106 - 0.0872*Affect Pre-course
+ 0.1398*Interest Post-course + 0.1586*Number of college statistics courses

(13) Average Post-course Lurking variable score = 3.0198 – 0.1822*Effort
Pre-course – 0.0016*Math SAT score

(14) Average Post-course Method score = -0.0672 + 0.2195*Gender
+ 0.0008*Total SAT score

(15) Average Post-course Reported statistic score = 0.2551 + 0.2028*Gender
+ 0.0011*Cumulative college points

(16) Average Post-course Variation score = 1.5369 – 0.1257*Affect Pre-course
– 0.1669*Effort Pre-course + 0.0953*Interest Post-course – 0.22*Number of
high school statistics courses

For models (9) and (10), no independent variables were significant predictors of the dependent variable. Gender was a significant predictor in models (11), (14), and (15). In all these models, females had a higher expected post-course score in comparison to males. For model (11), the expected average post-course Definition score for females was 0.2427 points higher than the average for males. The expected post-course Method score (model 14) for females was 0.2195 points higher than the expected score for males. For model (15), the expected post-course Reported Statistics for females was 0.2028 points higher than that for males.

Model (13) is the only regression model in which all regression coefficients were negative. This indicated that for every 1-point increase in pre-course Effort score there was a 0.1822 decrease in expected post-course Lurking variable score. For every 1-point increase in reported Math SAT score there was a 0.0016 decrease in expected post-course Lurking variable score.

Post-course Generalize scores (model 12) were expected to decrease by 0.1586 point for each additional college statistics course taken. For each additional high school statistic course taken, post-course Variation scores were expected to decrease by 0.22 point. Further research is needed to explore these relationships between significant independent variables and the average post-course statistical literacy components.

Multivariate multiple regression model analysis for post-course average SLCR component scores.

Since post-course statistical literacy components may be dependent, a multivariate multiple regression model for the vector of post-course average statistical literacy component scores was created using all significant independent variables from the regression models listed in Table 4.23. Results from the Wilk's lambda test to determine if any of these independent variables had a significant effect on the vector of post-course average statistical literacy component scores can be found in Table 4.16. Significance was determined at a level of 0.1 to reduce the chance of a Type II error. Significant independent variables included gender, pre-course Effort, post-course Interest, and Total SAT score. Total SAT score may indicate that the person had been academically successful and is a significant predictor of the vector of statistical literacy component scores.

Two SATS-36© attitude components, pre-course Effort and post-course Interest, were the only attitude components that were significant predictors on the vector of post-course statistical literacy components.

Summary of Results for Research Question 1b.

The average SATS-36© attitude component scores after taking SIEL were above a neutral value of 4 indicating positive attitudes for all attitude components. The highest average score was observed for Effort, and the lowest score was observed for Difficulty. This indicated that after taking SIEL students felt that not much effort was needed to learn statistics. “Good” internal consistency results were observed based on criteria from George and Malloy (2003).

For the Allstate advertisement, the mode for Definition and Lurking variable was 1. The mode for all other statistical literacy components was 0. For the LiveActive advertisement, Method and Definition had modes of 1, but the mode for all other statistical literacy components was 0. For the depression article, a mode of 1 was observed for Definition, Lurking variable, and Method. All other statistical literacy components had a mode of 0. Despite the differences in context of the media articles, a mode score of 1 for Definition was observed for the three media articles.

SLCR scores were averaged across the media articles to produce average SLCR scores. The average scores ranged from 0.07 to 1.27. The highest average post-course scores were observed for Definition, Lurking variable, and Method while Causality had the lowest average score. Based on the low average statistical literacy scores, minimal to no awareness of statistical literacy components was observed despite completing SIEL. As seen with pre-course statistical literacy component scores, results from post-course average scores for Generalize and Reported statistic should be interpreted with caution due to issues related to inter-rater reliability.

A significant positive correlation was observed between reported Math SAT scores and post-course Cognitive competence. Those who thought they had the ability to learn statistics tended to have a higher Math SAT score. Course grade was correlated with post-course Affect, post-course Cognitive competence, post-course Difficulty, and post-course Value. A positive correlation was observed between Course effort and post-course Effort. This result suggests that students were realistic in their assessment of the amount of effort they would expend in SIEL.

No significant correlations were observed between post-course average SLCR scores and demographic variables. In addition, no significant correlations were found between post-course average SLCR scores and pre-course or post-course SATS-36© attitude scores. More research is needed to further investigate the relationships between statistical literacy and attitudes toward statistics.

From the results from regression analyses, gender effects were observed for Definition, Method and Reported Statistics. Females had higher scores than males for these components. Although statistically significant results were observed, the regression coefficients for the models were small indicating minimal effect on the dependent variables. Even though the results were statistically significant, practical significance is in question.

For the multivariate multiple regression model, significant independent variables included gender, pre-course Effort, post-course Interest, and Total SAT score. Despite statistically significant results from the regression and multivariate multiple regression models, the average post-course statistical literacy scores represented minimal to no awareness of the statistical literacy components. Therefore, these results are of little practical significance.

Research Question 1c

Overview.

The focus of Research Question 1c was to determine the change in the level of statistical literacy for college students who have taken a course on statistical literacy. In addition, the effect of gender, attitude, aptitude, and background upon this level was investigated. These results are presented and discussed in this section.

Results for the change in SATS-36© attitude scores.

Table 4.24 provides a summary of the average change in SATS-36© attitude component scores from post-course to pre-course. Differences in the average component scores for all attitude components were positive except for Effort and Interest. Similar results were observed in Carnell (2008).

Table 4.24: Averages, standard deviations, standard errors, and p-values for the change in SATS-36© attitude component scores (post-course minus pre-course)

Attitude Component	Average	Standard deviation	Standard error	p-value
Affect	0.33	1.15	0.10	0.001
Cognitive Competence	0.28	0.94	0.08	0.0006
Difficulty	0.39	0.71	0.06	<0.0001
Effort	-0.51	0.85	0.07	<0.0001
Interest	-0.15	1.02	0.09	0.0714
Value	0.07	0.74	0.06	0.2607

Paired t-tests were conducted to compare pre-course and post-course SATS-36© attitude component scores. Using a significance level of 0.05, all changes in SATS-36© attitude component scores were significant except for Interest and Value. These results indicated no effect of the course on the change in Interest (level of interest in statistics) or Value (attitude toward the usefulness and worth of statistics in their lives). Students may have already been interested in and valued statistics so that a lack of change would be expected. In contrast, Faghihi et al. (1995) found no significant changes in attitudes toward statistics using SATS-28© after taking an introductory statistics course. Carnell (2008) also found no significant changes in SATS-36© attitudes toward statistics when a data collection project was incorporated into a statistics course.

Results shown in Table 4.24 suggests that after taking SIEL students gained positive attitudes toward Affect (positive feeling about statistics), Cognitive Competence (positive attitude about intellectual knowledge and skills needed for statistics), and

Difficulty (positive attitude about the difficulty of statistics meaning that statistics was not viewed as difficult). Similarly in Cashin and Elmore (2005), Affect was the only SATS-36© component in which a change in attitude was observed.

A negative change in average Effort score suggested that students believed that they needed to put more effort into learning statistics after taking SIEL as compared to before taking SIEL. Two factors may have contributed to this result. First, before taking SIEL, students may have felt that the course would not be rigorous since it is a sophomore level course. Second, since SIEL is focused on concepts rather than computations, some students might have found that more effort was required to learn course material.

Although there were significant changes in SATS-36© attitude components after taking SIEL, the absolute value of these changes ranged from 0.0699 to 0.5124. Based on the 7-point Likert scale of SATS-36©, these changes are not meaningfully significant.

Pre-course and post-course attitude component scores from SATS-36© were correlated. A complete list of correlation results can be found in Table 10 Appendix D. Since 66 comparisons were correlated and an alpha of 0.05 was used, the significance level using the Bonferonni correction was determined to be 0.0008. Table 4.25 provides a summary of all significant correlations. Correlations listed in Table 4.25 were all positive.

Table 4.25: Significant* correlations between pre-course and post-course SATS-36© attitude components

	Pre Affect	Pre Cognitive competence	Pre Difficulty	Pre Effort	Pre Interest	Pre Value	Post Affect	Post Cognitive Competence	Post Difficulty	Post Effort	Post Interest	Post Value
Pre Affect		0.85	0.63		0.54	0.46	0.56	0.54	0.36		0.43	0.34
Pre Cognitive Competence			0.64		0.42	0.46	0.58	0.66	0.48		0.34	0.31
Pre Difficulty						0.35	0.48	0.47	0.62			0.29
Pre Effort										0.53		
Pre Interest						0.49	0.29				0.59	0.38
Pre Value							0.32	0.32			0.44	0.60
Post Affect\								0.88	0.67		0.50	0.48
Post Cognitive competence									0.74		0.34	0.46
Post Difficulty												
Post Effort												
Post Interest												0.65
Post Value												

- Significant at the $p < 0.0008$ level.

For each SATS-36© attitude component, respective pre-course and post-course components were significantly correlated. This indicated that as the attitude component score before SIEL increased so did post-course attitude component score.

Pre-course Affect was correlated with pre-course and post-course Cognitive Competence, pre-course and post-course Value, pre-course and post-course Difficulty, and pre-course and post-course Interest. These results suggested that the more students liked statistics prior to taking SIEL the more positive an attitude they had about their intellect to do well in statistics both before and after taking SIEL. The more students liked statistics prior to taking SIEL, the more they valued the worth of statistics before and after taking SIEL and the less difficult they thought learning statistics would be before and after taking SIEL. The more students liked statistics prior to taking SIEL, the more interested they were in the subject prior to and after taking SIEL.

It is interesting to note that the SATS-36© attitude component Effort was not significantly correlated with pre-course or post-course Affect. These results suggested that whether a person liked statistics prior to the course that this had no relationship to the amount of effort was believed to be needed pre-course or post-course. Similar results were observed for pre-course Cognitive Competence.

Pre-course Cognitive Competence was correlated with both pre-course and post-course attitude components, Affect, Difficulty, Interest, and Value but not Effort. Pre-course Effort and post-course Effort were only significantly correlated with each other but no other attitude components. This result may indicate that the attitude toward Effort is a unique attitude component and unrelated to the other SATS-36© attitude components.

Pre-course Difficulty was correlated with pre-course and post-course Affect, Cognitive Competence, and Value but not pre-course and post-course Interest and Effort. How difficult statistics is believed to be prior to the course is not associated with how interesting statistics is believed to be or the amount of effort believed to be needed to succeed in the course.

Pre-course Interest was correlated with pre-course and post-course Affect and Value. Those who like statistics may feel this way because they are interested in it and see its value.

Pre-course Value was correlated with all attitude components except pre-course and post-course Effort and post-course Difficulty. The value placed on learning statistics prior to SIEL is unaffected by the amount of effort that students believe they will need to expend to learn statistics.

Post-course Affect was correlated with all pre-course and post-course attitude components except Effort. This was the same trend as seen with pre-course Affect. Post-course Cognitive Competence was correlated with all pre-course attitude components except Effort and Interest and post-course Effort. Post-course Value was correlated with all pre-course attitude components except for Effort but it was significantly correlated with all post-course attitude components except for Difficulty and Effort. Similar trends were observed with pre-course Value.

The changes in average SLCR component scores from pre-course to post-course are presented in Table 4.26. All average score changes were positive. The highest average SLCR score changes were observed for Definition and Method while Causality had the lowest average score change. The highest median score was observed for Definition.

Table 4.26: Averages, medians, standard deviations, and standard errors for average change (post-course minus pre-course) in SLCR component scores

Component	Average	Median	Standard deviation	Standard error
Bias	0.28009	0.3333	0.5514	0.046
Causality	0.00694	0	0.214	0.0178
Definition	0.53472	0.6667	0.5916	0.0493
Generalize	0.16204	0	0.5347	0.0446
Lurking variable	0.21991	0.3333	0.7098	0.0592
Method	0.43287	0.3333	0.6018	0.0502
Reported statistic	0.18056	0	0.4922	0.041
Variation	0.19907	0	0.5593	0.0466

The results of paired t-tests performed using SAS version 9.2 (Cary, NC) to determine if the average post-course SLCR component scores were higher than the average pre-course SLCR component scores are provided in Table 4.27. Statistical significance was determined at an alpha of 0.05.

Table 4.27: Significant* results of paired t-tests to determine if average post-course SLCR component scores were higher than average pre-course SLCR component scores

Component	Test statistic	p-value
Bias	6.10	< 0.00005*
Causality	0.39	0.34875
Definition	10.85	< 0.00005*
Generalize	3.64	0.0002*
Lurking variable	3.72	0.00015*

Method	8.63	< 0.00005*
Reported statistic	4.40	< 0.00005*
Variation	4.27	< 0.00005*

*Significant at $p < 0.05$ level

For all SLCR components except Causality, there was a highly significant increase in the average component score after taking SIEL. This indicates that increases in awareness of these statistical literacy components were observed after taking SIEL.

A significant increase in awareness after taking SIEL was not observed for Causality. Several reasons may explain this result. First, "causality" may be a concept that requires more than one statistical literacy course to understand. A longer "digestion period" may be needed for students to grasp this concept. Finally, perhaps emphasis should be placed on "causality" in SIEL. Although statistically significant changes were observed, these changes do not reflect meaningful changes in the level of awareness of statistical literacy components since both pre-course and post-course average scores indicated minimal to no awareness of the components.

Change (post-course – pre-course) in SLCR scores.

Counts and percentages of scores from SLCR for each component of statistical literacy can be found in Tables 4.28, 4.29, and 4.30. Table 4.31 provides a summary of SLCR scores in which students received a score other than 0 prior to and after taking SIEL for the three media articles.

Table 4.28: Summary of SLCR scores pre-course and post-course for Allstate advertisement (number for each score and percentage)

	Number of SLCR scores (%)			
Component	0	1	2	3
Bias pre-course	97 (67.36)	44 (30.56)	1 (0.69)	2 (1.39)
Bias post-course	72 (50)	54 (37.5)	15 (10.42)	3 (2.08)
Causality pre-course	132 (91.67)	11 (7.64)	1 (0.69)	0 (0)
Causality post-course	133 (92.36)	10 (6.94)	1 (0.69)	0 (0)
Definition pre-course	51 (35.42)	67 (46.53)	18 (12.5)	8 (5.56)
Definition post-course	14 (9.72)	69 (47.92)	43 (29.86)	18 (12.5)
Generalize pre-course	140 (97.22)	1 (0.69)	2 (1.39)	1 (0.69)
Generalize post-course	134 (93.06)	6 (4.17)	3 (2.08)	1 (0.69)
Lurking variable pre-course	68 (47.22)	59 (40.97)	13 (9.03)	4 (2.78)
Lurking variable post-course	49 (34.03)	52 (36.11)	25 (17.36)	18 (12.5)
Method pre-course	119 (82.64)	17 (11.81)	5 (3.47)	3 (2.08)
Method post-course	78 (54.17)	48 (33.33)	12 (8.33)	6 (4.17)
Reported statistic pre-course	91 (63.19)	30 (20.83)	21 (14.58)	2 (1.39)
Reported statistic post-course	60 (41.67)	52 (36.11)	31 (21.53)	1 (0.69)
Variation pre-course	130 (90.28)	2 (1.39)	12 (8.33)	0 (0)
Variation post-course	111 (77.08)	7 (4.86)	20 (13.89)	6 (4.17)

Table 4.29: Summary of SLCR scores pre-course and post-course for LiveActive advertisement (number for each score and percentage)

	Number of SLCR scores (%)			
Component	0	1	2	3
Bias pre-course	100 (69.44)	41 (28.47)	3 (2.08)	0 (0)
Bias post-course	82 (56.94)	52 (36.11)	10 (6.94)	0 (0)
Causality pre-course	128 (88.89)	16 (11.11)	0 (0)	0 (0)
Causality post-course	133 (92.36)	11 (7.64)	0 (0)	0 (0)
Definition pre-course	58 (40.28)	66 (45.83)	18 (12.5)	2 (1.39)
Definition post-course	10 (6.94)	97 (67.36)	32 (22.22)	5 (3.47)
Generalize pre-course	120 (83.33)	19 (13.19)	5 (3.47)	0 (0)
Generalize post-course	107 (74.31)	21 (14.58)	12 (8.33)	4 (2.78)
Lurking variable pre-course	69 (47.92)	61 (42.36)	11 (7.64)	3 (2.08)
Lurking variable post-course	66 (45.83)	53 (36.81)	22 (15.28)	3 (2.08)
Method pre-course	60 (41.67)	68 (47.22)	14 (9.72)	2 (1.39)
Method post-course	24 (16.67)	69 (47.92)	37 (25.69)	14 (9.72)
Reported statistic pre-course	127 (88.19)	10 (6.94)	7 (4.86)	0 (0)
Reported statistic post-course	108 (75)	26 (18.06)	9 (6.25)	1 (0.69)
Variation pre-course	140 (97.22)	2 (1.39)	1 (0.69)	1 (0.69)
Variation post-course	125 (86.81)	4 (2.78)	11 (7.64)	4 (2.78)

Table 4.30: Summary of SLCR scores pre-course and post-course for Depression article (number for each score and percentage)

	Number of SLCR scores (%)			
Component	0	1	2	3
Bias pre-course	108 (75)	25 (17.36)	11 (7.64)	0 (0)
Bias post-course	62 (43.06)	62 (43.06)	20 (13.89)	0 (0)
Causality pre-course	144 (100)	0 (0)	0 (0)	0 (0)
Causality post-course	141 (97.92)	0 (0)	0 (0)	3 (2.08)
Definition pre-course	83 (57.64)	42 (29.17)	18 (12.5)	1 (0.69)
Definition post-course	21 (14.58)	83 (57.64)	38 (26.39)	2 (1.39)
Generalize pre-course	113 (78.47)	9 (6.25)	17 (11.81)	5 (3.47)
Generalize post-course	95 (65.97)	10 (6.94)	34 (23.61)	5 (3.47)
Lurking variable pre-course	52 (36.11)	73 (50.69)	14 (9.72)	5 (3.47)
Lurking variable post-course	39 (27.08)	74 (51.39)	28 (19.44)	3 (2.08)
Method pre-course	49 (34.03)	68 (47.22)	27 (18.75)	0 (0)
Method post-course	21 (14.58)	75 (52.08)	46 (31.94)	2 (1.39)
Reported statistic pre-course	135 (93.75)	4 (2.78)	4 (2.78)	1 (0.69)
Reported statistic post-course	120 (83.33)	15 (10.42)	7 (4.86)	2 (1.39)
Variation pre-course	141 (97.92)	0 (0)	2 (1.39)	1 (0.69)
Variation post-course	134 (93.06)	1 (0.69)	5 (3.47)	4 (2.78)

For the Allstate advertisement, the mode score for Definition was 1 both prior to and after taking SIEL (Table 4.28). The mode for all other components was 0 both pre-course and post-course with the exception of Lurking variable post-course. Prior to SIEL, the mode for Lurking variable was at 0 and then increased to 1 after SIEL. Lurking variable was the only statistical literacy component in which the value of the mode increased from pre-course to post-course.

For the LiveActive advertisement, Definition and Method had a mode score of 1 both pre-course and post-course (Table 4.29). All other components had a mode of 0 both prior to and after taking SIEL. No changes in modes were observed from pre-course to post-course with the LiveActive advertisement.

For the Depression article, the modes for Method and Lurking variable were 1 both prior to and after SIEL (Table 4.30). The only increase in percentage of scores from pre-course to post-course was observed for Definition. Prior to SIEL, the mode was observed at 0 and after SIEL the mode was 1.

These results provided insight into what statistical literacy components have been affected by SIEL based on the media article. Of course the argument may be made that the media article topic may have affected questions that were asked. Since the goal was to assess statistical literacy, providing different types of articles that might be encountered in everyday life was believed to best simulate real life.

The preferred trend when comparing post-course to pre-course SLCR score percentages would be to see a decrease in scores of 0 and an increase in scores of 1, 2, and 3. For the Allstate advertisement, this trend was observed for all statistical literacy components except Causality (Table 4.28). For Causality, there was an increase in scores of “0” and a decrease in scores of 1 while scores of 2 and 3 remained unchanged.

Like the Allstate advertisement, only Causality did not exhibit the “preferred trend” for the LiveActive advertisement (Table 4.29). In contrast, for the Depression article, the “preferred trend” was observed for all statistical literacy components including Causality (Table 4.30). Combined these results suggested that Causality is a component which is not affected by taking SIEL or that it is a component of statistical literacy that involves more advanced conceptualization that is not developed by SIEL or that context may be influential in assessing statistical literacy.

Table 4.31 provides an alternative way to present results from Tables 4.28, 4.29, and 4.30. In Table 4.31, percentages of SLCR scores which are not 0 are presented for each statistical literacy component both prior to and after taking SIEL and for each of the media articles. A score of 0 indicated no awareness of the statistical literacy component. The percentages in Table 4.31 represent the percentages of students who had some awareness of the statistical literacy component. For all media articles and statistical literacy components except for the depression article and Causality, the percentage with awareness of the statistical literacy component increased from pre-course to post-course.

Table 4.31: A summary of the percentage of SLCR scores not "0" prior to and after taking SIEL for the three media articles

	Percentage of scores not 0 for Allstate ad		Percentage of scores not 0 for LiveActive ad		Percentage of scores not 0 for Depression article	
Component	Pre-course	Post-course	Pre-course	Post-course	Pre-course	Post-course
Bias	32.64	50	30.56	43.06	25	56.94
Causality	8.33	7.64	11.11	7.64	0	2.08
Definition	64.58	90.28	59.72	93.06	42.36	85.42
Generalize	2.08	6.94	16.67	25.69	21.53	34.03
Lurking variable	52.78	65.97	52.08	54.17	63.89	72.92
Method	17.36	45.83	58.33	83.33	65.97	85.42
Reported statistic	36.81	58.33	11.81	25	6.25	16.67
Variation	9.72	22.92	2.78	13.19	2.08	6.94

In Table 4.14, average pre-course and post-course SLCR component scores from the two advertisements and article are presented. In comparing post-course to pre-course average SLCR component scores, the range of average scores increased for components Variation, Reported statistic, Definition, and Bias. The ranges for average SLCR scores for Causality, Generalize, Lurking variable, and Method did not change from pre-course to post-course.

The mode of average SLCR scores increased for Bias, Definition, Method, and Reported statistic when comparing post-course to pre-course average scores. These results suggested that taking SIEL improved awareness of Bias, Definition, Method, and Reported statistic components.

Figures 4.1 through 4.8 display the distributions of pre-course and post-course average scores for statistical literacy components in SLCR from Table 4.14. For all pre-course and post-course average component scores, the distributions of average scores were skewed to the right. This would be expected for pre-course average component scores since it would not be anticipated that before SIEL there would be many higher average component scores.

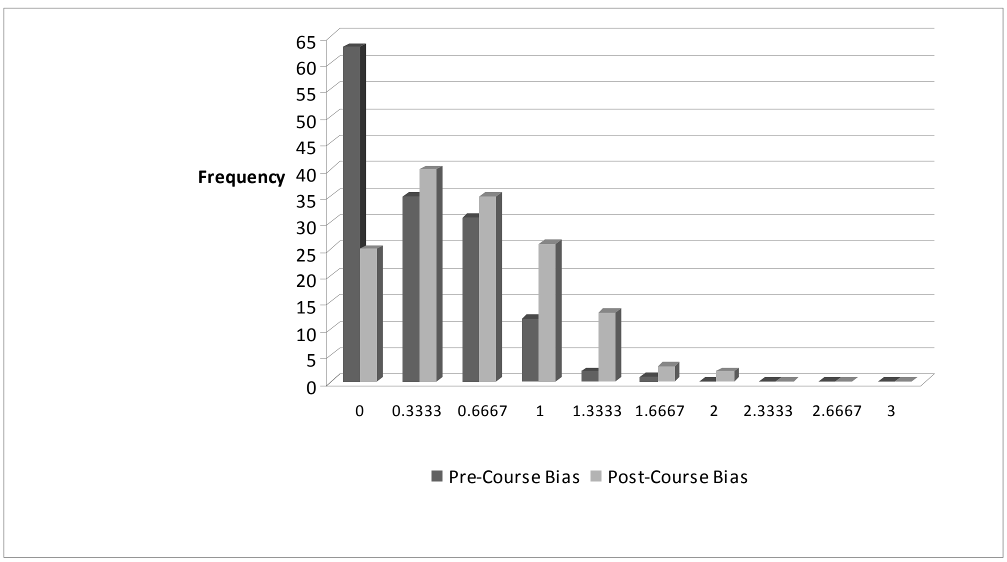


Figure 4.1: A comparison of pre-course and post-course average Bias scores

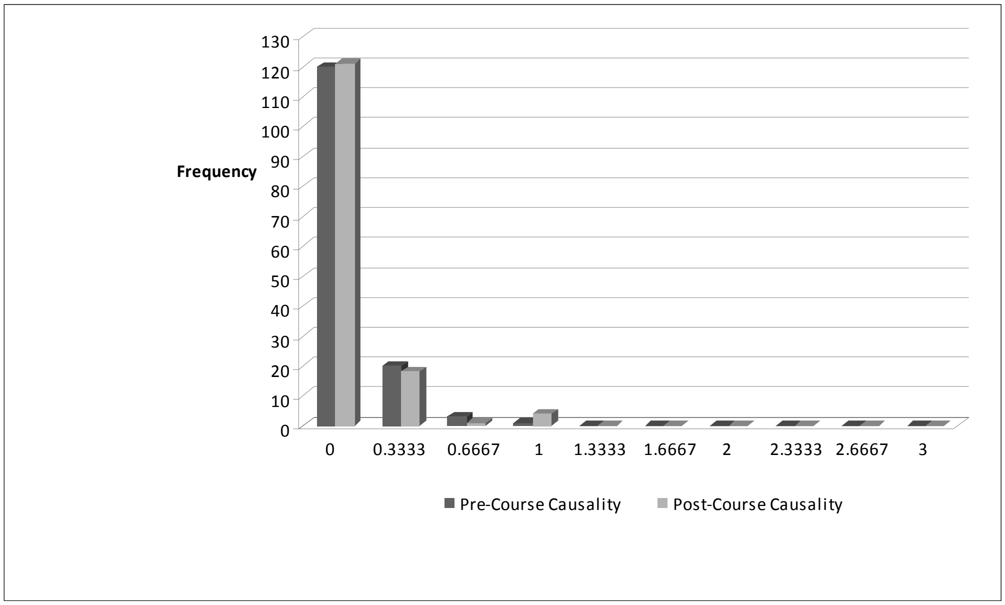


Figure 4.2: A comparison of pre-course and post-course average Causality scores

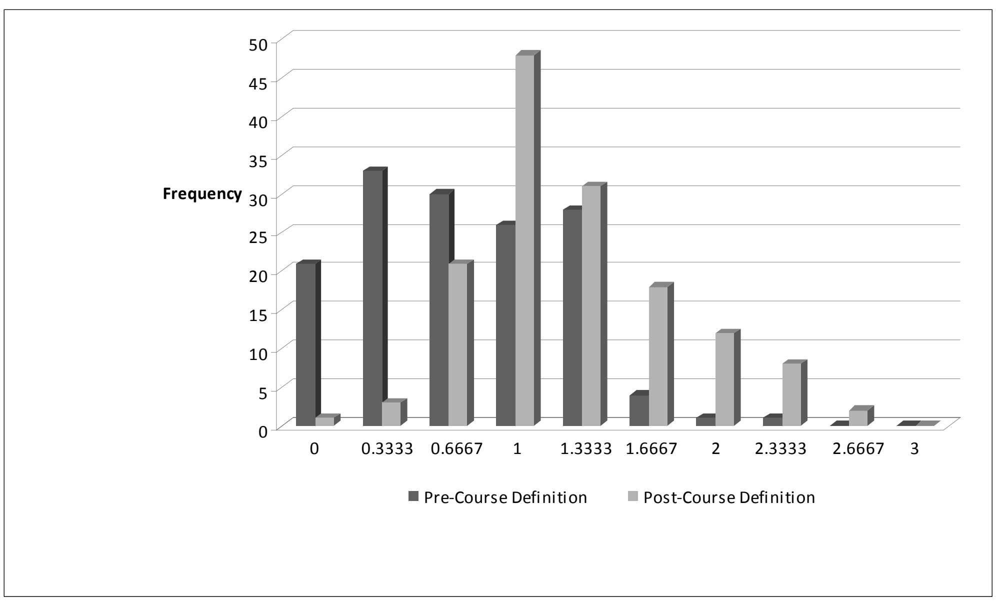


Figure 4.3: A comparison of pre-course and post-course average Definition scores

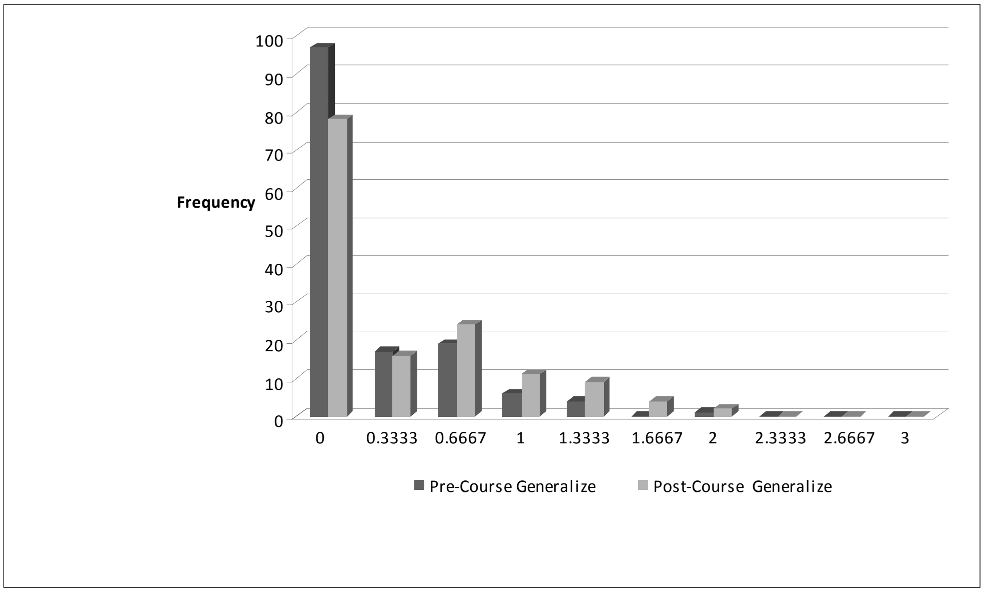


Figure 4.4: A comparison of pre-course and post-course average Generalize scores

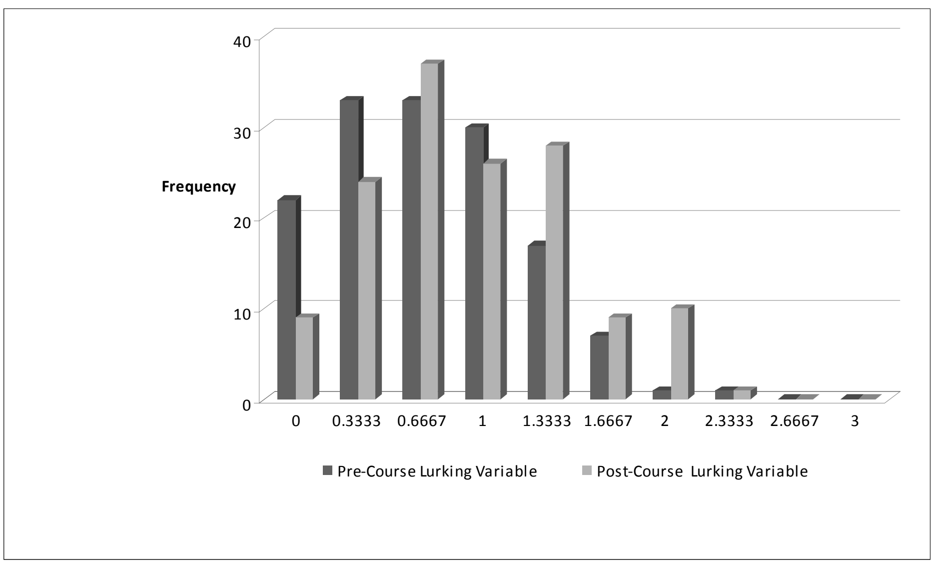


Figure 4.5: A comparison of pre-course and post-course average Lurking variable scores

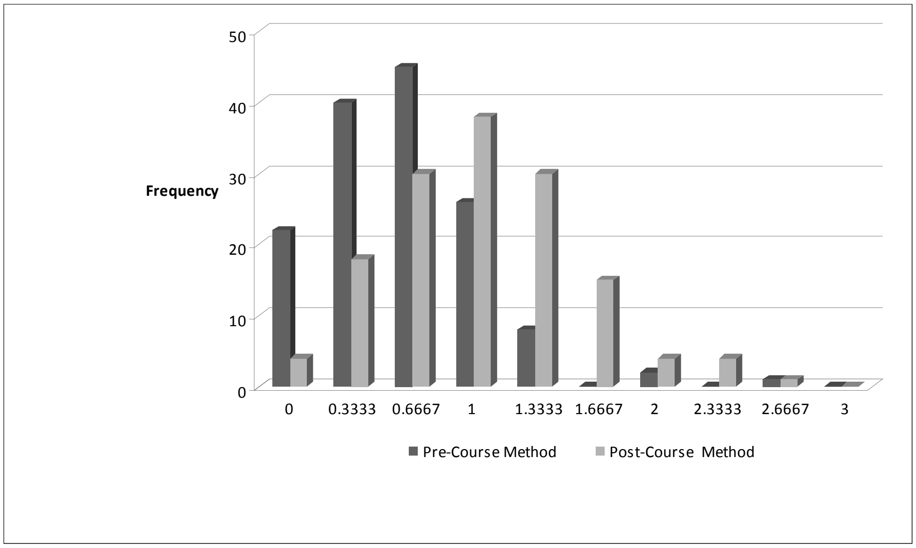


Figure 4.6: A comparison of pre-course and post-course average Method scores

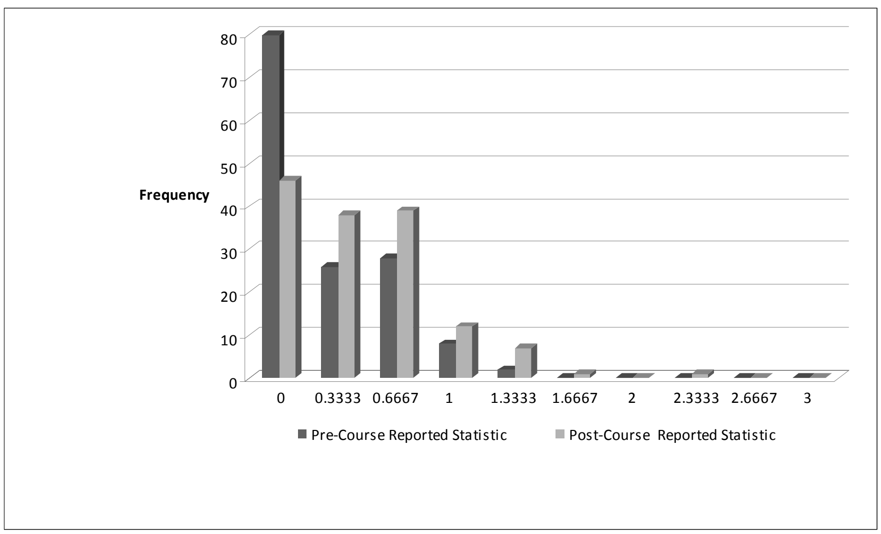


Figure 4.7: A comparison of pre-course and post-course average Reported statistic scores

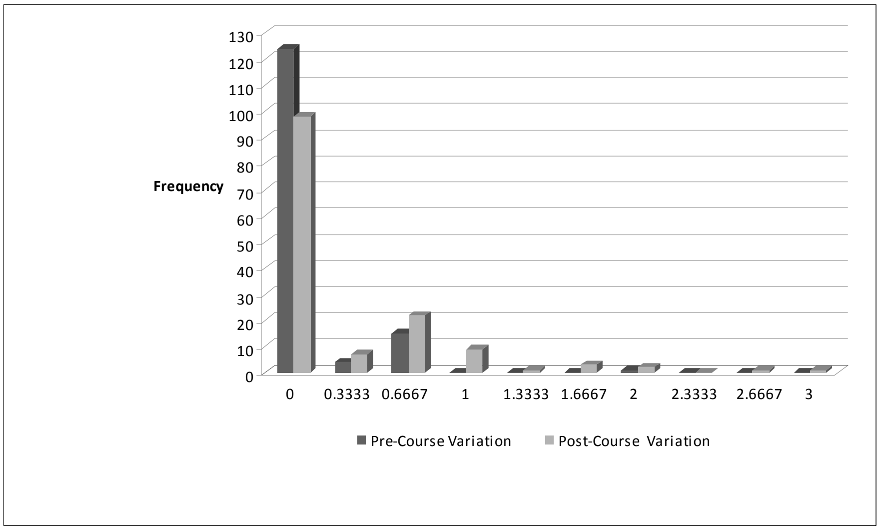


Figure 4.8: A comparison of pre-course and post-course average Variation scores

It was disappointing that, like pre-course average component scores, post-course average component scores were also skewed to the right indicating few high SLCR scores. After taking SIEL, it was hoped that the distributions of average post-course component scores would be skewed to the left indicating higher SLCR average scores and a higher level of awareness. Similar shapes for distributions between the average pre-course and post-course SLCR scores were observed for all components except Bias and Definition. Although the overall shapes for pre-course and post-course average Bias score distributions were skewed to the right, the average scores of 0 were not similar prior to and after taking SIEL. The post-course distribution of average SLCR scores for Definition was much less skewed than the pre-course distribution.

Correlation analyses involving the change in average SLCR component scores with demographic variables and SATS-36© attitude component scores.

The complete results from correlation analyses of demographic variables and the change (pre-course to post-course) in average SLCR statistical literacy component scores can be found in Table 11 Appendix D. A significance level of 0.004 was determined using the Bonferroni correction with an alpha of 0.05 and 120 comparisons. No significant correlations were observed.

The complete results from correlation analyses of pre-course SATS-36© attitude component scores and the change (post-course to pre-course) in average SLCR statistical literacy component scores can be found in Table 12 Appendix D. A significance level of 0.001 was determined using the Bonferroni correction in which alpha was set to 0.05 and 48 comparisons were performed. No significant correlations were observed.

The complete results from correlation analyses of post-course SATS-36© attitude component scores and the change (post-course to pre-course) in average SLCR statistical literacy component scores can be found in Table 13 Appendix D. The significance level used in Table 12 Appendix D of 0.001 was also used for Table 13 Appendix D. No significant correlations were observed. The small significance levels used in the correlation analyses had an effect on results.

Regression analyses for the change in average SLCR component scores.

To determine if independent variables were significant predictors of average changes in statistical literacy component scores, regression models were determined using the method described in Chapter 3. Results are displayed in Table 4.32. Statistical significance was determined at the 0.1 level to reduce the likelihood of Type II errors. The adjusted R^2 value, a measure of the percentage of variation in the dependent variable that is explained by the model, is provided for each regression model and appears in parentheses after the model.

Table 4.32: Significant* independent variables resulting from forward selection for the change in post-course and pre-course statistical literacy component scores

Statistical Literacy Component	Significant independent variable (p-value)	Adjusted R^2 value
Bias difference	Course Effort (0.0703) Number of high school statistics courses (0.0822) Effort Pre-course (0.0873)	0.0682
Causality difference	None	0
Definition difference	Gender (0.0147)	0.054
Generalize difference	Interest post-course (0.0091) Affect Pre-course (0.0012) Math SAT score (0.0761)	0.1787
Lurking variable difference	Effort Pre-course (0.0303) Difficulty Post-course (0.0365)	0.0765
Method difference	Total SAT score (0.0031) Gender (0.0374) Interest Pre-course (0.0578) Number of college statistics courses (0.0666)	0.1661
Reported statistic difference	Cumulative attempted credits (0.0809)	0.0227
Variation difference	Value Pre-course (0.02) Number of college math courses (0.0646) Affect Pre-course (0.0528) Gender (0.026) Number of college statistics courses (0.0831)	0.1629

*Significant at the $p < 0.1$ level.

Gender was a significant predictor of Definition difference, Method difference, and Variation difference. Bias difference was significantly predicted by the number of high school statistics courses taken. The number of college statistics courses taken was a significant predictor of Method difference, and the number of college math courses that were taken was a significant predictor of Variation difference. Attitude components from SATS-36© were significant predictors of Bias difference, Lurking variable difference, Generalize difference, Method difference, and Variation difference. Course effort was a significant predictor of Bias difference.

Despite statistically significant predictors of changes in statistical literacy components, the adjusted R^2 values for the regression models were low. This indicates that there are other predictors of the statistical literacy components not accounted for in this study.

The following are the regression model equations (numbered 17 through 24) for the average change in each statistical literacy component and significant independent variables listed in Table 4.31.

(17) Bias average difference = 0.2974 + 0.1265*Effort Pre-course
+ 0.2256*Number of high school statistics courses – 0.0104*Course Effort

(18) Causality average difference = -0.0036

(19) Definition average difference = 0.3778 + 0.3314*Gender

(20) Generalize average difference = 0.6415 – 0.1292*Affect Pre-course
+ 0.2059*Interest Post-course – 0.0014*Math SAT score

(21) Lurking variable average difference = 2.3647 – 0.2118*Effort Pre-course – 0.1888*Difficulty Post-course

(22) Method average difference = -1.2294 + 0.2288*Gender + 0.1006*Interest Pre-course + 0.0009*Total SAT score – 0.1524*Number of college statistics courses

(23) Reported statistics average difference = 0.09 + 0.0034*Cumulative attempted credits

(24) Variation average difference = 1.3928 – 0.2565*Gender – 0.155*Affect Pre-course – 0.1127*Value Pre-course + 0.1579*Number of college math courses – 0.1544*Number of college statistics courses

In models (19), (22), and (24), gender was a significant predictor of the dependent variable. For model (19), the expected change in average Definition score for females was 0.3314 point higher than that for males. Likewise, the expected change in average Method score for females was 0.2288 point higher than the score for males. The opposite relationship was observed for model (24) where the expected change in Variation score for females was 0.2565 point lower than the score for males.

A possible explanation for the result concerning the Definition component may be that this component involves verbal skills, and females may possess these skills more than males. An explanation for the Method and Variation components is not apparent.

Pre-course Effort had opposite effects in models (17) and (21). For model (17), for each 1-point increase in pre-course Effort score the change in average Bias score increased by 0.1265. For model (21) each 1-point increase in pre-course Effort score, the average change in Lurking variable component score decreased by 0.2118.

The number of college statistics courses taken was a significant predictor of the dependent variable in models (22) and (24). In both models, as the number of college statistics courses increased by one course, the average change in Method and Variation scores decreased. The amount of decrease in these models was similar. An explanation for this result may be that more statistics courses taken at higher levels may tend to emphasize statistical techniques and not statistical concepts such as evaluating the method used to obtain data or sources of variation.

The number of college mathematics courses taken was a significant predictor of the change in average Variation score. As the number of college mathematics courses increased by one course, the change in average Variation score increased by 0.1579 point.

The number of high school statistics courses taken and pre-course Effort were significant predictors of change in average Bias score (model 17). As the number of high school statistics courses increased by one course, the change in average Bias score increased by 0.2256 point. Bias may be a concept focused on in high school Statistics classes. As pre-course Effort increased by 1-point, the change in average Bias score decreased by 0.0104 point. The reason for this result is unclear.

Total SAT score was a significant predictor of the change in average Method score (model 22). As the Total SAT score increased by 1-point, the change in average Method score increased by 0.0009 points. This is an extremely small change. Cumulative attempted credits was a significant predictor of the change in average Reported statistic score (model 23). For each 1-point increase in cumulative attempted credit, the change in average Reported statistic score increased by 0.0034 points.

Several SATS-36© attitude component scores were significant predictors of change on average statistical literacy components. There was an inverse relationship between pre-course Affect score and the change in average Generalize (model 20) and Variation (model 24) scores. Inverse relationships were observed between post-course Difficulty and change in average Lurking variable score (model 21) and pre-course Value and change in average Variation score (model 24). For model 22, a 1-point increase in pre-course Interest scores the expected change in average Method score increased by 0.1006 point. Although statistically significant results were observed the actual change in SLCR scores is too small to be meaningful.

Multivariate multiple regression model analyses for the change in average SLCR component scores.

Since the change in average statistical literacy components scores may be dependent, a multivariate multiple regression model for the vector of changes in average statistical literacy component scores was created using all significant independent variables from the regression models listed in Table 4.32. Results from the Wilk's lambda test to determine if any of these independent variables had a significant effect on the vector of changes in average statistical literacy component scores can be found in Table 4.16. Significance was determined at a level of 0.1 to reduce the chance of Type II errors.

Significant variables included gender, post-course Interest, and Total SAT score. It is interesting to note that the vector of post-course statistical literacy components and the change in these components were significantly affected by the same variables except for pre-course Effort. Although the dependent variables used in the regression and multivariate multiple regression models represented statistically significant changes in scores, they are not meaningful changes in statistical literacy levels. Therefore, results from these analyses should be interpreted with caution.

Summary for Results for Research Question 1c.

SATS-36© attitude components, Affect, Cognitive competence, Difficulty, were found to have a significant positive changes after taking SIEL. This indicated that there were more positive attitudes concerning statistics (Affect), the intellectual knowledge and skills needed for statistics (Cognitive Competence), and difficulty of statistics (e.g. ease of statistics) (Difficulty). After taking SIEL, a significant negative change in Effort indicated that after taking SIEL students felt they needed to put more effort into the course in comparison to before taking SIEL. No significant changes were observed for Interest and Value. Although significant changes were observed for several SATS-36© attitude components, these changes were small and are of little practical significance.

Changes in average SLCR component scores were all positive and ranged from 0.007 to 0.535. The largest changes were observed for Definition and Method. All changes in statistical literacy components were statistically significant except Causality.

Although these changes were statistically significant, the magnitude of these changes was small which may not be of practical significance.

Pre-course and post-course SATS-36© attitude components were correlated. Effort was only correlated with itself and no other attitude components. More research concerning effort is needed to understand why this occurred. Pre-course Affect was positively correlated with all other attitude components except Effort. Post-course Value was positively correlated with all attitude components except Effort and post-course Difficulty.

No significant correlations were found between changes in SLCR component scores and demographic variables. Similarly, no significant correlations were found between changes in SLCR component scores and either pre-course or post-course SATS-36© component scores.

Results from regression analyses indicated that of the independent variables in this study none was predictors of the change in Causality score. Gender was a predictor of Variation change, Method change, and Definition change. For Method and Definition change, females had a higher score than males. Males had a higher score than females for Variation change. More research into effects of gender on statistical literacy components should be conducted. Results from the multivariate multiple regression analysis indicated that gender, post-course Interest, and Total SAT score had a significant effect on the vector of changes in statistical literacy components. Since both pre-course and post-course average scores represented at most minimal awareness of statistical literacy components, changes in these scores are not meaningful, and model results should be interpreted with caution.

Overall Summary for Research Question 1

The focus of Research Question 1 was to assess the level of statistical literacy prior to and after taking SIEL as well as changes in statistical literacy. The effect of attitude toward statistics, aptitude (such as Math, Verbal, and Total SAT scores), gender, and background (such as number of high school and college math and statistics courses taken) on the level of statistical literacy were investigated. The following are the summary of highlights from the results of this research question.

The highest average SATS-36© attitude component score both prior to and after taking SIEL was for Effort, and the lowest score was for Difficulty. The Effort attitude component was unique since it was the only attitude component in SATS-36© that was not correlated with any other attitude component. The change in average SATS-36© attitude component scores were all positive except for Effort and Interest. Significant changes in attitude components were observed for all components except Interest and Value.

No significant correlations were observed between average SATS-36© attitude component scores and average SLCR statistical literacy scores either prior to or after SIEL. In addition, no significant correlations were observed between post-course SLCR component scores or changes in SLCR scores and demographic variables. The only significant correlation was observed between the numbers of college statistics courses taken and average pre-course Variation score. Since many comparisons were made, the Bonferroni correction was used to reduce the chance of Type I errors. The small significance level used had an effect on correlation results.

Although the contexts of the advertisements and article varied, similar trends in scores for statistical literacy components were observed. The highest mode for any component was 1. Prior to SIEL, a mode of 1 was observed for Definition for the Allstate and LiveActive advertisements. In addition, a mode of 1 was observed for Method for the LiveActive advertisement and depression article. Lurking variable had a mode of 1 for the depression article. After SIEL, a mode of 1 for Definition was observed for all three media articles instead of just the advertisements as seen prior to SIEL. Method had a mode of 1 for the LiveActive advertisement and depression article; this was the same result as prior to SIEL. A mode of 1 was observed for Lurking variable for the Allstate advertisement and depression article. Prior to SIEL, a mode of 1 for Lurking variable was only observed for the depression article. Inter-rater reliability statistics indicated that caution should be taken when interpreting results concerning Generalize and Reported statistic.

Regression and multivariate multiple regression analyses involving pre-course, post-course and changes from post-course to pre-course average SLCR scores did detect significant predictors. However, most coefficients in these models were small and not of practical significance. In addition, the adjusted R^2 values for all regression models were low indicating other predictors of average SLCR scores not accounted for in this study.

Overall, results of Research Question 1 indicated that changes in attitudes toward statistics were observed after taking SIEL except for Interest and Value. Pre-course and post-course SLCR scores indicated minimal to no awareness of statistical literacy components. After taking SIEL, changes in all statistical literacy components except Causality were also observed. Although changes in statistical literacy scores were statistically significant, they were small in magnitude and are of little practical significance.

SIEL had an effect on attitudes toward statistics but only minimal to no effect on the level of awareness of statistical literacy components. Several factors may have contributed to this result. The difficulty of SLCR scoring with the open-ended format may have had an effect. To achieve more than a minimal level of statistical literacy, more than one course on statistical literacy may be required. Also, becoming statistically literate may require more time to develop than a few months during a semester.

In the next section, Research Question 2 is described and results are presented. Discussions of results are provided.

Research Question 2

Overview

Students were given two advertisements and an article prior to and after taking SIEL and asked what concerns they had about the information in the advertisements and article and why these concerns were important to ask. Students' responses were grouped into topic categories based on similar content.

The overall purpose of Research Question 2 was to observe concerns (topic categories) that students raised before and after taking a course on statistical literacy when confronted with claims from the media that make generalizations. The Allstate advertisement, LiveActive advertisement, and the article on depression (Appendix A) were the three media articles used for Research Question 2.

This research question was subdivided into four specific research questions. These included the following:

2a. Prior to taking a course on statistical literacy, what areas of concern (topic categories) were raised when viewing claims from the media that make generalizations?

2b. After taking a course on statistical literacy, what areas of concern (topic categories) were raised when viewing claims from the media that make generalizations?

2c. Were there changes in areas of concern (topic categories) that were raised from pre-course to post-course?

2d. Were these areas of concern (topic categories) different for males and females?

In the following sections, results from each of these questions are presented and discussed with respect to each media article. A summary of results for Research Question 2 is then provided.

Results for concerns to the Allstate advertisement.

Areas of concern prior to taking SIEL for the Allstate advertisement.

Table 4.33 provides a list of concerns (topic categories) for the Allstate advertisement that were observed prior to and after taking SIEL as well as changes from pre-course to post-course. The most common concern (58.33%) pertained to what definition was used for an "accident" or the type of accident (fender bender vs. cars totaled). Other common concerns included: what source was used to obtain "6 million" (29.17%), how the information was obtained (20.14%), what factors might have affected accidents (18.06%), what effect the definition of "car accident" might have had on the value "6 million" (15.97%), whether unreported accidents were included in the accident count (15.97%), whether people have Allstate insurance (11.11%), and whether Allstate had an agenda (10.42%).

Table 4.33: Percentage of responses for pre-course, post-course, and changes from post-course to pre-course for topic categories for the Allstate advertisement

Topic Category	% responded pre-course	% responded post-course	Change in %
Accuracy or Reliability of 6 million?	6.94	13.19	6.25
Accurate to count if not causes injury or damages?	3.47	0.69	-2.78
Accurate to compare to other countries?	2.08	0	-2.08
Accurate to count if not driver's fault?	0	0.69	0.69
Agenda for Allstate?	10.42	13.19	2.77
Better place definition?	9.03	4.86	-4.17
Biased?	4.17	8.33	4.16
Cause of accident?	8.33	2.08	-6.25
Correlation?	0.69	0.69	0
Coverage error?	0	1.39	1.39
Dark figure? (unreported cases)	0.69	4.86	4.17

Definition of America?	5.56	9.72	4.16
Definition of car?	1.39	8.33	6.94
Definition or type of accident?	58.33	84.72	26.39
Effect of definition of America on 6 million	2.08	4.17	2.09
Effect of definition of car accident on 6 million	15.97	30.56	14.59
Effect of definition of car on 6 million	0.69	2.08	1.39
Effect of having insurance on driving	1.39	0	-1.39
Effect of unreported accidents on 6 million	4.86	8.33	3.47
Exact number?	5.56	12.5	6.94
Factors affecting 6 million?	0.69	0.69	0
Factors affecting accidents?	18.06	12.5	-5.56
Generalizable?	0	0.69	0.69
Have Allstate?	11.11	3.47	-7.64
Have insurance and reduce accidents?	8.33	7.64	-0.69
How was information obtained?	20.14	60.42	40.28
Includes insured and uninsured drivers?	0.69	2.78	2.09
Location of accidents?	7.64	15.28	7.64
Lurking variable?	0	4.17	4.17
Measurement error?	0.69	2.78	2.09
Misconception	0.69	2.78	2.09
MOE or CI included?	0	10.42	10.42
Number of accidents varies	6.25	8.33	2.08
Number of cars or number of accidents?	5.56	4.17	-1.39
Number of cars, drivers, or accidents?	6.25	7.64	1.39
Number of reported accidents?	0.69	0	-0.69
Other ways to report?	3.47	1.39	-2.08
Outliers?	0	1.39	1.39
Parameter	0	0.69	0.69
Parameter or statistic?	0	1.39	1.39
Percent of cars or drivers on road	0.69	0.69	0
Percent of drivers not in accident?	0.69	0	-0.69
Picture in background	0.69	0	-0.69
Population or population size?	0.69	4.86	4.17
Probability of being in accident?	0	0.69	0.69
Question if could count all accidents	0.69	4.86	4.17
Question if number of accidents is really 6 million	5.56	6.94	1.38
Question wording	0	2.78	2.78

Reducing number of accidents	0.69	0	-0.69
Relativeness of 6 million	8.33	9.72	1.39
Relevance of 6 million	0	2.08	2.08
Representative?	2.08	1.39	-0.69
Round figure	7.64	13.89	6.25
Sample size?	1.39	4.17	2.78
Six million is an average	8.33	9.03	0.7
Six million is an estimate	4.17	8.33	4.16
Six million is correct	2.08	0	-2.08
Source of 6 million?	29.17	39.58	10.41
Too many definition	0.69	5.56	4.87
True for all insurance companies?	0.67	0	-0.67
Unreported accidents included?	15.97	30.56	14.59
Untrustworthy	2.08	4.86	2.78
Vague claim	0	0.69	0.69
Validity	4.17	6.25	2.08
Which years was study conducted?	7.64	15.28	7.64
World vs. America	3.47	0	-3.47

Areas of concern after taking SIEL for the Allstate advertisement.

After taking SIEL, the most common concern was questioning how "accident" was defined and what types of accidents were involved in this definition (fender benders vs. cars totaled) (84.72%). The percentage of responses concerning how the information in the advertisement was obtained was 60.42% while questioning the source of "6 million" was a concern for 39.58% of students. For topic categories, effect of definition of car accidents on "6 million" and whether unreported accidents were included in the number of accidents, 30.56% of students were concerned about each of these issues.

Change in areas of concern after taking SIEL for the Allstate advertisement.

The most common concern both prior to and after taking SIEL pertained to the definition of "accident" and the type of accident involved in this definition. The number of topic categories in which the percentage of responses was at least 10% increased from 8 to 13. This suggests that students' responses included more concerns about the advertisements after SIEL. Increases in percentages of responses from pre-course to post-course were observed (Table 4.33). The greatest improvement in response changes, denoted by the largest positive change in the percentage of responses from pre-course to post-course, was observed for the topic category in which how the information was obtained was questioned (40.28%). Other noteworthy improvements were seen for questioning the definition or type of accident (26.39%) and whether unreported accidents were included (14.59%). An increase of 10.42% was observed for questioning about the inclusion of the margin of error (MOE) or confidence intervals and the source of 6 million. The positive increases in percent change for these topic categories are important since these categories pertain to vital issues addressed in SIEL.

Although a positive percent change from pre-course to post-course indicated an increase in awareness of a topic category, this result was not applicable to topic categories that were less relevant to the media claims. For example, the largest decrease (-7.64%) in the percentage of responses from pre-course to post-course was observed for the topic category that involved questioning whether people had Allstate insurance. Questioning if people had Allstate insurance is a less relevant topic category to the claim.

To determine if there were changes in the patterns of concerns about topic categories from pre-course to post-course, McNemar's test was performed for each observed topic category in Table 4.33. More specifically, McNemar's tests determined if the proportions that did not mention the concern pre-course but did mention the concern post-course (p_{01}) were different from the proportions that did mention the concern pre-course but did not mention the concern post-course (p_{10}). The following hypotheses were tested:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

Table 4.34 provides a summary of all significant results from the McNemar's tests for Allstate topic categories. Statistical significance was determined at a level of 0.05. The complete list of results from McNemar's test for all Allstate topic categories can be found in Table 1 Appendix E.

Table 4.34: Significant† results from McNemar's tests for topic categories for the Allstate advertisement

Topic Category	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Cause of accident?	129	0	3	2.08	12	8.33	0.0352
Definition of car?	130	0	12	8.33	2	1.39	0.0129
Definition or type of accident?	14	76	46	31.94	8	5.56	<0.0001
Effect of definition of car accident on 6 million	90	13	31	21.53	10	6.94	0.0015
Have Allstate?	124	1	4	2.78	15	10.42	0.0192
How was information obtained?	48	20	67	46.53	9	6.25	<0.0001
Location of accidents?	115	4	18	12.50	7	4.86	0.0433
Lurking variable?	138	0	6	4.17	0	0.00	0.0313
MOE or CI included?	129	0	15	10.42	0	0.00	<0.0001
Source of 6 million?	71	26	31	21.53	16	11.11	0.04
Too many definition	135	0	8	5.56	1	0.69	0.0391
Unreported accidents included?	90	13	31	21.53	10	6.94	0.0015

† Significant at the $p < 0.05$ level
* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL

n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-values for comparing p_{01} and p_{10}

Topic categories for the Allstate advertisement that had significant changes observed from pre-course to post-course included: “cause of accident,” “definition of car,” “definition or type of accident,” “effect of definition of car accident on 6 million,” “have Allstate,” “how was information obtained,” “location of accidents,” “lurking variables,” “MOE or confidence interval,” “source of 6 million,” “too many definition,” and “unreported accidents included.” Of these topic categories, “cause of accident,” and “have Allstate” had a higher proportion who gave a response before SIEL but not after SIEL ($\hat{p}_{10}$) as compared to the proportion who did not give a response before SIEL but did give a response after SIEL ($\hat{p}_{01}$). Since these topic categories are not relevant, these results are viewed as positive changes.

For the remaining topic categories with a significant change from pre-course to post-course “definition or type of accident,” “definition of car,” “effect of definition of car accident on 6 million,” “how was information obtained,” “location of accidents,” “lurking variables,” “MOE or confidence interval,” “source of 6 million,” “too many,” and “unreported accidents,” the proportion who did not give a response before taking SIEL but did after SIEL ($\hat{p}_{01}$) was higher than the proportion who gave a response before taking SIEL but not after SIEL ($\hat{p}_{10}$). Topic categories, “definition or type of accident,” “definition of car,” and “definition of ‘too many’,” pertained to concerns centered on definitions. The category, “MOE or confidence interval,” referred to concerns about variation. Topics including “location of accidents,” “lurking variables,” “effect of definition of car accident on 6 million,” and “unreported accidents included” reflected issues that may have an effect on the value of the statistic. Categories such as “how was information obtained,” and “source of 6 million” pertained to data gathering issues. Definitions, data collection, variation, and factors affecting statistics are important issues discussed in SIEL. Being able to apply these concepts to everyday situations is a primary goal of SIEL.

Differences in concerns based on sex for the Allstate advertisement.

Pre-course differences in concerns for males and females for the Allstate advertisement..

Hypothesis tests were performed to determine if there were differences in the proportion of males and females who had concerns for each topic category. The following hypotheses were tested:

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

where p_f represented the proportion of females who mentioned the category pre-course for the Allstate advertisement and p_m represented the proportion of males who mentioned the category pre-course for the Allstate advertisement.

The complete list of results can be found in Table 2 Appendix E. A significance level of 0.05 was used. P-values were determined from Chi-square tests except when expected cell counts were fewer than five. In these cases, p-values from Fisher's exact tests were reported and designated in Table 2 Appendix E with an "F" after the p-value.

Prior to taking SIEL, there was only one topic category in which there was a significant difference in the proportion of responses for males and females. A significant difference in the proportion of males and females who had concerns dealing with "correlation" was observed (p-value=0.0092). Females had a higher percentage (1.19%) with concerns about "correlation" than males (0%). It is unclear why this result occurred. Since many statistical tests were performed, it would be expected that some of the tests would be significant. Future research should focus on whether sex differences actually exist.

Post-course differences in concerns for males and females for the Allstate advertisement.

To determine if there were differences in the proportion of males and females who had a concern about a particular topic category, the following hypotheses were tested:

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

where p_f represented the proportion of females who mentioned the category post-course for the Allstate advertisement and p_m represented the proportion of males who mentioned the category post-course for the Allstate advertisement.

Statistical significance was determined at a level of 0.05, and significant results are presented in Table 4.35. The complete list of results for post-course topic categories can be found in Table 3 Appendix E. P-values were determined from Chi-square tests except when expected cell counts were less than five. In these cases, p-values from Fisher's exact tests were reported and designated in Table 4.35 and Table 3 Appendix E with an "F" after the p-value.

Table 4.35: Significant* post-course percentages for topic categories by sex for the Allstate advertisement

Topic Category	Percent female	Percent male	p-value for difference**
Definition of America?	4.76	16.67	0.0174
Question if could count all accidents	8.33	0	0.0415 F

*Significant at a $p < 0.05$ level
**p-values from chi-square tests were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

After taking SIEL, there were two categories in which significant differences in the proportions of males and females that had concerns for the Allstate advertisement was observed (Table 4.35). It is unclear why these differences occurred even given the large number of tests performed.

Changes in concerns based on sex for the Allstate advertisement.

To determine if there were changes in the patterns of responses from pre-course to post-course for males and females, McNemar's test was performed for each topic category using the following hypotheses:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

where p_{01} represented the proportion of females/males who did not mention the category pre-course but did mention the category post-course for the Allstate advertisement and p_{10} represented the proportion of females/males who mentioned the category pre-course but did not mention the category post-course for the Allstate advertisement.

Statistical significance was determined at a level of 0.05, and all significant results can be found in Table 4.36. A complete list of results is presented in Table 4 Appendix E.

Table 4.36: Significant† results from McNemar's tests by sex for topic categories for the Allstate advertisement

Topic Category	Sex	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Cause of accident?	Female	73	0	1	1.19	10	11.90	0.0117
Dark figure? (unreported cases)	Female	78	0	6	7.14	0	0.00	0.0313
Definition of America	Male	50	4	6	10.00	0	0.00	0.0313
Definition or type of accident?	Female	7	44	28	33.33	5	5.95	<0.0001
	Male	7	32	18	30.00	3	5.00	<0.0001
Effect of definition of car accident on 6 million	Female	51	8	20	23.81	5	5.95	0.0041
Have Allstate insurance?	Female	70	1	3	3.57	10	11.90	0.0192
How was information obtained?	Female	28	12	41	48.81	3	3.57	<0.0001

	Male	20	8	26	43.33	6	10.00	<0.0001
Location of accidents?	Male	49	2	8	13.33	1	1.67	0.0391
MOE or CI included?	Female	74	0	10	11.90	0	0.00	0.002
	Male	55	0	5	8.33	0	0.00	<0.0001
Population or population size?	Female	78	0	6	7.14	0	0.00	0.0313
Question if could count all accidents?	Female	77	0	7	8.33	0	0.00	0.0156
Source of 6 million?	Female	40	18	19	22.62	7	8.33	0.029
Too many definition	Female	78	0	6	7.14	0	0.00	0.0313
Unreported accidents included?	Female	50	11	20	23.81	3	3.57	<0.0001

†Significant at a $p < 0.05$ level
* n_{00} represents the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represents the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represents the number of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
n_{10} represents the number of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represents the percentage of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represents the percentage of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

*** Exact p-values for comparing p_{01} and p_{10}

Based on the results from McNemar's tests, there were significant changes in response patterns for three topic categories for both males and females. These categories included "definition or type of accident," "how was information obtained," and "MOE or confidence interval." For each of these categories and for both males and females, the proportion that did not mention the category pre-course but did mention the category post-course ($\hat{p}_{01}$) was higher than the proportion that mentioned the category pre-course but did not mention the category post-course ($\hat{p}_{10}$). These categories represent important topics (eg. definitions, methodology, and variation) discussed in SIEL.

For topic categories, "definition of America" and "location of accidents," there were significant changes in the pattern of responses from pre-course to post-course for males only. For these categories, the proportion that did not mention the category pre-course but did mention the category post-course ($\hat{p}_{01}$) was higher than the proportion that mentioned the category pre-course but did not mention the category post-course ($\hat{p}_{10}$).

Significant changes in the pattern of responses were observed for several topic categories for females only. These categories included "cause of accidents," "dark figure," "effect of definition of car on 6 million," "population or population size," "question if could count all accidents," "source of 6 million," "too many definition," and "unreported accidents included." For all of these categories except "cause of accidents," the proportion that did not mention the category pre-course but did mention the category post-course ($\hat{p}_{01}$) was higher than the proportion that mentioned the category pre-course but did not mention the category post-course ($\hat{p}_{10}$). A similar result for "cause of accidents" was observed in Table 4.34.

It is unclear why significant results were observed for one sex versus the other sex. However, given the large number of tests that were performed, it is not surprising that some significant results were found. Future research may address whether these results reflect true gender differences.

Summary of results for the Allstate advertisement

The most common topic category for the Allstate advertisement both prior to and after taking SIEL concerned the definition and type of accident. Prior to SIEL, the only topic category with differences in the percentages of responses for males and females was observed for “correlation.” The percentage of female responses was higher than the percentage of male responses for this category. After SIEL, differences in the percentages of responses for males and females were observed for two Allstate topic categories, “Definition of America” and “Question if could count all accidents.”

Significant changes in response patterns from pre-course to post-course were observed for “cause of accident,” “definition of car,” “definition or type of accident,” “effect of definition of car accident on 6 million,” “have Allstate,” “how was information obtained,” “location of accidents,” “lurking variables,” “MOE or confidence interval,” “source of 6 million,” “too many definition,” and “unreported accidents included.”

For both males and females, significant changes in response patterns were observed for the following topic categories: “definition or type of accident,” “how was information obtained,” and “MOE or confidence interval.” These results suggested that SIEL influenced student questioning of these important topic categories regardless of gender. These topic categories are fundamental concepts covered in SIEL.

More topic categories were observed to have significant changes in response patterns for females as compared to males. For males, there was a significant change in response patterns for only two topic categories: “definition of America” and “location of accidents.” In comparison for females, there was a significant change in response patterns for eight topic categories: “cause of accidents,” “dark figure,” “effect of definition of car on 6 million,” “population or population size,” “question if could count all accidents,” “source of 6 million,” “too many definition,” and “unreported accidents included.” These results indicated that different topic categories were questioned by males and females and that females questioned about more topic categories than males. Future research should focus on difference in what topic categories are questioned based on sex.

Results for concerns to the LiveActive advertisement

As with the Allstate advertisement, the LiveActive advertisement was given to students to voice concerns about both prior to and after SIEL. Students’ concerns about the advertisement were grouped into topic categories based on similar content. The results for the LiveActive advertisement are presented and discussed in this section.

Areas of concern prior to taking SIEL for the LiveActive advertisement.

Table 4.37 lists percentages of responses for topic categories that were raised by students prior to and after taking SIEL, as well as changes in these responses from pre-course to post-course. The most common topic category (concern) students had regarding the LiveActive advertisement prior to SIEL pertained to factors that might affect irregularity (45.14%) and how information from the study was obtained (42.36%) (Table 4.37). Other common topic categories included questions about the sample size used (34.03%), the premise of the advertisement (27.08%), and the source of the value “4 out of 5” (24.31%).

Table 4.37: Percentage of responses for pre-course, post-course, and change from post-course to pre-course for topic categories for the LiveActive advertisement

Topic Category	% responded pre-course	% responded post-course	Change in %
Accuracy and Reliability?	14.58	13.19	-1.39
Agenda?	9.03	11.81	2.78
Amount of cottage cheese to eat?	0	0.69	0.69
Be that other woman definition	0	0.69	0.69
Bias	6.94	13.89	6.95
Causality	11.11	7.64	-3.47
Effect of definition occasional irregularity on 4 of 5	5.56	19.44	13.88
Effect of definition of occasional on 4 of 5	4.17	6.94	2.77
Effect of factors on 4 of 5	18.06	18.75	0.69
Effect of how information obtained on 4 of 5	8.33	27.08	18.75
Effect of location on irregularity	3.47	5.56	2.09
Effect of lurking variable on 4 of 5	0	2.78	2.78
Factors affecting irregularity	45.14	40.97	-4.17
Generalize?	7.64	6.94	-0.7
How did they get information?	42.36	78.47	36.11
Length of study?	0	0.69	0.69
Location?	11.11	15.97	4.86
Lurking variable	0.69	13.89	13.2
Measurement error	3.47	10.42	6.95
Men not included	0.69	1.39	0.7
Misconception	6.25	9.72	3.47
Misrepresenting	0.69	0	-0.69
Misunderstand ad	7.64	9.72	2.08
MOE/CI included?	0.69	11.81	11.12
Non-response	0	7.64	7.64
Other woman question	2.08	2.78	0.7
Population	2.78	6.25	3.47
Question premise of article	27.08	22.22	-4.86
Question wording	0.69	15.97	15.28
Relativeness of 4 of 5	1.39	0	-1.39
Relevance of 4 of 5	0	0.69	0.69
Representative?	7.64	16.67	9.03
Round number	0	2.08	2.08
Sample size?	34.03	43.06	9.03
Sensitive topic	0	6.94	6.94

Source of 4 of 5	24.31	18.75	-5.56
Spoon size	0	1.39	1.39
Statistic and parameter	0	2.78	2.78
Untrustworthy	0.69	1.39	0.7
Vague claim	4.17	9.03	4.86
Validity	9.72	6.94	-2.78
Variation	1.39	1.39	0
When study conducted?	0.69	2.78	2.09
Which yogurt for comparison?	1.39	5.56	4.17
Why study conducted?	0.69	1.39	0.7

Areas of concern after taking SIEL for the LiveActive advertisement.

After taking SIEL, the most common topic category students had regarding the LiveActive advertisement pertained to how the advertisers obtained the information (78.47%) (Table 4.37). This concern was by far the most common since topic categories with the next highest percentages were related to the size of the sample (43.06%) and factors that affected irregularity (40.97%). Other common topic categories included the effect of how information was obtained on "4 of 5" (27.08%), questions concerning the premise of the advertisement (22.22%), and questioning the source of "4 of 5" (18.75%). The topic categories with the three highest percentages (how information was obtained, sample size, and factors affecting irregularity) are methodological concerns and are topics stressed in SIEL.

Change in areas of concern after taking SIEL for the LiveActive advertisement.

In comparing pre-course and post-course percentages found in Table 4.37, the number of topic categories in which the percentage of responses was at least 10%

increased from 9 to 17. This indicates a broader scope of topic categories in students' responses after SIEL.

The most notable improvements in the percentages of students who had concerns about topic categories, denoted by large positive changes, included how information was obtained (36.11%), the effect of how information was obtained on "4 of 5" (18.75%), and concerns about question wording (15.28%) (Table 4.37). These topic categories focused on method.

Based on the information contained in Table 4.37, the largest decrease (-5.56%) in the percentage of responses from pre-course to post-course was observed for the topic category that involved questioning the source of the value "4 out of 5." Although the source of the information is important to understanding motivations of the study and agendas, a decrease in the percentage questioning this topic category after SIEL is not seen as negative since the source is not an especially important topic category.

To determine if there were changes in the pattern of responses from pre-course to post-course, McNemar's test was conducted for each topic category. The following hypotheses were tested:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

where p_{01} represented the proportion who did not mention the category pre-course but did mention the category post-course for the LiveActive advertisement and p_{10} represented the proportion of who mentioned the category pre-course but did not mention the category post-course for the LiveActive advertisement.

Statistical significance was determined at a level of 0.05. Table 4.38 displays all significant results of the McNemar's tests. A complete list of results can be found in Table 5 Appendix E.

Table 4.38: Significant† results from McNemar's tests for topic categories for the LiveActive advertisement

Topic Category	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Effect of definition of occasional irregularity on 4 of 5	113	5	23	15.97	3	2.08	<0.0001
Effect of how information obtained on 4 of 5	100	7	32	22.22	5	3.47	<0.0001
How did they get information?	20	50	63	43.75	11	7.64	<0.0001
Lurking variable	123	0	20	13.89	1	0.69	<0.0001
Measurement error	126	2	13	9.03	3	2.08	0.0213
MOE/CI included?	127	1	16	11.11	0	0.00	<0.0001
Non-response	133	0	11	7.64	0	0.00	<0.0001
Question wording	120	0	23	15.97	1	0.69	<0.0001
Representative?	113	4	20	13.89	7	4.86	0.0192
Sensitive topic	134	0	10	6.94	0	0.00	0.002

†Significant at a $p < 0.05$ level

* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL

n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL

n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL

n_{10} represented the number of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL

$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-values for comparing p_{01} and p_{10}

There were significant changes in patterns of responses from pre-course to post-course for ten topic categories. These categories included "effect of definition of occasional irregularity on 4 of 5," "effect of how information was obtained on 4 of 5," "how did they get information," "lurking variables," "measurement error," "MOE or confidence interval," "non-response," "question wording," "representative," and "sensitive topic." For all of these categories, the proportion that did not have a response pre-course but did have a response post-course ($\hat{p}_{01}$) was higher than the proportion that did have a response pre-course but did not have a response post-course ($\hat{p}_{10}$). These topic categories concern effects of definitions and methods on statistics, study methods, factors not considered in the study such as lurking variables and non-response, variation, and the sensitive nature of the topic. These are important topics stressed in SIEL.

Differences in concerns based on sex for the LiveActive advertisement.

Pre-course differences in concerns for males and females for the LiveActive advertisement..

Hypothesis tests were performed to determine if there were differences in the proportions of males and females who had concerns about each of the observed topic categories from the LiveActive advertisement prior to taking SIEL. The following hypotheses were tested:

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

where p_f represented the proportion of females who mentioned the category pre-course for the LiveActive advertisement and p_m represented the proportion of males who mentioned the category pre-course for the LiveActive advertisement.

A significance level of 0.05 was used. Significant results from McNemar's tests are presented in Table 4.39. A complete list of results can be found in Table 6 Appendix E. P-values reported in Table 4.39 and Table 6 Appendix E were determined from Chi-square tests. When the expected cell counts were less than five, Fisher's exact test was performed. P-values from Fisher's exact test were denoted with an "F" after the p-value in both tables.

Table 4.39: Significant† pre-course percentages for topic categories by sex for the LiveActive advertisement

Topic Category	Percent female	Percent male	p-value for differences*
Causality	5.95	18.33	0.0198
Effect of factors on 4 of 5	23.81	10	0.0337
Factors affecting irregularity	55.95	30	0.002

†Significant at a $p < 0.05$ level
*Exact p-values were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

Prior to SIEL, there were significant differences in the proportions of males and females who had concerns for "causality," "effect of factors on 4 out of 5," and "factors affecting irregularity." For the two latter topic categories, the percentage of females with concerns was higher than the percentage of males. In contrast, the percentage of males was higher than the percentage of females for the causality topic category. It is unclear why these differences existed. Again, given the large number of tests, we would expect some to be significant. Future research should determine whether gender differences actually exist.

Post-course differences in concerns for males and females for the LiveActive advertisement.

Hypothesis tests were performed to determine if there were differences in the proportion of males and females who had concerns about each observed topic category for the LiveActive advertisement after taking SIEL. The following hypotheses were tested:

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

where p_f represented the proportion of females who mentioned the category post-course for the LiveActive advertisement and p_m represented the proportion of males who mentioned the category post-course for the LiveActive advertisement.

A significance level of 0.05 was used. Significant results are listed in Table 4.40 with a complete list of results in Table 7 Appendix E. P-values reported in Table 4.40 and Table 7 Appendix E were determined from Chi-square tests. When the expected cell counts were less than five, Fisher's exact test was performed. P-values from Fisher's exact test were denoted with an "F" after the p-value in both tables.

Table 4.40: Significant† post-course percentages for topic categories by sex for the LiveActive advertisement

Topic Category	Percent female	Percent male	p-value for differences*
Effect of factors on 4 of 5	26.19	8.33	0.0068
How did they get information?	84.52	70	0.0366
Representative?	22.62	8.33	0.0233

†Significant a $p < 0.05$ level
*Exact p-values were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

After taking SIEL, there were significant differences between the proportions of males and females who had concerns about the "effect of factors on 4 of 5", "how information in the advertisement was obtained" and if the sample was "representative." These topic categories pertain to method. For each of these topic categories, the percentage of females with concerns was higher than the percentage of males. It is unclear why these differences were observed or whether it was due to the large number of tests performed.

Changes in concerns based on sex for the LiveActive advertisement.

To determine if there were changes in patterns of responses from pre-course to post-course for males and females, McNemar's test was performed for each topic category using the following hypotheses:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

where p_{01} represented the proportion of females/males who did not mention the category pre-course but did mention the category post-course for the LiveActive advertisement and p_{10} represented the proportion of females/males who mentioned the category pre-course but did not mention the category post-course for the LiveActive advertisement.

Statistical significance was determined at a level of 0.05. Table 4.41 presents all significant results from McNemar's tests for comparing topic category responses for males and females. A complete list of results can be found in Table 8 Appendix E.

Table 4.41: Significant† results from McNemar's tests by sex for topic categories for the LiveActive advertisement

Topic Category	Sex	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Effect of definition of occasional irregularity on 4 of 5	Female	66	2	14	16.67	2	2.38	0.0042
	Male	47	3	9	15.00	1	1.67	0.0215
Effect of how information obtained on 4 of 5	Female	54	3	24	28.57	3	3.57	<0.0001
How did they get information?	Female	9	34	37	44.05	4	4.76	<0.0001
	Male	11	16	26	43.33	7	11.67	0.0013
Lurking variable	Female	69	0	14	16.67	1	1.19	0.001
	Male	54	0	6	10.00	0	0.00	0.0313
Measurement error	Female	72	2	9	10.71	1	1.19	0.0215
MOE/CI included?	Female	74	0	10	11.90	0	0.00	0.002
	Male	53	1	6	10.00	0	0.00	0.0313
Non-response	Female	76	0	8	9.52	0	0.00	0.0078
Question wording	Female	68	0	16	19.05	0	0.00	<0.0001
Representative?	Female	63	3	16	19.05	2	2.38	0.0013
Sensitive topic	Female	77	0	7	8.33	0	0.00	0.0156

†Significant at a $p < 0.05$ level
* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-value for comparing p_{01} and p_{10}

Topic categories in which there was a significant change in patterns of responses from pre-course to post-course for both males and females included: "effect of definition of occasional irregularity on 4 of 5," "how did they get their information," "lurking variable," and "MOE or confidence interval." For these categories, the proportion that did not have a response for the topic category before SIEL but did have a response for the topic category after SIEL ($\hat{p}_{01}$) was higher than the proportion that did have a response before taking SIEL but did not have a response after taking SIEL ($\hat{p}_{10}$).

No significant differences in the percentage of responses for a topic category were observed for males only. In contrast, there were several topic categories in which significant changes in pattern of responses from pre-course to post-course were observed for females. These topics included "effect of how information obtained on 4 of 5," "measurement error," "non-response," "question wording," "representative," and “sensitive topic.” For these categories, the proportions of females that did not have a response before taking SIEL but did have a response after taking SIEL ($\hat{p}_{01}$) was higher than the proportion of females that had a response before taking SIEL but did not have a response after taking SIEL ($\hat{p}_{10}$).

The significant topic categories represent important concepts covered in SIEL, including method, lurking variable, and variation issues. It is unclear why significant results were observed for one sex versus the other. Given the large number of tests that were performed it is not surprising that some significant results were found. Future research should focus on understanding differences in statistical literacy for males and females.

Summary of results for the LiveActive advertisement

The most common topic categories prior to SIEL were factors affecting irregularity and how the information for the advertisement was obtained. After SIEL, the most common topic category was how the information in the advertisement was obtained.

Prior to SIEL, there were differences in the percentages of responses for several topic categories for males and females. These categories included “causality,” “effects of factors on 4 of 5,” and “factors affecting irregularity.” For the causality topic category, the percentage of males who had a response in this category was higher than the percentage of females. For categories, “effects of factors on 4 of 5” and “factors affecting irregularity,” the percentage of responses for females was higher than the percentage of responses for males.

After SIEL, there were differences in the percentages of responses for several topic categories for males and females. These categories included the “effect of factors on 4 of 5,” “how information was obtained,” and whether the sample was “representative.” For each of these categories, the proportion of females with responses was higher than the proportion of males.

It is noteworthy that both prior to and after SIEL that gender differences were observed for the topic category pertaining to the “effects of factors on 4 of 5.” It is unclear why this occurred. One explanation may be that since the advertisement was targeted at women females may have related to this issue and had questions concerning this category. Since many tests were performed and the Bonferroni correction was not used, the probability of a Type I error was inflated. Future research could be directed at whether advertisements targeted at a particular sex have different effects on results by sex.

Changes in patterns of responses were observed for the following topic categories: “effect of definition of occasional irregularity on 4 of 5,” “effect of how information was obtained on 4 of 5,” “how they obtained information,” “lurking variables,” “MOE or confidence interval,” “non-response,” “question wording,” “representative,” and “sensitive topic.” For all of these categories, the proportion that did not have a response pre-course but did have a response post-course ($\hat{p}_{01}$) was higher than the proportion that did have a response pre-course but did not have a response post-course ($\hat{p}_{10}$).

Changes in the pattern of responses for topic categories for both males and females were observed. These topic categories included: “effect of definition of occasional irregularity on 4 of 5,” “how did they get their information,” “lurking variable,” and “MOE or confidence interval.” For these categories, the proportion that did not have a response for the topic category before SIEL but did have a response for the topic category after SIEL ($\hat{p}_{01}$) was higher than the proportion that did have a response before taking SIEL but did not have a response after taking SIEL ($\hat{p}_{10}$).

No significant differences in the percentage of responses were observed for males only. Several topic categories had significant changes in patterns of responses for females only. These topics included “effect of how information obtained on 4 of 5,” “measurement error,” “non-response,” “question wording,” “representative,” and “sensitive topic.”

Results for concerns to the depression article

As with the Allstate and LiveActive advertisements, an article concerning depression during teen years was given to students prior to and after SIEL. Students were asked to voice concerns about the information in the article. These concerns were grouped into topic categories based on similar content. The results are presented in the following section.

Areas of concern prior to taking SIEL for the depression article.

Table 4.42 provides a list of concerns for topic categories that students raised with the depression article prior to taking SIEL. The most common concern prior to SIEL pertained to "how the information in the article was obtained" (52.08%). Other common concerns included "factors affecting depression" (42.36%), "definition of major depressive episode" (38.89%), "location of the study" (27.08%), and the "number of boys and girls in the study" (25.69%). For three topic categories, "effect of factors on results," "effect of how information obtained on the statistics in the article," "questioning the premise of the article," the percentage for each was 23.61%.

Table 4.42: Percentage of responses for pre-course, post-course, and change from post-course to pre-course for topic categories for the depression article

Topic Category	% responded pre-course	% responded post-course	Change in %
Accuracy/Reliability	18.06	13.89	-4.17
Agenda	1.39	4.17	2.78
At start of the study	0.69	0.69	0
Average of 4 years	0.69	0	-0.69
Bias	6.25	14.58	8.33
Boys depression rate	2.78	3.47	0.69
Correlation - cause and effect	0	2.08	2.08
Coverage error	0	3.47	3.47
Dark figure	0	2.08	2.08
Definition of adolescence	0	2.08	2.08
Definition of anxiety	0.69	0	-0.69
Definition of major	0	0.69	0.69
Definition of major depressive episode	38.89	81.94	43.05
Definition of substantial	0.69	0.69	0
Effect of definition of depression	16.67	45.14	28.47
Effect of factors on results	23.61	22.92	-0.69
Effect of girls admit more than boys	4.17	2.08	-2.09
Effect of how information obtained	23.61	39.58	

on statistics			15.97
Effect of location on results	18.75	20.14	1.39
Effect of lurking variables	0	6.25	6.25
Effect of number of boys and girls	18.75	21.53	2.78
Effect of non-teen years included	2.08	3.47	1.39
Effect of non-response	0	1.39	1.39
Effect of question asked on results	4.86	20.83	15.97
Effect of same number of gender by age	5.56	4.86	-0.7
Effect of when study done on results	0	4.86	4.86
Errors	0	4.86	4.86
Exact number or percent?	0	3.47	3.47
Extraneous variable	0	2.08	2.08
Factors affecting depression	42.36	36.81	-5.55
Generalize	12.5	17.36	4.86
Girls more likely to admit depression	14.58	25.69	11.11
Girls vs. boys social situation	1.39	0	-1.39
How was information obtained?	52.08	73.61	21.53
Includes non teen years	2.78	5.56	2.78
Inconsistent	0.69	0	-0.69
Journal credibility	0	0.69	0.69
Last statement concerns	2.08	2.78	0.7
Length of study?	1.39	2.08	0.69
Location of study?	27.08	29.86	2.78
Lurking variable	0.69	13.19	12.5
Measurement error	12.5	31.94	19.44
Misconception	5.56	6.94	1.38
MOE/CI included?	1.39	7.64	6.25
Multiple depressive episodes definition	0.69	0	-0.69
Non-response error	0	5.56	5.56
Not understand statistics given	1.39	1.39	0
Number of boys and girls?	25.69	43.75	18.06
One in 10 each year during study?	0.69	0	-0.69
Population or population size?	0.69	4.86	4.17
p-value and alpha	0	0.69	0.69
Qualification of researcher?	0.69	2.08	1.39
Question 2 times statistic	1.39	4.17	2.78
Question premise of article	23.61	6.94	-16.67
Questions asked?	6.94	28.47	21.53

Relevance?	0.69	0.69	0
Representative?	6.94	18.06	11.12
Results for different age groups?	0	0.69	0.69
Same number of boys and girls in each age?	6.25	5.56	-0.69
Sample size?	6.25	9.72	3.47
Sensitive subject	0	1.39	1.39
Simpson's Paradox	0	0.69	0.69
"Some time in her life" concern	0	0.69	0.69
Statistic vs. Parameter	0	0.69	0.69
Trustworthy?	1.39	2.08	0.69
US vs. Canada depression rate	1.39	1.39	0
Vague claim	0.69	3.47	2.78
Validity	5.56	10.42	4.86
When did study occur?	1.39	5.56	4.17
Who funded/conducted study?	2.08	4.17	2.09

Areas of concern after taking SIEL for the depression article.

After taking SIEL, the most common concerns were the definition of "major depressive episode" (81.94%) and "how the information in the article was obtained" (73.61%) (Table 4.42). Other frequently expressed concerns included questioning the "effect of the definition of depression on the results" (45.14%), the "number of boys and girls in the study" (43.75%), the "effect of how the information was obtained on the resulting statistics" (39.58%), "what factors might have affected depression" (36.81%), "measurement error" (31.94%), the "location of the study" (29.86%), "what questions were asked" (28.47%), "girls more likely to admit to depression than boys" (25.69%), the "effect of the number of boys and girls on results" (21.53%), and the "effect of the questions asked on results" (20.83%).

Change in areas of concern after taking SIEL for the depression article.

The number of topic categories with percentage of responses at least 10% increased from 14 to 20. This suggests a broader scope of topic categories observed from pre-course to post-course (Table 4.42).

When comparing pre-course and post-course percentages of responses from Table 58, the largest improvement in percentages of responses, denoted by a large positive increase in difference in percentages, was observed for the topic category dealing with the "definition of major depressive episode" (43.05%). Other improvements in topic categories were observed for questioning the "effect of definition of depression" (28.47%), "how information was obtained" (21.53%), "questions asked" (21.53%), and "measurement error" (19.44%). These topic categories pertain to important concepts concerning methods and definitions and their effects on results.

The largest decrease (-16.67%) in the percentage of responses from pre-course to post-course was observed for the topic category that involved "questioning the premise of the article." This result was important since it indicated a decrease in concerns about topics that are irrelevant to the claims being made in the article.

To determine if there were changes in patterns of responses from pre-course to post-course, McNemar's test was performed for each topic category. The following are the hypotheses that were tested:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

where p_{01} represented the proportion who did not mention the category pre-course but did mention the category post-course for the depression article and p_{10} represented the proportion of who mentioned the category pre-course but did not mention the category post-course for the depression article.

Statistical significance was determined at the 0.05 level. The statistically significant results are presented in Table 4.43. A complete list of results can be found in Table 9 Appendix E.

Table 4.43: Significant† results from McNemar's tests for topic categories for the depression article

Topic Category	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Bias	117	3	18	12.50	6	4.17	0.0227
Definition of major depressive episode	26	56	62	43.06	0	0.00	<0.0001
Effect of definition of depression	72	17	48	33.33	7	4.86	<0.0001
Effect of how information obtained on statistics	70	17	40	27.78	17	11.81	0.0032
Effect of lurking variables	135	0	9	6.25	0	0.00	0.0039
Effect of question asked on results	110	3	27	18.75	4	2.78	<0.0001
Effect of when study done on results	137	0	7	4.86	0	0.00	0.0156
Errors	137	0	7	4.86	0	0.00	0.0156
Girls more likely to admit depression	98	12	25	17.36	9	6.25	0.009
How was information obtained?	25	62	44	30.56	13	9.03	<0.0001
Lurking variable	124	0	19	13.19	1	0.69	<0.0001
Measurement error	89	9	37	25.69	9	6.25	0.0000
MOE/CI included?	132	1	10	6.94	1	0.69	0.0117
Non-response error	136	0	8	5.56	0	0.00	0.0078

Number of boys and girls?	66	22	41	28.47	15	10.42	0.0007
Question premise of article	104	4	6	4.17	30	20.83	<0.0001
Questions asked?	99	6	35	24.31	4	2.78	<0.0001
Representative?	113	5	21	14.58	5	3.47	0.0025

†Significant at a $p < 0.05$ level
* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-value for comparing p_{01} and p_{10}

There were statistically significant changes in the pattern of responses from pre-course to post-course for the following topic categories: “bias,” “definition of major depressive episode,” “effect of definition of depression,” “effect of how information obtained on statistics,” “effect of lurking variables,” “effect of question asked on results,” “effect of when the study was done on results,” “errors,” “girls more likely to admit depression,” “how was information obtained,” “lurking variable,” “measurement error,” “MOE or confidence interval included,” “non-response error,” “number of boys and girls,” “question premise of advertisement,” “questions asked,” and “representative.” Many of these topic categories pertained to effects of definitions or factors related to data collection on the statistics. Other topics related to bias and errors such as measurement error, margin of error, or non-response error.

The category, “question premise of advertisement,” was the only category in which the proportion that had responses in a topic category before SIEL but not after ($\hat{p}_{10}$) was higher than the proportion that did not have responses in a topic category before SIEL but did have a response after SIEL ($\hat{p}_{01}$). For the other categories that had a significant difference, the proportion that did not have a response before taking SIEL but did have a response after taking SIEL ($\hat{p}_{01}$) was higher than the proportion that did have a response before SIEL but not a response after SIEL ($\hat{p}_{10}$). These topic categories were discussed in SIEL so these results suggest more awareness of these concepts.

Differences in concern based on sex for the depression article.

Pre-course differences in concerns for males and females for the depression article.

The proportions of responses for males and females were compared to determine if differences in responses existed prior to SIEL. The following hypotheses were tested:

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

where p_f represented the proportion of females who mentioned the category pre-course for the depression article and p_m represented the proportion of males who mentioned the category pre-course for the depression article.

A significance level of 0.05 was used. Results for these hypothesis tests are presented in Table 10 Appendix E. P-values were determined using chi-square tests. When expected cell counts were less than five, Fisher's exact tests were performed. P-values from Fisher's exact test are denoted with an "F" after the p-value in Table 10 Appendix E.

Results from Table 10 Appendix E indicated that, despite the large number of tests performed, there were still no significant differences in the proportions of responses for males and females for any of the topic categories that were observed for the depression article prior to taking SIEL. These results may suggest that there was not a gender effect on topic category responses prior to taking SIEL.

Post-course differences in concerns for males and females for the depression article.

The proportions of responses for males and females were compared to determine if differences in responses for topic categories existed for the depression article after taking SIEL. The following hypotheses were tested:

$$H_0 : p_f = p_m$$
$$H_a : p_f \neq p_m$$

where p_f represented the proportion of females who mentioned the category post-course for the depression article and p_m represented the proportion of males who mentioned the category post-course for the depression article.

A significance level of 0.05 was used. All significant results are presented in Table 4.44. A complete list of results are found in Table 11 Appendix E. P-values were determined using Chi-square tests. When expected cell counts were less than five, Fisher's exact tests were performed. P-values from Fisher's exact test are denoted with an "F" after the p-value in these tables.

Table 4.44: Significant† post-course percentages for topic categories by sex for the depression article

Topic Category	Percent female	Percent male	p-value for differences*
Effect of factors on results	29.76	13.33	0.0208
Effect of how information obtained on statistics	51.19	23.33	0.0008
How was information obtained?	86.9	55	<0.0001

†Significant at $p < 0.05$ level
*Exact p-values were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

After taking SIEL, there were three topic categories in which significant differences in the proportion of responses for males and females were observed. These categories included the "effect of factors on results," the "effect of how the information was obtained on the resulting statistics," and "how the information in the article was obtained." For each of these topic categories, females had a higher percentage of responses than males. It is unclear why these results were observed. Given the large number of tests that were performed, we would expect some to be significant. Future research should be conducted to investigate whether sex differences actually exist.

Changes in concerns based on sex.

To determine if there were significant changes in responses from pre-course to post-course for males and females, McNemar's tests were performed on responses to each topic category. The following are the hypotheses that were tested:

$$H_0 : p_{01} = p_{10}$$
$$H_a : p_{01} \neq p_{10}$$

where p_{01} represented the proportion of females/males who did not mention the category pre-course but did mention the category post-course for the depression article and p_{10} represented the proportion of females/males who mentioned the category pre-course and did not mention the category post-course for the depression article.

Statistical significance was determined at a level of 0.05. All significant results from the McNemar's tests are presented in Table 4.45. A complete list of results can be found in Table 12 Appendix E.

Table 4.45: Significant† results from McNemar's tests by sex for topic categories for the depression article

Topic Category	Sex	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Bias	Female	65	2	13	15.48	4	4.76	0.049
Definition of major depressive episode	Female	11	33	40	47.62	0	0.00	<0.0001
	Male	15	23	22	36.67	0	0.00	<0.0001
Effect of definition of depression	Female	41	12	28	33.33	3	3.57	<0.0001
	Male	31	5	20	33.33	4	6.67	0.0015
Effect of how information obtained on statistics	Female	30	13	30	35.71	11	13.10	0.0043
Effect of question	Female	64	2	17	20.24	1	1.19	0.0001

asked on results								
How was information obtained?	Female	5	40	33	39.29	6	7.14	<0.0001
Location of study?	Male	32	8	15	25.00	5	8.33	0.0414
Lurking variable	Female	71	0	12	14.29	1	1.19	0.0034
	Male	53	0	7	11.67	0	0.00	0.0156
Measurement error	Female	57	3	20	23.81	4	4.76	0.0015
	Male	32	6	17	28.33	5	8.33	0.0169
Number of boys and girls?	Male	31	9	17	28.33	3	5.00	0.0026
Question premise of article	Female	65	3	3	3.57	13	15.48	0.0213
	Male	39	1	3	5.00	17	28.33	0.0026
Questions asked?	Female	58	3	21	25.00	2	2.38	<0.0001
	Male	41	3	14	23.33	2	3.33	0.0042
Representative?	Male	47	1	10	16.67	2	3.33	0.0386
Sample size?	Female	71	1	10	11.90	2	2.38	0.0386

†Significant at a $p < 0.05$ level
* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-value for comparing p_{01} and p_{10}

For several topic categories there were significant differences in pre-course and post-course responses for both males and females. These categories include "definition of major depressive episode," "effect of definition of depression," "lurking variable," “measurement error,” “question premise of advertisement,” and “questions asked.” For all of these categories except “question premise of advertisement,” the proportion that did not have a response in a category before SIEL but did have a response after SIEL ($\hat{p}_{01}$) was higher than the proportion that did have a response in a category before SIEL but did not have a response in a category after SIEL ($\hat{p}_{10}$). As mentioned earlier, “question the premise of the ad” is less relevant to the claim so this pattern of responses is favorable.

For several categories, there was a significant change in responses pre-course to post-course for females only. These categories included “bias,” “effect of how information obtained on statistics,” “effect of question asked on results,” “how was information obtained,” and “sample size.” For these categories, the proportion that did not have a response in a category before SIEL but did have a response after SIEL ($\hat{p}_{01}$) was higher than the proportion that did have a response in a category before SIEL but did not have a response in a category after SIEL ($\hat{p}_{10}$).

Significant changes in responses for several topic categories from pre-course to post-course were observed for males only. These categories included “location of study,” “number of boys and girls,” and “representative.” For these categories, the proportion that did not have a response in a category before SIEL but did have a response after SIEL ($\hat{p}_{01}$) was higher than the proportion that did have a response in a category before SIEL but did not have a response in a category after SIEL ($\hat{p}_{10}$).

Summary of results for the depression article

Prior to and after SIEL, the most common topic categories concerned "how the information in the article was obtained" and the "definition of major depressive episode." No differences in responses by sex were observed prior to SIEL, but after SIEL there were differences in responses for males and females. Females had a higher percentage of responses than males for the following topic categories: effect of factors on results, effect of how information was obtained on results, and how information was obtained.

Changes in the pattern of responses were observed for the following topic categories: "bias," "definition of major depressive episode," "effect of definition of depression," "effect of how information obtained on statistics," "effect of lurking variables," "effect of question asked on results," "effect of when the study was done on results," "errors," "girls more likely to admit depression," "how was information obtained," "lurking variable," "measurement error," "MOE or confidence interval included," "non-response error," "number of boys and girls," "question premise of advertisement," "questions asked," and "representative."

For males and females, changes in pattern of responses were observed for the following topic categories: "definition of major depressive episode," "effect of definition of depression," "lurking variable," "measurement error," "question premise of advertisement," and "questions asked." Females also had a significant change in pattern of responses for bias, effect of how information was obtained, effect of questions asked on results, how the information was obtained, and sample size. For males, topic categories in which significant changes in patterns of responses were observed included location of the study, the number of boys and girls in the study, and whether the sample was representative.

Summary of Results for Research Question 2.

Results for Research Question 2 suggested that SIEL had an influence on the pattern of responses for some of the topic categories that were questioned when comparing pre-course to post-course topic categories. The topic category that had one of the largest changes in percentages from pre-course to post-course for the two advertisements and article concerned how the information was obtained. Based on significant changes in the patterns of responses observed from McNemar's tests, other topic categories that were common to the media articles and were influenced by SIEL included effects of definitions or the effect of how information was obtained on statistics in the media article, concerns about the margin of error or confidence interval, and lurking variables (Table 4.46). These topic categories were stressed in SIEL and are important for statistical literacy.

Table 4.46: A comparison of topic categories that had significant results from McNemar's tests for the three media articles

Allstate advertisement	LiveActive advertisement	Depression article
Cause of accident	Effect of definition of occasional irregularity on 4 of 5	Bias
Definition of car	Effect of how information was obtained on 4 of 5	Definition of major depressive episode
Definition or type of accident	How they obtained information	Effect of definition of depression
Effect of definition of car accident on 6 million	Lurking variables	Effect of how information obtained on statistics

Have Allstate	MOE or confidence interval included	Effect of lurking variables
How was information obtained	Non-response	Effect of question asked on results
Location of accidents	Question wording	Effect of when the study was done on results
Lurking variables	Representative	Errors
MOE or confidence interval included	Sensitive topic	Girls more likely to admit depression
Source of 6 million		How was information obtained
Too many definition		Lurking variable
Unreported accidents included		Measurement error
		MOE or confidence interval included
		Non-response error
		Number of boys and girls
		Question premise of advertisement
		Questions asked
		Representative

Differences in topic categories that were questioned based on sex were investigated. Prior to SIEL, there were no significant differences in the percentage of responses to topic categories for males and females for the article on depression. In contrast, differences in responses for males and females were observed for the Allstate and LiveActive advertisements prior to SIEL. After SIEL, significant differences in the percentage of responses to topic categories between males and females were observed for all media articles.

These results suggested the importance of using several media articles with varied content because the focus of the media article may have an influence on the topic categories that were questioned. For example, a possible reason sex differences were observed for the LiveActive advertisement may be that the focus of the advertisement was female related. Females may have felt more comfortable questioning the claim in this advertisement. More research involving different types of media is needed to understand differences in responses for males and females.

Table 4.47 provides a comparison of topic categories in which there was a significant change in response patterns pre-course to post-course for the three media articles based on sex. This table includes topic categories in which significant results from McNemar's tests were observed for females only, males only, and both females and males. Topic categories in common for males and females included questioning about definitions or effects of definitions. More research concerning differences in response patterns for topic categories based on sex and articles of varied context is needed.

Table 4.47: A comparison of significant results from McNemar's tests by sex

	Allstate	LiveActive	Depression
Females only	Cause of accidents Dark figure Effect of definition of car on 6 million Population or population size Question if could count all accidents Source of 6 million Too many definition Unreported accidents included	Effect of how information obtained on 4 of 5 Measurement error Non-response Question wording Representative Sensitive topic	Bias Effect of how information was obtained Effect of questions asked on results How the information was obtained Sample size
Males only	Definition of America Location of accidents		Location of the study The number of boys and girls in the study Representative

Both females and males	Definition or type of accident How was information obtained MOE or confidence interval included	Effect of definition of occasional irregularity on 4 of 5 How did they get their information Lurking variable MOE or confidence interval included	Definition of major depressive episode Effect of definition of depression Lurking variable Measurement error Question premise of advertisement Questions asked

Although there were changes in patterns of questioning of important topic categories stressed in SIEL, there were some important topic categories in which significant changes in patterns of response were not observed. These categories included bias, sample size, generalization, correlation, and types of errors including measurement error, coverage error, and non-response error. The lack of statistical significance for these topic categories may indicate that more emphasis should be placed on these topics in SIEL or that questioning related to these topic categories requires more time (a longer "digestion time") than a single course offers in order to have questions about these categories. A longitudinal study could be conducted to determine if changes in response patterns were observed a year or more after taking SIEL.

In the next section, Research Question 3 is described and results are presented. Discussion of results and summaries are also included.

Research Question 3

Overview

At the end of the semester, students enrolled in SIEL were required to complete a reflection paper (Appendix B) as part of their final course grade. The reflection paper included questions concerning their past experiences and experiences during the semester in SIEL. Selected students' responses were evaluated and grouped into categories based on similar content. The method of selecting students' papers that were evaluated was described in Chapter 3.

The overall purpose of Research Question 3 was to determine what patterns related to the experience of taking SIEL and past experiences were observed from responses to this reflection paper. This overall question was subdivided into two more specific research questions. These included the following:

3a. Were these patterns observed from responses to the reflection paper different for males and females?

3b. Were these patterns observed from responses to the reflection paper different for the three classifications of "effort?"

For Research Question 3b, students' effort levels were determined based on the percentage of completed activities during the semester. These effort levels were categorized into one of three levels: low, moderate, or high as described in Chapter 3.

For Research Question 3, there were three foci. Responses to reflection paper questions were categorized based on similar content. First, overall percentages of responses were determined and differences in proportions of responses to categories for each reflection paper question were investigated. The second focus pertained to determining if associations between sex and categories for each question existed (Research Question 3a) while the third focus dealt with determining if associations existed between effort levels and categories for each question (Research Question 3b).

For each of these hypothesis tests, a significance level of 0.05 was used. Results for Research Question 3 are presented in three sections corresponding to the three foci. Within each section, results for each reflection paper are presented and discussed.

Focus 1: Overall Proportions of Responses

This section focused on determining if there were differences in the proportion of responses to categories from each reflection paper question. The following hypotheses were tested:

$$H_0 : p_1 = p_2 ... = p_n$$
$$H_a : \text{Not all proportions are equal}$$

To test these hypotheses, Chi-square tests were performed. In the section that follows, results for each reflection paper are presented and discussed.

Please describe past experiences in other mathematics or statistics courses that you have taken.

Based on remarks to this question, students' views concerning their past experiences were categorized as "indifferent," "positive," "negative" or "mixed" (a combination of positive and negative views). Examples from each of these categories are provided.

The following is an example of a remark from the "indifferent" view:

> "I have taken two other statistic based courses. The first one was based on learning the basics of statistics; it was a more mathematical class. The second course pulled all the basic concepts of stats together and taught us to analyze data at a higher level. At the end of the course we had to create our own hypothesis and carry out all the testing, data analysis, and conclusions on our own."

A representative example of a student's remark from the "mixed" view is as follows:

> "I have taken many math courses. I have always loved math. In high school I took geometry, 2 algebras, and calculus. I did not take a statistics course. When I got to this University I was required take a Math course. I took a second level math science statistics course and was very confused the entire time. The teacher did not explain the different tests well enough to apply them. I basically memorized formulas and barely passed through the class. Other than that I have not taken any more math courses here at this University."

An example of a student's comment that was categorized in the "negative" view:

> "I have always struggled in math courses; it has never been my strong suit. I have always felt as if I was behind everyone else in math class, regardless of how much time/effort I put into it. This has always been extremely frustrating for me, as I have went [sic] as far as seeking out private tutors and extra help but I still seem to come up short."

The following represents an example of a student's comment from the "positive" view:

> "I have taken many math courses previously. In high school I took many honors courses and also an AP statistics class my senior year. At this University I have taken three math courses, and currently taking math course and SIEL. My past experiences in my other mathematics courses have been pretty good. I really enjoy math, so I do not mind taking math courses. I really like solving equations and formulas. Although most of the math courses that I have taken at this University have been a bit challenging I still enjoy it very much. I really like dealing with numbers, because for me it challenges in a fun way. I would much rather take a math course over an English course any day."

In Table 4.48, percentages of these categories are listed. More than one-third had a positive view and more than one-third had a negative view. The fewest percent of students had mixed views of past experiences.

Table 4.48: Percentages for overall view of past experiences in other mathematics or statistics courses taken

View of Past Experience	Percentage
Indifferent View	18.33%
Mixed View	8.33%
Negative View	35.00%
Positive View	38.33%

Results from a Chi-square test indicated that not all of the proportions of categories of views of past experiences in mathematics or statistics courses were equal (p-value= 0.0024). These results suggested that there were different experiences with previous mathematics or statistics courses but that fewer students expressed either mixed or indifferent views. Views of past experiences with mathematics or statistics classes were polarized. The categories that responses were grouped into allowed for only one category option and this explains why the percentages in Table 4.48 added to 100%.

Explain how your experiences in SIEL were similar to your past experiences in mathematics or statistics courses.

Students' responses to similarities between SIEL and their past experiences in mathematics or statistics courses were grouped into five categories based on response content. These categories included anxiety, content, course format, no similarities, and problem solving. Examples of the five categories of similarities between SIEL and past mathematics or statistics courses are given.

The following represents a student's comment from the "anxiety" category:

> "There are more differences than similarities between SIEL and my past experiences in math. One similarity may be… the confusion if not panic when a new topic is introduced..! And then there's the practical similarity of dealing with dependent or independent probabilities again, but that's really all I can think of."

A student's remark representative of the "content" category follows:

> "SIEL is extremely similar to AP statistics in that many of the formulas are used in both classes. In my AP stat class we dealt a lot with standard deviation, the normal curve, null, and alternative hypothesis."

The following is a student's remark from the "course format" category.

> "This class was similar in the way that we had the lessons each day, which involved going over the material and working out an example problem. I like this routine because it keeps me organized. We also did class work, not everyday, but enough for me to learn what was expected like in my previous math classes."

A student's comment from the "none" category is presented as follows.

> "I did not feel that this statistics course was similar to any of the math classes I had previously taken. I did however enjoy this one very much!"

An example for the "problem solving" category is illustrated in the following student's remark.

> "SIEL has been similar to some of my past math experiences because it requires a lot of logical and strategic thinking, and as well problem solving skills. The word problems in this class have been similar to a lot of the previous math classes I have taken such as algebra, geometry and pre-calculus."

In Table 4.49, percentages for each category are listed. A majority of students (55%) believed that the content of SIEL was most similar with past mathematics or statistics courses. Only 3.33% of respondents believed that the problem-solving skills involved with SIEL and previous mathematics or statistics courses were similar. The results from the Chi-square test indicated that there were differences in the proportions of responses concerning similarities between SIEL and past mathematics or statistics courses (p-value=0.0024).

Table 4.49: Percentages for similarities of past experiences in mathematics or statistics courses with SIEL

Similarities with other courses	Percentage
Anxiety	6.67%
Content	55.00%
Course format	30.00%
None	5.00%
Problem solving	3.33%

Explain how your experiences in SIEL were different from your past experiences in mathematics or statistics courses.

Students' responses to differences in experiences between SIEL and their past mathematics or statistics courses were grouped into four categories. These differences included analysis, application to everyday life, course format, and first statistics course. Examples for each of the four categories are presented.

The following is a representative example from the "analysis" category.

> "My prob/stats class in high school was very mathematics-based; it was mostly problems and worksheets with very little information to analyze. I prefer the approach that SIEL has taken because it makes me think about what I'm doing in-depth, and I believe this has made me more successful at it."

An example from the "application to everyday life" category follows.

> "In this course, the material was based on how statistics is used in everyday life. In my statistics class in high school we never really even talked about applications of statistics in life. I enjoyed SIEL because the math work we did would be applied to life and it actually made sense. I hate how in many math courses, like calculus, I feel that the majority of what I am learning I will never use again. With SIEL I did not feel this way."

The following represents a student's comment from the "course format" category.

> "None of my other math courses taught from power point slides. All of the notes were hand written on the board. All my other math classes were also in smaller lecture settings than this particular course which made asking questions a little easier but the material was also much more difficult that what we did in SIEL. Also, none of my other math courses had activities for us to participate in and they had a lot more calculations to do. I have also never received a formula sheet for an exam before."

The category with the highest percentage of responses was course format (Table 4.50). This was to be expected since the format of SIEL represented an unconventional approach to presenting statistical concepts. Based on Chi-square test results, there were differences in proportions of categories of differences between SIEL and past mathematics or statistics courses (p-value < 0.0001). As with similarities of past experiences and SIEL, students' responses concerning differences could be classified into one difference category so that percentages in Table 4.50 sum to 100%.

Table 4.50: Percentages for differences of past experiences in mathematics or statistics courses with SIEL

Differences with other courses	**Percentage**
Analysis	20.00%
Application to everyday life	33.33%
Course format	43.33%
First statistics course	3.33%

Explain what "statistical literacy" means to you.

Based on students' remarks to what statistical literacy meant to them, responses were categorized as students having a basic understanding of statistical literacy or not. To be categorized as having a basic understanding of statistical literacy, students' responses indicated that statistical literacy involved not only understanding basic statistics but either understanding the application of statistics or challenging statistics. Examples of each classification are given below.

The following represents a student's comment that was categorized as having a basic understanding of statistical literacy.

> "Statistical literacy is how well someone understands statistics. It's their ability to comprehend what the ideas that are being given and for them to copy and understand the same idea. For example, If someone that was "statistical literate" was looking at an article and came across a statistic, they would be able to think about the problems with the statistic and know WHY there are problems with that statistic. This person has to be able to know the terms and why's associated with statistics."

The following represents a student's comment that was categorized as not having a basic understanding of statistical literacy.

> "To me, statistical literacy is talking about a person's ability to understand what statistics is about. To me it also means a person's ability to determine a statistic."

Table 4.51 presents the results from categorizing students' responses. Based on criteria described, over two-thirds of students had a basic understanding of statistical literacy. A chi-square test was performed, and results indicated that the proportion with a basic understanding of statistical literacy was statistically different from the proportion without a basic understanding of statistical literacy (p-value = 0.0055). In addition, a 95% confidence interval for the proportion with a basic understanding of statistical literacy was (0.5656, 0.801). This confidence interval indicated that the majority had a basic understanding of statistical literacy.

Table 4.51: Percentages of students' understanding of the concept of statistical literacy

Basic understanding of statistical literacy	**Percentage**
No	31.67%
Yes	68.33%

Do you feel that you are "statistically literate?"

Students were asked to respond to whether they believed they were statistically literate with a response of "yes" or "no." Table 4.52 lists the results. Chi-square test results indicated that the proportion who thought they were statistically literate was higher than the proportion who did not think they were statistically literate (p-value < 0.0001). A 95% confidence interval for the proportion who believed they were statistically literate was (0.8949, 1.00). This confidence interval indicated that a majority believed they were statistically literate.

Table 4.52: Percentages of students' who believe they are statistically literate

Statistical literate?	**Percentage**
No	5.00%
Yes	95.00%

Has SIEL changed the way in which you look at the statistics you encounter in everyday life?

Students were asked to respond to whether SIEL had changed the way they looked at statistics encountered in everyday life with response choices of either "yes" or "no." The results are presented in Table 4.53. The vast majority (86.67%) of students responded that SIEL had changed the way they looked at statistics encountered in everyday life.

Based on the results from a Chi-square test, the proportion who believed SIEL had changed the way they looked at statistics encountered in everyday life was statistically different from the proportion who said that SIEL had not changed the way they looked at statistics in everyday life (p-value < 0.0001). In addition, the 95% confidence interval for the proportion who believed that SIEL had changed the way they looked at statistics encountered in everyday life was (0.7807, 9527). This confidence interval indicated that a majority believed that SIEL changed the way they looked at statistics.

Table 4.53: Percentages of students' who say that the course has changed the way they look at statistics encountered in everyday life

Changed way look at everyday life statistics?	Percentage
No	13.33%
Yes	86.67%

Attributes students believe are necessary to be successful in SIEL.

Students were asked to list attributes they believed were necessary to be successful in SIEL. A complete list of attributes and percentages of students who said the attribute was necessary for success in SIEL can be found in Table 1 Appendix F. The most common attributes students believed were necessary to be successful in SIEL included the following: attend class (35%), do homework or practice (28.33%), have an open mind (25%), be able to perform basic mathematics or have basic mathematics skills (23.33%), pay attention in class (20%), and have problem solving or logical thinking skills (20%). The given percentages do not total to 100% because many students listed multiple attributes.

Attributes listed in Table 1 Appendix F were grouped into four main categories based on similar attribute characteristics. These categories included attributes related to abilities, class behavior, student qualities, and thinking capabilities. Results are presented in the following section.

Attributes to be successful in SIEL: Abilities.

The following attributes were grouped into the “abilities” category: accept past mistakes, apply definitions, do basic math or math skills, learn definitions, learn or use formulas, memorize, understand and apply concepts, understand basics, understand big picture, understand graphs, understand marketing and advertising, understand statistics before course, and understand theories. Table 4.54 presents the percentages of students who indicated that an attribute in the “abilities” category was necessary to be successful in SIEL. Based on the chi-square test, there was not evidence to say that there was a difference in the proportion who said and did not say that having an attribute in the abilities category was necessary to be successful in SIEL (p-value = 0.4386).

Table 4.54: Percentages of students who said "ability" attribute was needed to be successful in SIEL

Abilities?	Percentage
No	45.00%
Yes	55.00%

Attributes to be successful in SIEL: Class behavior.

The following attributes were classified in the "class behavior" attribute category: ask questions, attend class, do homework or practice, participate in activities or class, pay attention in class, print notes for class, study, take good notes, and work well in groups. Table 4.55 lists the percentages of students who said that an attribute in the "class behavior" category was necessary to be successful in SIEL. Results of a Chi-square test indicated that there was not enough evidence that there was a difference in the proportions of students who said and did not say that the "class behavior" attribute was necessary to be successful in SIEL (p-value = 0.1967).

Table 4.55: Percentages of students who said "class behavior" attribute was needed to be successful in SIEL

Class behavior?	Percentage
No	41.67%
Yes	58.33%

Attributes to be successful: Student quality.

The "student quality" category included the following attributes listed in Table 1 Appendix F: apply yourself, appreciation for statistics, be excited about course, be prepared or willing to learn new things, conscientiousness, creative, dedication, desire, determination or driven, do not underestimate course, do not need to be good or interested in statistics, effort, good listeners, intelligence, interest, like math, motivation, observant, open minded, organized, patience, perceptive, responsibility, self-control, serious attitude, visual learner, want to learn, and work hard or work ethic. Table 4.55 lists the percentages of students who said that the "student quality" attribute was necessary for success in SIEL. A Chi-square test to compare these percentages indicated that the proportion who said that student quality attribute was necessary for success in SIEL was significantly different from the proportion who did not say this attribute category was necessary (p-value = 0.0389). A 95% confidence interval for the proportion who believed "student quality" was needed to be successful in SIEL was (0.5114, 0.7552), and this interval indicated that a majority believed that "student quality" was needed to be successful in SIEL.

Table 4.56: Percentages of students who said "student quality" attribute was needed to be successful in SIEL

Student Quality?	Percentage
No	36.67%
Yes	63.33%

Attributes to be successful: Thinking capabilities.

Attributes from Table 1 Appendix F in the "thinking capabilities" category included analytical, common sense, comprehend word problems, different way of thinking, do not think analytically, good memory, problem solving or logical thinking skills, question statistics and think on your feet. The percentages of students who believed that the "thinking capabilities" category was necessary for success in SIEL can be found in Table 4.57. A Chi-square test revealed that there was not enough evidence to indicate there was a difference in the proportions who believed and did not believe that the "thinking capabilities" category was necessary for success in SIEL (p-value = 0.1967).

Table 4.57: Percentages of students who said thinking capabilities attribute was needed to be successful in SIEL

Thinking Capabilities?	Percentage
No	58.33%
Yes	41.67%

Even after taking a concept based statistics course, it is difficult to believe that a majority of selected students would believe that thinking capabilities would not be necessary in order to be successful in SIEL. An explanation for this result may be that students do not associate understanding and applying statistical concepts with "thinking capabilities." They might only relate "thinking capabilities" with performing computational techniques.

During the course were you presented with material that challenged your beliefs?

Students were asked if material in SIEL had challenged their beliefs. Examples of students' explanations as to whether SIEL challenged their beliefs are presented.
The following represents a student's comment when the student felt that SIEL had not challenged his/her belief.

> "To me 'challenged your beliefs' means that something offended you and nothing in this course offended me at all or made me feel uncomfortable."

The following student's comment represents an example of when a student felt that SIEL had challenged his/her belief.

> "My immediate answer was no, but then after giving this greater thought I must admit that my belief in trusting others may have changed. I have always trusted the 'experts' when reading or hearing research information presented, but now I have the knowledge to look at information deeper and apply the principles of SIEL to evaluate their information for validity."

The results of responses to this question can be found in Table 4.58. Almost all students (93.33%) said that SIEL had not challenged their beliefs. Results of a Chi-square test indicated that the proportion that did not feel that SIEL challenged their beliefs was higher than the proportion who did feel that SIEL had challenged their beliefs (p-value < 0.0001). A 95% confidence interval for the proportion who did not feel that SIEL challenged their beliefs was (0.8702, 0.9964). This interval indicated that a majority did not feel that SIEL challenged their beliefs.

Results may have been attributed to the wording of the question; the use of the phrase "challenging your beliefs" may have been too strong. Students may have interpreted this phrase as meaning a life altering change or change in religious beliefs or personal philosophy.

Table 4.58: Percentages of students who say that during the course were you challenged with material that challenged your beliefs

Beliefs challenged?	**Percentage**
No	93.33%
Yes	6.67%

How would you describe your attitude toward statistics – good, moderate, or poor?

Students were asked to describe their attitude toward statistics as good, moderate, or poor. The results are listed in Table 4.59. None of the respondents indicated that their attitude toward Statistics was poor. A majority of students classified their attitude toward statistics as “moderate.” A Chi-square test indicated that there were differences in the proportions of attitude categories (p-value < 0.0001).

Table 4.59: Percentages of students’ who describe their attitude toward statistics as good, moderate, or poor

Attitude	Percentage
Good	48.33%
Moderate	51.67%
Poor	0%

As a result of this class, has your attitude toward statistics changed?

Students were asked if they believed that their attitude toward statistics had changed as a result of SIEL. Percentages of students’ responses can be found in Table 4.60. A majority of students believed their attitude toward statistics had changed because of SIEL. A Chi-square test revealed that the proportion who believed their attitude toward statistics had changed because of SIEL was significantly different from the proportion who did not feel this way (p-value = 0.0045). The 95% confidence interval for the proportion who believed that SIEL changed their attitude toward statistics was (0.5656, 0.801) which indicated that a majority held such a belief.

Table 4.60: Percentages of students' who say that their attitude toward statistics has changed as a result of this course

Attitude changed?	Percentage
No	31.67%
Yes	68.33%

Students were asked to expand on their answer concerning whether they believed that SIEL had changed their attitude toward statistics. Classifications of students' attitudes were based on these descriptions. Analyses were performed separately based on whether attitudes had or had not changed as a result of SIEL.

For those students who said that their attitude had not changed as a result of this course, how did they describe their attitude?

For those students who said that their attitude had not changed as a result of SIEL, the description of their attitudes was classified as negative, positive, or neutral. Examples of each classification are presented.

An example from the "negative attitude" classification follows.

> "Although this class was not overly challenging, I understand that statistics can get much, much more complicated. I still have a little bit of statistical anxiety when I think about the classes I will have to take in the future. I know that I can understand the material if I try, but statistics are still a little bit scary because I do not want to underachieve do [sic] to a lack of understanding them. In other words, before this class, my attitude towards statistics could be described as an uneasy trepidation. Now, I am more confident about my abilities to work through any challenges that statistics may present in the future, but I am still uneasy about what those challenges will be."

The following is a representative example for the "neutral attitude" classification.

"My attitude has not changed toward statistics because I never had an attitude towards it before and I still don't really have an attitude towards it. This class isn't part of my major so it isn't a part of my main focus. I just took it because I had to fill a general education requirement. I hate to say it, but I kind of just go to class then put it on the back burner until the next class or test. I can't really focus on it that much because I have too many other classes."

An example from the "positive attitude" classification follows.

"My attitude towards statistics hasn't changed because I always felt that statistics was a pretty easy course to study and I've done well in it both in the past and now. Maybe if I continue to take statistic courses and they begin to get more difficult to understand my attitude will change, but as of right now I still have a positive attitude."

The results can be found in Table 4.61. The majority of students (52.63%) who said that their attitude had not changed as a result of SIEL indicated that they had positive attitudes toward statistics. The Chi-square test indicated that there was not enough evidence that there were differences in the proportions of attitude classifications of negative, positive, or neutral (p-value = 0.0759).

Table 4.61: Description of attitude as a result of course for those who said that their attitude was not changed

Attitude	**Percentage**
Negative	10.53%
Neutral	36.84%
Positive	52.63%

If attitude had changed as a result of this course, in what way did it change?

For those students who said that their attitude had changed as a result of SIEL, descriptions of attitude changes were categorized as negative to positive or neutral to positive. None of the selected students' responses indicated that their attitude changed from either positive or neutral to negative. The following statements from students reflect each of these categories.

An example from the "negative to positive" change in attitude category follows.

> "As I mentioned above, I had a very negative attitude toward statistics, math in general for that matter, going into this course. I did not understand chi-squared tests, how to pick out the parameters, run a successful hypothesis test, NOTHING! But after taking this course, I really started to enjoy how statistics were involved in my everyday life, and how I am now able to completely understand the data presented. I know a lot more on the subject than I ever thought I would be able to comprehend, and am very pleased to say that my high grades reflect my attitude towards this class"

The following represents a student's comment from the "neutral to positive" change in attitude classification.

> "Yes, because after taking this course I can say I know more about statistics in everyday life and I am able to apply to everyday stats that I see in magazines I read, or TV shows I watch! This class from day one has been very beneficial to me and will be beneficial to me for the rest of my life. My attitude towards statistics has changed for the better because of this class because coming into this class I thought it was just going to be just another math class I took and got nothing from, but it was the complete opposite. Thanks for making this class fun and interactive so I could enjoy and learn and honestly say that I can apply what I learned to everyday stats I see in the future!"

Results for this question can be found in Table 4.62. The higher percentage of attitude changes occurred from neutral to positive (85.37%). Chi-square test results indicated that the proportion of attitude changes from negative to positive was different from the proportion of attitude changes from neutral to positive (p-value < 0.0001).

Table 4.62: Percentages of direction of students' attitude change toward statistics as a result of this course

Attitude change	Percentage
Neutral to Negative	0%
Positive to Negative	0%
Negative to Positive	14.63%
Neutral to Positive	85.37%

List three adjectives you would use to describe statistics.

Students were asked to provide three adjectives they would use to describe statistics. Overall percentages for these adjectives are located in Table 2 Appendix F. The most common adjectives used to describe statistics included challenging (10%), complex (10%), informative (10%), important (11.67%), interesting (23.33%), and useful (25%). Although most of these adjectives have positive connotations in describing statistics, "challenging" and "complex" may also be considered as having negative connotations.

Summary for Results for Focus 1: Overall Proportions

Results from analyses for investigating overall proportions of responses indicated that there were differences in past experiences with mathematics or statistic. Past experiences were polarized with highest percentages of views of past experiences categorized as either positive or negative. The content of SIEL was observed to be most similar to past courses while course format was seen as most different with past courses. An understanding of statistical literacy was observed from students' comments, and a majority felt they were statistically literate. Results suggested that SIEL changed the way statistics from everyday life were viewed.

The highest percentage of students (35%) believed that attending class was a necessary attribute for being successful in SIEL. When attributes necessary for being successful in SIEL were grouped into similar categories, the only attribute category with a significant difference between those who believed and did not believe the category was necessary for success in SIEL was the "student quality" attribute. Attributes in this category were those related to students' work ethics and abilities.

Although results suggested that SIEL had not challenged their beliefs, evidence indicated that students believed that attitudes toward statistics had changed as a result of SIEL. A higher proportion of changes in attitudes from neutral to positive as compared to negative to positive. The most popular adjectives used to describe statistics had positive connotations.

Focus 2: Effect of Sex on Responses

The focus of this section was to determine if there were associations between students' sex and the response categories for each reflection paper question. The following hypotheses were tested:

H_0: There is no association between sex and response categories

H_a: There is an association between sex and response categories

Chi-square tests were performed for each research question. Results for each reflection paper question can be found in Tables 3 through 17 in Appendix F. Statistical significance were determined at a level of 0.05. P-values from chi-square tests were recorded unless the expected cell count was less than five in which case p-values from Fisher's exact test were given.

Of the Chi-square tests that were performed on responses by sex, only two reflection paper questions had statistically significant results. Table 4.63 provides the percentages of categories of differences in SIEL and their previous mathematics or statistics courses for males and females. There was a significant association between sex and differences between SIEL and past experiences with other mathematics or statistics courses (p-value=0.0145). A majority of females believed course format was the primary difference between past experiences and SIEL. For males, 36% believed differences were due to analysis and application to everyday life. More research on differences in past experiences for males and females is needed.

Table 4.63: Percentages for differences of past experiences in mathematics or statistics courses with SIEL based on student sex

	Differences with other courses			
Sex	Analysis	Application to everyday life	Course format	First Statistics course
Female	8.57%	31.43%	57.14%	2.86%
Male	36%	36%	24%	4%

The percentages of male and female students who believed that the class behavior attribute was necessary to be successful in SIEL can be found in Table 4.64. The chi-square test indicated that there was an association between sex and belief that the class behavior attribute was necessary to be successful in SIEL (p-value=0.0149). A majority of females believed the class behavior attribute was necessary for success in SIEL while a majority of males felt just the opposite. It is not clear why this disparity exists.

Table 4.64: Percentages of students who said class behavior attribute was needed to be successful in SIEL by sex

	Class behavior?	
Sex	No	Yes
Female	28.57%	71.43%
Male	60%	40%

Tables 18 and 19 in Appendix F provide lists of adjectives used to describe statistics from females and males, respectively. Table 4.65 lists the most popular adjectives to describe statistics given by females and males and percentages for each adjective. Two common adjectives for both males and females included “interesting” and “useful.” A higher percentage of females described statistics with these adjectives than males. The most popular adjectives given by females have positive connotations with the exception of “challenging” which may be interpreted as having positive or negative connotations. Two of the most common adjectives given by males, “complex” and “misleading,” could be interpreted as having negative connotations while the rest of the popular adjectives given by males have positive connotations.

Table 4.65: A comparison of popular adjectives used to describe statistics by sex

Popular adjectives given by females (%)	Popular adjectives given by males (%)

Useful (31.43%)	Complex (16%)
Interesting (28.57%)	Important (16%)
Challenging (11.43%)	Interesting (16%)
Informative (11.43%)	Useful (16%)
	Misleading (12%)

Summary for Results for Focus 2: Effect of Sex on Responses

There was no evidence of an association between sex and responses for all reflection paper questions except for an association between sex and views of past experiences with mathematics or statistics courses and belief that class behavior was necessary for success in SIEL. The most common adjectives to describe statistics for males and females were, in general, favorable.

Focus 3: Effect of Effort on Responses

The focus of this section was to determine if there was an association between effort levels and the response category for each reflection paper question. The following hypotheses were tested:

H_0: There is no association between effort level and response categories
H_a: There is an association between effort level and response categories

Effort levels were determined based on the percentage of activities completed during the semester as described in Chapter 3.

Chi-square tests were performed for each research question. Results for reflection paper questions can be found in Tables 20 through 34 in Appendix F. Statistical significance was determined at a level of 0.05. P-values from Chi-square tests were recorded unless the expected cell count was less than five. For these situations, p-values from Fisher's exact test were given.

No significant effects were found between effort levels and any of the categories for the reflection paper questions. Although this result implies that there is no association between levels of effort and reflection paper question categories such as those pertaining to their past experiences and attitudes toward statistics, more research including experiments is needed to investigate these relationships.

Adjectives used to describe statistics for students classified at high, moderate, and low effort levels can be found in Tables 35, 36, and 37 in Appendix F, respectively. A summary of the most popular adjectives (more than 10%) used to describe statistics for each effort level is displayed in Table 4.66. It is interesting to note that the number of adjectives with response rates above 10% increased as the effort level decreased from high to low. For the high effort group, the most popular adjectives all had positive connotations while there were some adjectives in the moderate and low effort groups which had either mixed or negative connotations such as "difficult," "misleading," and "confusing." Students classified at the high effort level routinely attended class to be present to perform the class activities. Those who attended class might have more of an interest in statistics and associate positive adjectives with it.

Table 4.66: A summary of popular adjectives used to describe statistics and percentages for the three effort levels

Popular adjectives for High Effort (%)	Popular adjectives for Moderate Effort (%)	Popular adjectives for Low Effort (%)

Interesting (22.86%)	Useful (36.84%)	Challenging (50%)
Important (17.14%)	Complex (15.79%)	Interesting (50%)
Useful (17.14%)	Interesting (15.79%)	Useful (33.33%)
Informative (11.43%)	Complicated (10.53%)	Analytical (16.67%)
	Difficult (10.53%)	Complex (16.67%)
	Insightful (10.53%)	Conceptual (16.67%)
	Misleading (10.53%)	Confusing (16.67%)
	Subjective (10.53%)	Entertaining (16.67%)
	Ubiquitous (10.53%)	Exciting (16.67%)
		Helpful (16.67%)
		Informative (16.67%)
		Insightful (16.67%)
		Meticulous (16.67%)

Summary for Results for Focus 3: Effect of Effort on Responses

Based on statistical analyses of results, effort levels were not associated with any of the response categories for any of the research questions. The number of popular adjectives used to describe statistics, indicated by a response rate of more than 10%, increased as the effort level decreased from high to moderate to low levels. For those with a high effort level, the popular adjectives all had positive connotations while the adjectives for moderate and low effort levels had both positive and negative connotations.

Summary for Research Question 3

Several trends emerged from the three foci for Research Question 3. Results indicated that views of students' past experiences with mathematics or statistics courses were polarized with most students having either positive or negative views. There were also differences between SIEL and past experiences with mathematics or statistics courses for males and females. The highest percent of females said that course format was different while analysis and application to everyday life were the most common differences for males.

Students were asked to indicate the attributes necessary to be successful in SIEL. Overall, the "student quality" attribute was believed to be necessary to be successful in SIEL. Differences in views between males and females as to whether the "class behavior" attribute was necessary to be successful in SIEL did emerge. A higher percentage of females (71.43%) as compared to males (40%) believed that the "class behavior" attribute was necessary for success in SIEL.

Overall, students believed that, after the course, they were statistically literate and that they understood what statistical literacy is. Although students felt that SIEL did not challenge their beliefs, responses indicated that SIEL had changed the way they look at statistics and did change their attitudes toward statistics.

By asking students to list adjectives to describe statistics, insight into how students felt about statistics might be gained. Overall, the most popular adjectives (more than 10% response rate) to describe statistics included useful, interesting, important, informative, complex, and challenging. For females, the most popular adjectives to describe statistics had positive connotations while for males adjectives with both positive and negative connotations were observed. For those classified in the high effort level group, all popular adjectives to describe statistics had positive connotations. The number of adjectives with negative connotations was more for moderate and low effort levels than for high effort level. This result suggests that a high effort level may be associated with a more positive view of statistics.

It may be argued that since students' responses to the reflection paper were not anonymous their responses may not have reflected their true feelings. Examples of responses in this section provide illustrations of how forthright responses were.

Summary of Research Questions

Based on responses to the reflection paper, students indicated that they believed SIEL had changed the way they looked at statistics. Likewise, paired t-tests indicated that there were significant changes in SLCR component scores except for Causality. The statistical literacy components with the highest average change in score were Definition and Method. Although these changes were statistically significant, practical significance is in question since the changes, overall, were small in magnitude. In addition, inter-rater reliability results indicated that caution should taken when interpreting results for statistical literacy components Generalize and Reported statistic.

Higher percentages of students who challenged important topic categories were observed after taking SIEL as compared to before taking SIEL. In addition, females challenged more topic categories than males.

Results from McNemar's tests indicated that there were significant changes in patterns of responses for several topic categories for the three media articles. It is also noteworthy to mention that there were changes to patterns of responses to topic categories that were less relevant to the statistical claim made in the media article. Differences in patterns of responses were observed for males and females for several topic categories. Differences were observed for more topic categories for females as compared to males.

From the reflection paper, students also indicated that they believed they were statistically literate. Responses indicated that a majority had a basic understanding of what statistically literacy is. Although students understood the concept of statistical literacy, average SLCR scores indicated, at most, a basic awareness of statistical literacy components either prior to or after taking SIEL.

Polarized views of past experiences were observed from reflection paper responses. Differences in students' past experiences with other mathematics or statistics courses and SIEL were observed for males and females. For females, course format was observed to be the most common difference between SIEL and past courses while analysis and application to everyday life were the most common difference for males. Prior to SIEL, differences in topic categories that were questioned were observed for males and females and these may be due to differences in past experiences.

In addition, context may have been another factor in determining which topic categories were questioned prior to and after SIEL for males and females. For example, no statistically significant differences in any topic categories were observed for males and females prior to SIEL for the depression article yet there were differences observed for males and females for Allstate and LiveActive.

Although not a direct research question, significant changes in SATS-36© attitude component scores were observed from pre-course to post-course although these changes were small and are of little practical significance. No significant changes were observed for attitude components Interest and Value. For the attitude component Effort a significant negative change was observed which indicated that students believed they needed to expend more effort to learn statistics after SIEL than they believed they needed prior to SIEL. Similarly, responses to the reflection paper indicated that students believed that their attitudes had changed after taking SIEL. For those who indicated that their attitudes had changed from either negative to positive or from neutral to positive. Some students indicated that changes in attitudes were related to the course format that focused on how statistics is applicable to everyday life.

Results of correlation analyses indicated that pre-course and post-course Effort component scores were correlated only with each other and no other SATS-36© attitude components. Course effort was correlated with SATS-36© attitude component Effort, which suggested that students were realistic in the assessment of the effort they would expend in the course. In comparison, the responses from the reflection paper indicated that effort levels were not significantly correlated with any categories of the reflection paper. The most common adjectives used to describe statistics differed in connotation based on effort levels. The most common adjectives for those classified as high effort all had positive connotations whereas adjectives for moderate and low effort levels had positive and negative connotations.

In Chapter 5, an overview of the research questions and results are provided. In addition, assets and limitations of the study are discussed as well as future research to be conducted. Final thoughts concerning this study conclude this section.

CHAPTER FIVE

CONCLUSIONS

This chapter provides an overall summary of results from the three research questions. In addition, assets and limitations of the study and future research endeavors are discussed. The chapter concludes with final thoughts about the study.

Overall Summary of Results

The purpose of this research was to determine the effects of taking a statistical literacy course on the challenges that college students make to statistical claims in media articles. Two advertisements and an article were given to students prior to and after taking Statistics in Everyday Life 200 (SIEL). Students were asked to provide questions concerning the information in the media articles and to explain why these questions were important to ask. In addition, at the end of the semester, students were required to complete a reflection paper concerning their past experiences with mathematics or statistics courses and their experiences in SIEL.

Three main research questions were asked. The first research question pertained to determining differences in the levels of awareness of eight components of statistical literacy prior to and after taking SIEL as well as changes in the levels of awareness of these components. The eight components of statistical literacy included Bias, Causality, Definition, Generalize, Lurking variable, Method, Reported statistic, and Variation derived from Utts (2002). The levels of awareness for each component were evaluated using the Statistical Literacy Components Rubric (SLCR). In addition, the effects of attitude toward statistics assessed using SATS-36© and demographic variables (gender, aptitude, and background attributes) on the level of awareness of the statistical literacy components were investigated.

Prior to SIEL, average and median statistical literacy component scores indicated essentially no awareness of the components. After taking SIEL, post-course average and median scores for Definition, Lurking variable and Method indicated minimal awareness while there was still essentially no awareness of the other five components of statistical literacy.

Definition and Method had the highest average post-course scores while Causality and Variation had the lowest. Although significant changes in all statistical literacy component scores except Causality were observed, these changes were small and not of practical significance.

Correlation analyses were performed between average pre-course and post-course SLCR component scores and attitudes toward statistics and demographic variables. For correlation analyses between pre-course average SLCR scores and demographic variables, only the average pre-course Variation score and the number of college statistics courses taken were significantly correlated. In contrast, no significant correlations were observed between post-course SLCR component scores and demographic variables. In addition, no significant correlations were found between either pre-course or post-course SLCR scores and pre-course and post-course SATS-36© attitude components. Although more research is needed, there were no significant linear relationships between statistical literacy and attitudes toward statistics. For correlation analyses, the Bonferroni correction was used and provided a conservative level of significance which affected what variables were significant. If fewer comparisons had been made, more statistically significant correlations would have been detected.

Although independent variables were found to be significant predictors of the level of awareness of statistical literacy components and on the vector of changes in the level of awareness of statistical literacy component scores, average SLCR scores were at or below minimal awareness. In addition, low adjusted R^2 values were observed for each regression model indicating that other important predictors of levels of statistical literacy components were not measured. Therefore the results from these analyses should be interpreted with caution.

Furthermore, changes in all SATS-36© attitude components were observed with the exception of Interest and Value. Of the attitude components in which there was a significant change, Effort was the only component in which the score change from pre-course to post-course decreased. This indicated that students felt that they needed to put more effort into SIEL after taking it than before taking it. Although significant changes in attitude components were observed, these changes were small in magnitude and did not represent meaningful changes. Effort was a unique attitude component because it was the only attitude component that was correlated only with itself and none of the other attitude components. This result suggests that Effort is not related to the other attitude components.

The second research question focused on the questions students asked concerning the three media articles prior to and after taking SIEL as well as changes in the patterns of responses for these questions. The questions were categorized into topic categories. Differences in topic categories for males and females were also investigated.

Overall, significant changes in patterns of important topic categories that were questioned were observed in which these topic categories were not asked before SIEL but asked after SIEL. The topic category that had one of the largest changes in percentages from pre-course to post-course for the two advertisements and article concerned how the information was obtained. Also, differences in topic categories that were questioned were observed for males and females. In every topic category where these differences were observed except two, females questioned the topic category more than males. These results were seen regardless of the media article. The topic categories in which males questioned more than females included “causality” in the LiveActive advertisement prior to SIEL and “definition of America” in the Allstate advertisement after taking SIEL. Topic categories in common for males and females included questioning about definitions or effects of definitions.

The context of the media articles may have contributed to the topic categories questioned by males and females. For the Allstate advertisement, there were eight topic categories in which the patterns of responses were different for females only while there were two topic categories for males only. For the LiveActive advertisement, there were six topic categories in which the patterns of responses were different for females only while there were no categories for males only. More similar numbers of topic categories in which the patterns of responses for males and females differed were observed for the depression article – four for females and three for males. It is interesting to note that the discrepancy for these numbers of these topic categories by gender was greater for the advertisements than for the article. This suggests that with less information given, as with the advertisements, females were more apt to make challenges about nuances of the statistical claim. In addition to media article format, context may have contributed to these results. The greatest discrepancy between the number of topic categories in which the pattern of responses were different for males and females was for the LiveActive advertisement. This advertisement pertained to a “woman’s issue” and this may have affected males’ willingness to question issues which they felt did not affect or pertain to them.

The third research question pertained to responses on a reflection paper that students completed at the end of SIEL concerning their past experiences with mathematics or statistics courses and their experiences in SIEL. The reflection paper was composed of direct, specific questions, including some yes/no questions. Responses were categorized into topic categories. Differences in topic categories were investigated based on sex and effort level. Students were classified into one of three effort levels based on their percentage of completed activities during the semester.

Based on students’ responses to the reflection paper, students showed a wide variation in prior experiences. Most of the views were either positive (38.33%) or negative (35%) with fewer being indifferent (18.33%) or mixed (8.33%). The most common similarity between SIEL and past experiences was content (55%) and the most common differences were analysis (20%), application to everyday life (33.33%), and course format (43.33%).

In the reflection paper, students were asked to describe what statistical literacy is. Based on these responses, 68.33% of students demonstrated that they had a basic understanding of what statistical literacy is, and 95% of students believed they were statistically literate. A majority (86.67%) of students also believed that SIEL had changed the way that they looked at statistics encountered in everyday life. A common reason for this change was due to the course's emphasis on showing the applicability of statistics to everyday life. These reflection paper results are in direct contrast to results from SLCR scores which indicated that even after taking SIEL students had at most minimal awareness of statistical literacy components. It is fascinating that such a large percentage of students believed they were statistically literate and seemed to have an understanding of what statistical literacy is and yet demonstrated such a lack of awareness of statistical literacy components. Although more research into why this inconsistency existed is needed, students' understanding of statistical literacy may have represented the second tier of the Hierarchical Model (Watson, 1997) since it measured only a "basic understanding" of statistical literacy while SLCR was evaluating statistical literacy at the third tier, the highest level.

A majority (68.33%) of students said that SIEL had changed their attitudes toward statistics. Of those students who said their attitude had changed, 14.63% of students' attitudes changed from negative to positive and 85.37% changed from neutral to positive. The results from this research support the results from the reflection paper concerning attitude changes. There were significant changes after taking SIEL in all but two SATS-36© attitude components. Perhaps now with positive attitudes, which affect performance and possibly statistical literacy, students will be in a position to become more aware of statistical literacy components and approach statistical information with an open mind.

Differences in responses for males and females concerning differences of past experiences in mathematics or statistics courses and SIEL were observed. Most females (57.14%) said that the course format of SIEL was different from past course experiences while 36% of males said that both analysis and applications to everyday life associated with SIEL were different from past experiences.

Males and females differed in responses concerning whether class behavior was an important attribute to be successful in SIEL. While 71.43% of females believed that class behavior was an important attribute to be successful in SIEL, 60% of males believed that it was not an important attribute for success in SIEL. Since the class behavior attribute included “being able to ask questions in class,” this result provides insight into gender differences observed for Research Questions 1 and 2. If males did not believe that asking questions was an important attribute for success in the course, then this may have affected their willingness to ask questions of the three media articles. Males did not see the relevance of making such challenges.

For effort levels, no significant differences were observed for any of the reflection paper questions. Like the SATS-36© attitude component Effort, it seems that the amount of work expended was a unique characteristic in that it did not have an effect on students’ experiences in SIEL or related to past experiences in mathematics or statistics courses.

In the reflection paper, students were asked to provide three adjectives to describe statistics. This was asked as a way to understand students’ underlying perceptions of statistics. Interesting trends were observed based on sex and effort levels. The most common adjectives used by females to describe statistics had positive connotations while for males adjectives of positive and negative connotations were given. The choice of adjectives used to describe statistics provides insight into students' view of statistics and may have contributed to gender differences observed in Research Questions 1 and 2. Negative views of statistics may have contributed to an unwillingness to challenge the articles' statistical claims.

For students classified in the high effort level, the most common adjectives given had positive connotations. Those in the moderate and low effort levels gave adjectives with both positive and negative connotations. Although no statistically significant results were found for responses to reflection paper questions based on effort level, more probing questions in the reflection paper may have uncovered differences based on effort level.

Although it cannot be ruled out that one course focused on statistical literacy is not enough to see dramatic changes in awareness of statistical literacy components, the results obtained for Research Question 1 may be a product of how the study was conducted. In retrospect, it is hypothesized that having students make challenges and explain why these were important to ask for three media articles created fatigue and did not show the depth of awareness of the statistical literacy components. When faced with three media articles to provide challenges and explanations, students may have felt overwhelmed by this exercise and just wanted to give a response for each article instead of deeply pondering the claims in each article and giving thoughtful challenges with explanations. In spite of randomizing the media articles as a means to avoid bias, questioning claims in three media articles may have been too mentally taxing. If randomly assigning only one of the articles to each student had been done, students would have had more time to concentrate on the issues of that particular media article the results for Research Question 1 may have shown higher levels of awareness and more questions may have been asked enhancing results for Research Question 2.

The focus of this research was to assess the highest level of the Hierarchical Model (Watson, 1997). This level pertained to questioning statistical claims and evaluated using SLCR. According to this model, to attain this highest level one must have already attained the first two levels. Although the average SLCR scores were low, these results indicated that students had already attained the first two levels of the three level hierarchies. After taking SIEL, students may have reached the second tier of the hierarchy and what was measured in this research was movement into the third tier.

Gender differences were observed in all three research questions. Although before taking SIEL gender was not a significant predictor of any of the statistical literacy components, gender was a significant predictor for components Definition, Method, and Reported statistic after taking SIEL. This result implies that males and females have different levels of awareness to certain components of statistical literacy.

Gender differences were observed with respect to the percentage of males and females who challenged different aspects of the media claims as well as pattern of responses. There were more challenges and significant changes in patterns of responses for females as compared to males.

In addition, when asked to give adjectives to describe statistics, the most common adjectives given by males included those with negative connotations while adjectives given by females had positive connotations. This is more evidence that the course had more of an effect on females than it did for males. The larger question is why did this happen. Why would males be less likely than females to challenge statistical claims made in the media after taking SIEL? One possibility may be that females have healthier attribution, believing they have more control of themselves and therefore are more willing to make challenges.

Future Research

This research represents the building block for other studies on statistical literacy. More research using SLCR is needed using additional media contexts to determine the effect of context on statistical literacy.

Instead of having students ask questions, in general, about the statistic in the article, an alternative approach may be to have students ask questions about the eight statistical literacy components. Although this approach may direct students to think about specific statistical literacy components, it may provide information about students' thought processes specifically related to the particular components prior to and after SIEL.

Researchers may consider comparing the level of awareness of statistical literacy components for a course focused on statistical literacy and a traditional statistics methods course. This research could focus on differences in the topic categories that are questioned for the two types of course formats.

Longitudinal studies could be conducted to determine if there are additional changes in the levels of awareness of statistical literacy components as well as questions that are asked over time. For a longitudinal study, either a particular statistical literacy component or all eight components could be studied over time. This is important because some statistical literacy components may take time for students to grasp. For example, "Understanding of complex concepts like random and variation takes time to develop and may be associated by their application in various contexts" (Watson, 2006, p. 267).

Since some students may have had difficulty in expressing in writing why questions were important to ask, interviews could be conducted as another way to assess levels of awareness of statistical literacy components instead of having students write the questions they would ask. In addition, interviews could be conducted at different grade levels from elementary to high school to assess what questions students ask at different stages of development. This would provide insight for educators in developing lessons to promote statistical literacy and also what questions they may develop without formal training.

Assessing levels of awareness of statistical literacy components and questions asked of media articles for graduate students is another study to be conducted. Differences in backgrounds of graduate students may contribute to higher levels of awareness.

In addition, research to address the effect of demographic variables such as political party affiliation on the questions that are asked might shed light on how different world views influence the way in which people interpret many of the statistics which they see everyday. Asking students to provide questions concerning statistical information in the media that has a particular political slant and comparing these questions based on students' political party affiliations may indicate the effect of context and political persuasion on statistical literacy.

A wide-scale study with several universities could focus on assisting teachers and educators develop and teach courses focused on statistical literacy. If funded, a comparison of levels of awareness of statistical literacy components and questions asked about media articles from several universities in the United States will be proposed. In addition, a comparison of levels of awareness of statistical literacy components from several countries could be conducted to determine differences due to culture.

Although more research should be conducted to confirm this, statistical literacy may evolve in stages with statistical literacy components representing "stepping stones" or a hierarchy of components. Definition, Lurking variable, and Method may represent the first steps or bottom tier towards statistical literacy while Causality may represent a higher stage in statistical literacy development. Results from this research provide a basis for this hierarchy. Definition, Lurking variable, and Method were the statistical literacy components with the highest levels of awareness after taking SIEL. In addition, significant changes in the level of awareness of all statistical literacy components were observed for all components except Causality. This suggests that Causality is a statistical literacy component which either requires more time to develop or a higher level of processing.

Assets of Study

There were several positive aspects of this study. First, the students in the study represented a broad range of majors, and comparable numbers of males and females made up the sample. The use of an open-ended format to obtain questions students might have about the statistical information in the media articles was an attempt to simulate what might happen in real life because it would not lead students to consider any specific aspects of the media article.

Although an argument might be made that SIEL and those who taught it were ineffective and that might be a reason for low levels of awareness of statistical literacy components even after the course, student evaluations of SIEL have consistently been excellent regardless of the instructor. Both instructors who taught during this study were developers of SIEL and have taught the course for many semesters. Furthermore, these instructors have histories of excellence in teaching and have been recognized for their teaching by being honored with national teaching awards so there is a low probability that the poor results in regard to students' statistical literacy was due to poor instruction.

Limitations of Study

There were several limitations of this study. First, the sample was not randomly selected. Those in the course don't represent a random sample of all college students, though the demographics do suggest the class is representative. Students in the sample group had to complete both pre-course and post-course SATS-36©, give questions to the media articles prior to and after SIEL, and complete the reflection paper. Of the 195 students who completed the course, 73.85% (144 students) completed all aspects of assignments to be included in the sample. There may be important differences in students in the sample and those who did not complete all assignments and excluded from the sample. Based on the results from this study, inferences can only be made concerning students at this particular university who take SIEL and willing to participate in the study.

The results from this study pertained to three specific media articles so context of articles chosen might have had an effect on results. An attempt to reduce this effect was employed with the use of varied contexts of media articles that required limited literacy abilities and might be seen in everyday life experiences. Nevertheless, we do not know how students would react to articles with different contexts.

Although the use of an open-ended format to obtain students' questions to the media articles was an attempt to simulate what might be experienced in real life, students may have had difficulty expressing themselves in this format. Also, students were asked to formulate questions to three media articles and they might have become fatigued completing this for three articles. This may have contributed to students not elaborating on their responses.

Levels of awareness to eight components of statistical literacy were measured using SLCR. To make more consistent evaluations of levels, more directions in determining score levels should be given to evaluators. For this study, the evaluators who were also instructors of SIEL during this study did not have much training using SLCR. This was deliberately done in order not to prejudice the results. However, future studies may include response examples of each score level and media article as a reference for scoring.

Final Thoughts

Based on the results of this study, taking SIEL was not enough to see a significant effect on the level of awareness of the statistical literacy components. Although this one course helped students reach the second tier of the Hierarchical Model (Watson, 1997), it is believed that to get virtually everyone to the third tier would take many experiences.

With an increase in available statistical information through the Internet and other forms of media, statistical literacy, like general literacy, should start in grade school and continue throughout one's educational experience. Time should be given for statistical literacy to develop. In elementary school, lesson plans involving questioning aspects of advertisements could be used to facilitate a "questioning attitude" at a young age. For example, asking questions about an advertisement such as one that indicates the proportion of kids who like a particular brand of cereal could be used as a basis for the lesson.

In middle school mathematics classes, students could be required to find advertisements that make statistical claims to present to the class, and each student would be assigned to write ten questions about the statistical information in the advertisements. Class discussions could focus on these questions and possible answers. Students could then write about their questions and explain why their questions were important to ask. The "advertisement of the month" could be placed on the teacher's website with a count down of the top ten questions.

In high school statistics classes, students could be given current events articles that includes statistical statements or claims. Students could write an essay concerning questions they have about the statistical information and explain why the questions were important to ask. Possible answers to the questions could be discussed during class. In addition, students could be given a current events article and be assigned to either provide justifications or evidence against the statistical information in the article. Then students could present both sides in a "statistical debate." After the debate, students would then evaluate the statistical information in the article and write about their evaluations. Teachers could have students write this paper before and after the debate to determine if there were changes in how the statistical information was evaluated by the students.

In undergraduate and graduate statistics courses, students could be assigned to find journal articles in their field of study and write a paper which focuses on questions they would ask about the statistical information in the article. In addition, students could create a presentation about the journal article and then other students in the class could ask questions about the statistical information. This format would simulate what might happen at a professional meeting or graduate thesis or dissertation defense.

In teacher education classes, prospective teachers could be taught about statistical literacy and activities to promote it. An assignment for these prospective teachers could be to create other activities to enhance statistical literacy. Courses could be developed in teacher education departments that specifically focus on how to promote statistical literacy.

So much time in today's education is spent finding the "right answer" that the ability to question or ask questions is overlooked or even suppressed. I challenge educators to change the focus to one in which the environment to question is encouraged so that it will foster a "questioning attitude."

Statistical literacy should be considered a necessary component of one's educational experience not just for academic endeavors but for citizenship as well. "Freedom of access to statistical information is a way to attenuate information asymmetries and contribute to the progress of a democratic knowledge society, but is ineffective if citizens are not literate, i.e., if they lack the tools to read, understand and analyze statistics; these tools are acquired in school and in lifelong learning, with an interaction between different subject matters and between theory and practice" (Barbieri and Giacche′, 2006, p. 2).

Through the development and teaching of Statistics in Everyday Life, I was hopeful that my students would not see the world in the same way again – they would not readily believe statistics they encounter in their everyday life. They would question statistics they hear about on the news and read on the Internet, newspapers, and magazines and achieve the third tier of the Hierarchical Model (Watson, 1997). Once they have these questions, I hope they have gained the ability and desire to find out the answers to help them make more informed decisions in their lives. Though this one course has improved their attitudes and raised awareness, it was not sufficient to accomplish this goal. Clearly, it takes more.

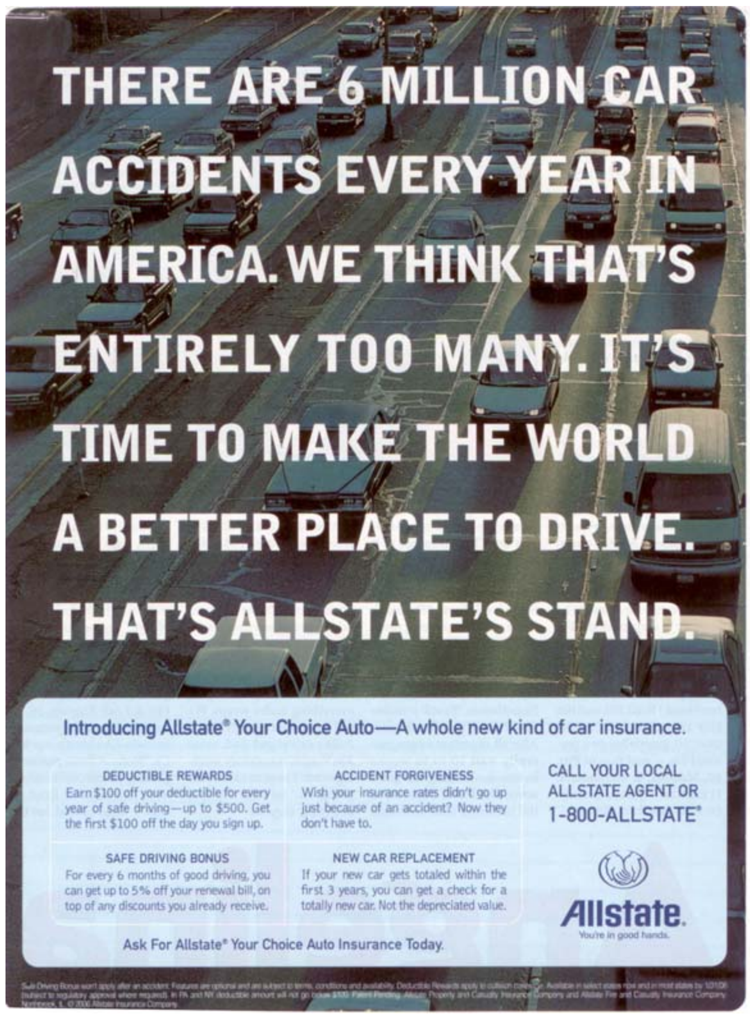
APPENDICES

Appendix A

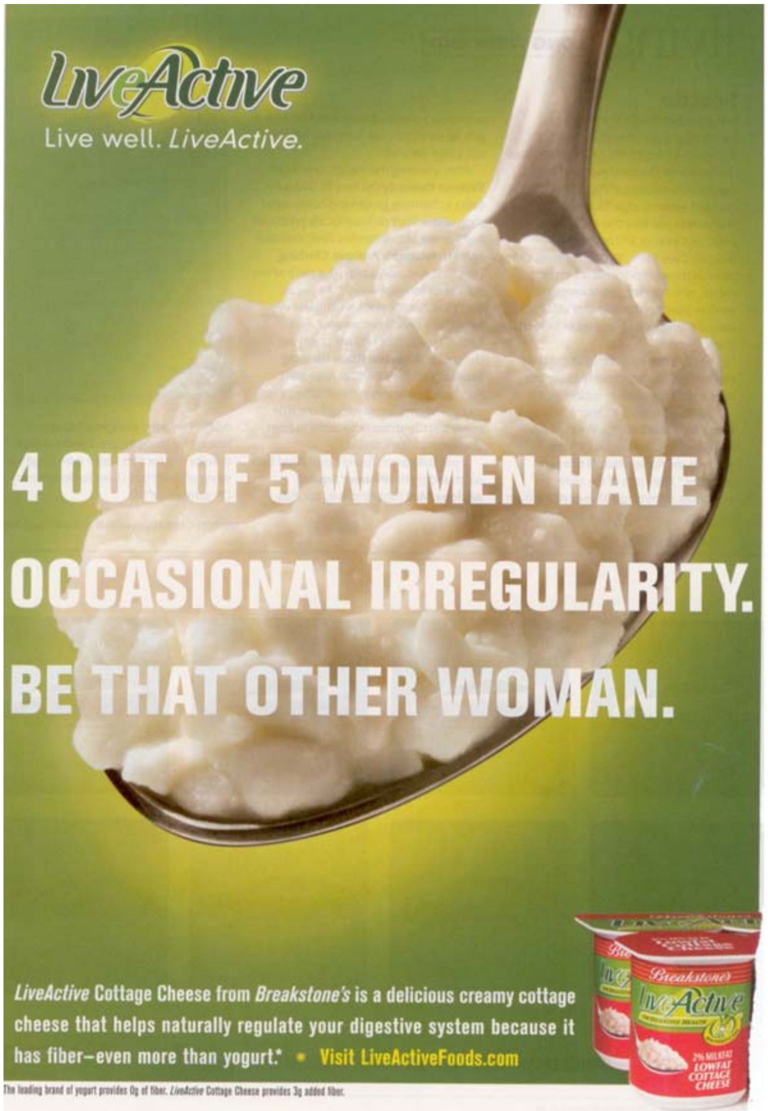
Advertisements and Article for Evaluating Statistical Literacy

Name: ______________________________________

Section: _________



What questions might you have concerning Allstate's claim that there are "6 million car accidents every year in America?" Explain why each question is important to ask.



What questions might you have concerning LiveActive's claim that "4 out of 5 women have occasional irregularity?" Explain why each question is important to ask.

Depression Hits 1 in 10 Teen Girls Each Year

Teen Girls Twice as Likely to Suffer Depression as Boys

By Daniel DeNoon, WebMD Medical News
Reviewed By Brunilda Nazario, MD Wednesday, February 04, 2004

Feb. 4, 2004 -- About one in 10 teen girls suffer a major depression each year, a Canadian study shows.

The study, led by University of Alberta researcher Nancy L. Galambos, PhD, analyzed four years of data from 1,322 boys and girls. At the start of the study, th teens ranged in age from 12 to 19 years old.

Not surprisingly, more than one in five girls admitted to having been depressed sometime in her life. Only one in 10 boys said they'd ever been depressed.

But more probing questions showed that during each year of the study, nearly one in 10 teen girls had a major depressive episode -- about twice the rate of boy

"This is a substantial number of young Canadian women who should be identified as depressed and treated," Galambos says in a news release.

"Very substantial proportions of young people will experience a major depressive episode at some point as they move through adolescence," Galambos and colleagues write in the January issue of the *International Journal of Behavioral Development*. "About twice as many females as males will be so affected."

Galambos notes that depression puts teen girls at risk of anxiety, eating disorders, conduct problems, academic failure, and trouble with relationships.

What questions might you have concerning the above article? Explain why each question is important to ask.

Appendix B

Reflection Paper Questions

Name:

Section:

In this assignment, you will write a reflection paper concerning your experiences in SIEL as well as past experiences. Type your responses to the following questions in the gray boxes. The gray boxes will accommodate as much text as you type. Your responses should be typewritten using **Times New Roman**, a **font size of 12** and **single spaced**. In order to receive credit for this assignment, you must answer each question thoroughly.

1. Please describe past experiences in other mathematics or statistics courses that you have taken.

2. Explain how your experiences in SIEL were similar to your past experiences in mathematics or statistics courses.

3. Explain how your experiences in SIEL were different from your past experiences in mathematics or statistics courses.

4. Explain what "statistical literacy" means to you.

5. Do you feel that you are "statistically literate?" Answer "Yes" or "No" in the space provided.

Explain why you gave this answer.

6. Has SIEL changed the way in which you look at the statistics you encounter in everyday life? (Answer “Yes” or “No” in the space provided.)

7. If your answer to question 6 is “yes,” describe a **specific occurrence** in which you looked at statistics encountered in everyday life differently than you would have before taking SIEL.

8. If your answer to question 6 is “no,” explain why you do not think SIEL has changed the way you look at statistics encountered in everyday life?

9. What attributes do you think are necessary for a student to have in order to do well in SIEL? List these attributes in the space provided.

Explain why you believe these attributes are important to doing well in SIEL.

10. During the course, were you confronted with material that challenged your beliefs? (Answer “Yes” or “No” in the space provided.)

Explain your response in the space below.

11. List 3 adjectives you would use to describe "statistics."

1.

2.

3.

Explain why you selected these adjectives.

12. How would you describe your attitude toward statistics - good, moderate or poor? (Answer "Good", "Moderate" or "Poor" in the space provided.)

Explain what you believe has contributed to your attitude toward statistics.

13. As a result of this class, has your attitude toward statistics has changed?
(Answer “Yes” or “No” in the space provided.)

Explain your response above.

Appendix C

Statistical Literacy Component Rubric (SLCR)

Statistical Literacy Component	**0** **No awareness of component***	**1** **Demonstrated basic (minimal) awareness of component***	**2** **Demonstrated proficient (moderate) awareness of component***	**3** **Demonstrated advanced awareness of component***
Questions concerning the method How was the study conducted? Sampling method used? Type of study… Experiment or Observational? Sample size? Response rate?				
Questions concerning bias Who conducted the study? Agenda? Voluntary response Wording of question(s) Coverage error?				
Questions concerning the reported statistic How was the statistic obtained? Is it a parameter or statistic? Is mean or median appropriate for data if quantitative?				
Questions concerning definitions How is it defined? What does it mean? Are there other definitions? Can you measure what it is defined? (validity) Reliability				
Questions concerning variation What is the MOE? Variation associated with statistic Statistic is not exact. Statistic is an estimate.				
Questions concerning generalizability Can you use information to generalize about population? To what population would be				

appropriate to generalize?				
Questions concerning lurking variable(s) Any information left out? Other factors that may affect results?				
Questions concerning causality Does one factor cause another? Is it implying just correlation or causation?				

Appendix D

Correlation Tables for Research Question 1

Table 1: Correlations and p-values for demographic variables

Variables	Pearson Correlation Coefficient	p-value
Gender*Reported Math SAT score	-0.30311	0.0015
Gender*Reported Verbal SAT score	0.00363	0.9704
Gender*Reported Total SAT score	-0.06221	0.5052
Gender*Number of high school math courses taken	-0.2155	0.0097
Gender*Number of high school statistics courses taken	-0.03721	0.6591
Gender*Number of college math courses taken	-0.17623	0.0346
Gender*Number of college statistics courses taken	-0.11513	0.1694
Gender*Cumulative attempted credits	-0.05494	0.5131
Gender*Cumulative earned credits	-0.052	0.5359
Gender*Cumulative points	-0.02459	0.7698
Gender*Cumulative GPA	0.13719	0.1011
Gender*Course grade	0.08582	0.3064
Gender*Course Effort	0.07371	0.3799
Gender*Reported GPA	0.13164	0.1197
Reported Math SAT score*Reported Verbal SAT score	-0.01613	0.8697
Reported Math SAT score*Total SAT score	0.4169	<0.0001
Reported Math SAT score*Number of high school math courses taken	0.40601	<0.0001
Reported Math SAT score*Number of high school statistics courses taken	0.01299	0.8949
Reported Math SAT score*Number of college math courses taken	0.25389	0.0083
Reported Math SAT score*Number of college statistics courses taken	0.05355	0.5839
Reported Math SAT score*Cumulative attempted credits	0.15808	0.1039
Reported Math SAT score*Cumulative earned credits	0.19527	0.0438
Reported Math SAT score*Cumulative points	0.22182	0.0217

Reported Math SAT score*Cumulative GPA	0.35026	0.0002
Reported Math SAT score*Course grade	0.34057	0.0003
Reported Math SAT score*Course Effort	0.11388	0.2428
Reported Math SAT score*Reported GPA	0.20306	0.0368
Reported Verbal SAT score*Reported Total SAT score	0.56891	<0.0001
Reported Verbal SAT* Number of high school math courses taken	-0.10346	0.2913
Reported Verbal SAT *Number of high school statistics courses taken	-0.14059	0.1506
Reported Verbal SAT *Number of college math courses taken	-0.10375	0.2876
Reported Verbal SAT *Number of college statistics courses taken	-0.06471	0.5079
Reported Verbal SAT *Cumulative attempted credits	0.04304	0.6598
Reported Verbal SAT *Cumulative earned credits	0.11098	0.2551
Reported Verbal SAT *Cumulative points	0.08275	0.3968
Reported Verbal SAT *Cumulative GPA	0.29361	0.0021
Reported Verbal SAT *Course grade	0.0954	0.3284
Reported Verbal SAT *Course Effort	-0.22864	0.0179
Reported Verbal SAT *Reported GPA	0.23784	0.0141
Reported Total SAT score * Number of high school math courses taken	0.00584	0.9504
Reported Total SAT score *Number of high school statistics courses taken	-0.18723	0.0442
Reported Total SAT score *Number of college math courses taken	-0.00918	0.9217
Reported Total SAT score *Number of college statistics courses taken	-0.0984	0.2912
Reported Total SAT score *Cumulative attempted credits	-0.00635	0.9458
Reported Total SAT score *Cumulative earned credits	0.08925	0.3386
Reported Total SAT score	0.03678	0.6938

*Cumulative points		
Reported Total SAT score *Cumulative GPA	0.30219	0.0009
Reported Total SAT score *Course grade	0.06166	0.509
Reported Total SAT score *Course Effort	-0.11651	0.2109
Reported Total SAT score *Reported GPA	0.20890	0.0244
Number of high school math courses taken *Number of high school statistics courses taken	0.39614	<0.0001
Number of high school math courses taken *Number of college math courses taken	0.20132	0.0159
Number of high school math courses taken *Number of college statistics courses taken	0.41713	<0.0001
Number of high school math courses taken *Cumulative attempted credits	0.12943	0.1234
Number of high school math courses taken *Cumulative earned credits	0.11497	0.1715
Number of high school math courses taken *Cumulative points	0.1022	0.2245
Number of high school math courses taken *Cumulative GPA	0.12653	0.1321
Number of high school math courses taken *Course grade	0.02509	0.7661
Number of high school math courses taken *Course Effort	0.00459	0.9566
Number of high school math courses taken *Reported GPA	0.05386	0.5273
Number of high school statistics courses taken *Number of college math courses taken	0.01413	0.867
Number of high school statistics courses taken *Number of college statistics courses taken	0.30708	0.0002
Number of high school statistics courses taken *Cumulative attempted credits	-0.11191	0.1833
Number of high school statistics courses taken *Cumulative earned	-0.05611	0.5057

credits		
Number of high school math courses taken *Cumulative points	-0.15708	0.0610
Number of high school statistics courses taken *Cumulative GPA	-0.09741	0.2471
Number of high school statistics courses taken *Course grade	-0.03114	0.7120
Number of high school statistics courses taken *Course Effort	0.00996	0.9060
Number of high school statistics courses taken *Reported GPA	-0.00626	0.9415
Number of college math courses taken*Number of college statistics courses taken	0.42299	<0.0001
Number of college math courses taken *Cumulative attempted credits	0.47443	<0.0001
Number of college math courses taken *Cumulative earned credits	0.53057	<0.0001
Number of college math courses taken *Cumulative points	0.46366	<0.0001
Number of college math courses taken *Cumulative GPA	0.06258	0.4562
Number of college math courses taken *Course grade	0.11889	0.1558
Number of college math courses taken *Course Effort	-0.01398	0.8680
Number of college math courses taken *Reported GPA	-0.02670	0.7533
Number of college statistics courses taken *Cumulative attempted credits	0.34525	<0.0001
Number of college statistics courses taken *Cumulative earned credits	0.45286	<0.0001
Number of college statistics courses taken *Cumulative points	0.33175	<0.0001
Number of college statistics courses taken *Cumulative GPA	0.01980	0.8138
Number of college statistics courses taken *Course grade	0.01625	0.8467
Number of college statistics courses taken *Course Effort	-0.03557	0.6721
Number of college statistics courses taken *Reported GPA	0.10442	0.2178
Cumulative attempted	0.84056	<0.0001

points*Cumulative earned credits		
Cumulative attempted points*Cumulative points	0.95905	<0.0001
Cumulative attempted points*Cumulative GPA	0.16655	0.0460
Cumulative attempted points*Course grade	-0.05759	0.4930
Cumulative attempted points*Course Effort	-0.06270	0.4554
Cumulative attempted points*Reported GPA	-0.0323	0.7037
Cumulative earned credits *Cumulative points	0.83699	<0.0001
Cumulative earned credits *Cumulative GPA	0.24307	0.0033
Cumulative earned credits *Course grade	-0.01682	0.8414
Cumulative earned credits *Course Effort	-0.07588	0.3661
Cumulative earned credits *Reported GPA	0.12091	0.1532
Cumulative points*Cumulative GPA	0.30683	0.0002
Cumulative points *Course grade	0.09787	0.2432
Cumulative points *Course Effort	-0.00339	0.9678
Cumulative points *Reported GPA	0.16872	0.0455
Cumulative GPA*Course grade	0.42530	<0.0001
Cumulative GPA*Course Effort	0.19592	0.0186
Cumulative GPA*Reported GPA	0.69079	<0.0001
Course grade*Course Effort	0.36065	<0.0001
Course grade*Reported GPA	0.54274	<0.0001
Effort*Reported GPA	0.22342	0.0077

Table 2: Correlations for demographic variables with pre-course attitude component scores from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Gender*Affect	-0.25601	0.0024
Gender*Cognitive competence	-0.16657	0.046
Gender*Value	-0.09928	0.2381
Gender*Difficulty	-0.0815	0.3332
Gender*Interest	-0.10693	0.2021
Gender*Effort	0.25504	0.0022
Reported Math SAT score*Affect	0.39589	<0.0001
Reported Math SAT score*Cognitive competence	0.37077	<0.0001
Reported Math SAT score*Value	0.12942	0.184
Reported Math SAT score*Difficulty	0.29205	0.0024
Reported Math SAT score*Interest	0.13949	0.1519
Reported Math SAT score*Effort	-0.28494	0.0031
Reported Verbal SAT score*Affect	-0.06752	0.5001
Reported Verbal SAT* Cognitive competence	-0.05954	0.5424
Reported Verbal SAT *Value	-0.07633	0.4346
Reported Verbal SAT *Difficulty	-0.0195	0.8427
Reported Verbal SAT *Interest	-0.09589	0.3258
Reported Verbal SAT *Effort	-0.17177	0.0783
Reported Total SAT score * Affect	0.15103	0.1119
Reported Total SAT score *Cognitive competence	0.14218	0.1262
Reported Total SAT score *Value	-0.04432	0.6351
Reported Total SAT score *Difficulty	0.08495	0.3646
Reported Total SAT score *Interest	-0.05936	0.5249
Reported Total SAT score *Effort	-0.23223	0.0121

Number of high school math courses taken *Affect	0.1314	0.1259
Number of high school math courses taken *Cognitive competence	0.06497	0.4408
Number of high school math courses taken *Value	0.10026	0.2352
Number of high school math courses taken *Difficulty	0.03481	0.6809
Number of high school math courses taken *Interest	0.00556	0.9474
Number of high school math courses taken *Effort	-0.07118	0.4016
Number of high school statistics courses taken *Affect	0.0369	0.6686
Number of high school statistics courses taken *Cognitive competence	-0.04055	0.6306
Number of high school statistics courses taken *Value	0.12961	0.1242
Number of high school statistics courses taken *Difficulty	-0.00492	0.9536
Number of high school math courses taken *Interest	-0.02881	0.7327
Number of high school statistics courses taken *Effort	-0.0751	0.3761
Number of college math courses taken*Affect	0.3468	<0.0001
Number of college math courses taken *Cognitive competence	0.34418	<0.0001
Number of college math courses taken *Value	0.20808	0.0126
Number of college math courses taken *Difficulty	0.29572	0.0003
Number of college math courses taken *Interest	0.1631	0.0508
Number of college math courses taken *Effort	-0.15323	0.0687
Number of college statistics courses taken *Affect	0.07132	0.4058
Number of college statistics courses taken *Cognitive competence	0.06203	0.4601

Number of college statistics courses taken *Value	0.12929	0.1238
Number of college statistics courses taken *Difficulty	-0.10783	0.1999
Number of college statistics courses taken *Interest	0.14455	0.0839
Number of college statistics courses taken *Effort	-0.05506	0.5152
Cumulative attempted points*Affect	0.05963	0.4872
Cumulative attempted points*Cognitive competence	0.05993	0.4755
Cumulative attempted points*Value	0.03893	0.6444
Cumulative attempted points*Difficulty	-0.00928	0.9124
Cumulative attempted points*Interest	-0.02524	0.7639
Cumulative attempted points*Effort	0.02399	0.7768
Cumulative earned credits *Affect	0.0701	0.414
Cumulative earned credits *Cognitive competence	0.05082	0.5453
Cumulative earned credits *Value	0.11288	0.1795
Cumulative earned credits *Difficulty	-0.02392	0.7767
Cumulative earned credits *Interest	0.01551	0.8536
Cumulative earned credits *Effort	-0.05266	0.5337
Cumulative points*Affect	0.07992	0.3515
Cumulative points *Cognitive competence	0.0981	0.2421
Cumulative points *Value	0.08014	0.3414
Cumulative points *Difficulty	0.03445	0.683
Cumulative points *Interest	0.01957	0.8159
Cumulative points *Effort	0.00686	0.9355
Cumulative GPA*Affect	0.5236	0.5419
Cumulative GPA*Cognitive competence	0.16841	0.0436
Cumulative GPA*Value	0.0304	0.7185

Cumulative GPA*Difficulty	0.08311	0.3237
Cumulative GPA*Interest	-0.0391	0.6418
Cumulative GPA*Effort	-0.13811	0.1012
Course grade*Affect	0.12368	0.1484
Course grade*Cognitive competence	0.25312	0.0022
Course grade*Value	0.19899	0.0172
Course grade*Difficulty	0.23639	0.0045
Course grade*Interest	0.04993	0.5523
Course grade*Effort	-0.0969	0.2513
Course Effort*Affect	-0.00724	0.9329
Course Effort*Cognitive competence	0.00504	0.9522
Course Effort*Value	0.05041	0.5499
Course Effort*Difficulty	-0.04753	0.573
Course Effort*Interest	0.1146	0.1714
Course Effort*Effort	0.19165	0.0223
Reported GPA*Affect	-0.05969	0.4916
Reported GPA*Cognitive competence	0.0088	0.9168
Reported GPA*Value	0.06848	0.4215
Reported GPA*Difficulty	0.0502	0.5558
Reported GPA*Interest	0.04346	0.6089
Reported GPA*Effort	-0.0903	0.2904

Table 3: Correlations and p-values between demographic variables and pre-course average statistical literacy component scores from the Statistical Literacy Components Rubric

Variables	Pearson Correlation Coefficient	p-value
Pre-course Bias*Gender	-0.12955	0.1217
Pre-course Bias*Reported Math SAT score	0.13868	0.1543
Pre-course Bias*Reported Verbal SAT score	0.01305	0.8939
Pre-course Bias*Reported Total SAT score	0.03947	0.6726
Pre-course Bias*Number of high school math courses taken	0.08174	0.3318
Pre-course Bias*Number of high school statistics courses taken	-0.01511	0.8578
Pre-course Bias*Number of college math courses taken	0.05292	0.5287
Pre-course Bias*Number of college statistics courses taken	-0.01272	0.8797
Pre-course Bias* Cumulative attempted credits	0.06460	0.4417
Pre-course Bias* Cumulative earned credits	0.01889	0.8222
Pre-course Bias* Cumulative points	0.05565	0.5076
Pre-course Bias* Cumulative GPR	0.05903	0.4821
Pre-course Bias*Course grade	-0.10726	0.2007
Pre-course Bias*Course Effort	-0.04606	0.5836
Pre-course Bias*Reported GPR	0.01196	0.8880
Pre-course Causality*Gender	-0.11165	0.1828
Pre-course Causality *Reported Math SAT score	0.07223	0.4597
Pre-course Causality *Reported Verbal SAT score	-0.01297	0.8945
Pre-course Causality *Reported Total SAT score	0.02158	0.8174
Pre-course Causality *Number of high school math	0.01125	0.8939

courses taken		
Pre-course Causality *Number of high school statistics courses taken	-0.02510	0.7660
Pre-course Causality *Number of college math courses taken	-0.07678	0.3604
Pre-course Causality *Number of college statistics courses taken	-0.10529	0.2091
Pre-course Causality * Cumulative attempted credits	-0.12484	0.1360
Pre-course Causality * Cumulative earned credits	-0.20297	0.0147
Pre-course Causality * Cumulative points	-0.12726	0.1285
Pre-course Causality * Cumulative GPR	0.08768	0.2960
Pre-course Causality *Course grade	0.05346	0.5245
Pre-course Causality *Course Effort	0.17735	0.0335
Pre-course Causality *Reported GPR	0.01062	0.9005
Pre-course Definition*Gender	-0.10601	0.2060
Pre-course Definition *Reported Math SAT score	0.0978	0.3162
Pre-course Definition *Reported Verbal SAT score	0.28285	0.0032
Pre-course Definition *Reported Total SAT score	0.12568	0.1770
Pre-course Definition *Number of high school math courses taken	-0.04199	0.6186
Pre-course Definition *Number of high school statistics courses taken	-0.02920	0.7292
Pre-course Definition *Number of college math courses taken	0.04539	0.5891
Pre-course Definition *Number of college statistics courses taken	0.08930	0.2871

Pre-course Definition * Cumulative attempted credits	0.09779	0.2436
Pre-course Definition * Cumulative earned credits	0.17544	0.0354
Pre-course Definition * Cumulative points	0.14497	0.0830
Pre-course Definition * Cumulative GPR	0.09125	0.2767
Pre-course Definition *Course grade	0.26698	0.0012
Pre-course Definition *Course Effort	-0.02546	0.7620
Pre-course Definition *Reported GPA	0.25094	0.0027
Pre-course Generalize*Gender	0.03216	0.7020
Pre-course Generalize *Reported Math SAT score	0.16823	0.0832
Pre-course Generalize *Reported Verbal SAT score	-0.04809	0.6228
Pre-course Generalize *Reported Total SAT score	0.01036	0.9118
Pre-course Generalize *Number of high school math courses taken	0.13020	0.1212
Pre-course Generalize *Number of high school statistics courses taken	-0.08668	0.3033
Pre-course Generalize *Number of college math courses taken	0.23524	0.0045
Pre-course Generalize *Number of college statistics courses taken	0.20282	0.0148
Pre-course Generalize * Cumulative attempted credits	0.06587	0.4328
Pre-course Generalize * Cumulative earned credits	0.12438	0.1375
Pre-course Generalize * Cumulative points	0.08236	0.3264
Pre-course Generalize * Cumulative GPA	0.10631	0.2047
Pre-course Generalize *Course grade	0.14100	0.0919

Pre-course Generalize *Course Effort	-0.00859	0.9186
Pre-course Generalize *Reported GPA	0.09406	0.2673
Pre-course Lurking variable*Gender	0.21778	0.0087
Pre-course Lurking variable *Reported Math SAT score	-0.22291	0.021
Pre-course Lurking variable *Reported Verbal SAT score	0.05797	0.5531
Pre-course Lurking variable *Reported Total SAT score	0.01096	0.9066
Pre-course Lurking variable *Number of high school math courses taken	-0.05454	0.5176
Pre-course Lurking variable *Number of high school statistics courses taken	0.03645	0.6656
Pre-course Lurking variable *Number of college math courses taken	-0.09757	0.2447
Pre-course Lurking variable *Number of college statistics courses taken	-0.003	0.9716
Pre-course Lurking variable * Cumulative attempted credits	-0.10898	0.1935
Pre-course Lurking variable * Cumulative earned credits	-0.09555	0.2546
Pre-course Lurking variable * Cumulative points	-0.10008	0.2327
Pre-course Lurking variable * Cumulative GPA	-0.03551	0.6726
Pre-course Lurking variable *Course grade	0.10183	0.2246
Pre-course Lurking variable *Course Effort	0.13425	0.1087
Pre-course Lurking variable *Reported GPA	-0.00323	0.9697
Pre-course Method*Gender	-0.00715	0.9322
Pre-course Method *Reported Math SAT score	0.0306	0.7544
Pre-course Method *Reported Verbal SAT score	0.04666	0.6332

Pre-course Method *Reported Total SAT score	0.02686	0.7738
Pre-course Method *Number of high school math courses taken	0.02355	0.7801
Pre-course Method *Number of high school statistics courses taken	0.14055	0.0941
Pre-course Method *Number of college math courses taken	-0.09963	0.2348
Pre-course Method *Number of college statistics courses taken	0.06578	0.4335
Pre-course Method * Cumulative attempted credits	-0.02246	0.7893
Pre-course Method * Cumulative earned credits	0.0102	0.9035
Pre-course Method * Cumulative points	0.007	0.9336
Pre-course Method * Cumulative GPA	0.01945	0.8170
Pre-course Method *Course grade	0.12799	0.1263
Pre-course Method *Course Effort	-0.06863	0.4137
Pre-course Method *Reported GPA	0.03729	0.6607
Pre-course Reported statistic*Gender	0.03430	0.6831
Pre-course Reported statistic *Reported Math SAT score	0.03067	0.7538
Pre-course Reported statistic *Reported Verbal SAT score	-0.03108	0.7506
Pre-course Reported statistic *Reported Total SAT score	-0.06945	0.4568
Pre-course Reported statistic *Number of high school math courses taken	0.00719	0.9321
Pre-course Reported statistic *Number of high school statistics courses taken	-0.08186	0.3311
Pre-course Reported statistic *Number of college math	-0.13034	0.1195

courses taken		
Pre-course Reported statistic *Number of college statistics courses taken	-0.10834	0.1962
Pre-course Reported statistic * Cumulative attempted credits	-0.01503	0.8581
Pre-course Reported statistic * Cumulative earned credits	-0.04153	0.6211
Pre-course Reported statistic * Cumulative points	0.00425	0.9596
Pre-course Reported statistic * Cumulative GPA	-0.00553	0.9475
Pre-course Reported statistic *Course grade	-0.0707	0.3998
Pre-course Reported statistic *Course Effort	-0.05610	0.5042
Pre-course Reported statistic *Reported GPA	-0.05091	0.5488
Pre-course Variation*Gender	-0.00598	0.9433
Pre-course Variation *Reported Math SAT score	0.09979	0.3065
Pre-course Variation *Reported Verbal SAT score	-0.00444	0.9638
Pre-course Variation *Reported Total SAT score	0.03008	0.7475
Pre-course Variation *Number of high school math courses taken	0.15635	0.0622
Pre-course Variation *Number of high school statistics courses taken	0.11499	0.1715
Pre-course Variation *Number of college math courses taken	0.07372	0.3799
Pre-course Variation *Number of college statistics courses taken	0.29808	0.0003
Pre-course Variation * Cumulative attempted credits	-0.06929	0.4092
Pre-course Variation * Cumulative earned credits	0.04952	0.5555
Pre-course Variation * Cumulative points	-0.07724	0.3575

Pre-course Variation * Cumulative GPA	-0.05173	0.5381
Pre-course Variation *Course grade	-0.00608	0.9424
Pre-course Variation *Course Effort	-0.11652	0.1643
Pre-course Variation *Reported GPA	-0.02966	0.7270

Table 4: Correlations and p-values between pre-course average statistical literacy components from the Statistical Literacy Components Rubric and pre-course attitude component scores from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Pre-course Bias*Pre-course affect	0.04999	0.5604
Pre-course Bias*Pre-course cognitive competence	0.01448	0.8632
Pre-course Bias*Pre-course value	0.00696	0.9342
Pre-course Bias*Pre-course difficulty	0.06972	0.4080
Pre-course Bias*Pre-course interest	-0.00588	0.9442
Pre-course Bias*Pre-course effort	-0.14170	0.0925
Pre-course Causality*Pre-course affect	0.04574	0.5942
Pre-course Causality *Pre-course cognitive competence	0.04019	0.6325
Pre-course Causality *Pre-course value	-0.07033	0.4039
Pre-course Causality *Pre-course difficulty	-0.04057	0.6305
Pre-course Causality *Pre-course interest	-0.04317	0.6074
Pre-course Causality *Pre-course effort	-0.07617	0.3676
Pre-course Definition*Pre-course affect	0.01856	0.8290
Pre-course Definition *Pre-course cognitive competence	0.07725	0.3574
Pre-course Definition *Pre-course value	0.14672	0.0804
Pre-course Definition *Pre-course difficulty	0.14469	0.0847
Pre-course Definition *Pre-course interest	0.05161	0.5390
Pre-course Definition *Pre-course effort	-0.02555	0.7628
Pre-course Generalize*Pre-course affect	0.19947	0.019

Pre-course Generalize *Pre-course cognitive competence	0.23077	0.0054
Pre-course Generalize *Pre-course value	0.07429	0.3779
Pre-course Generalize *Pre-course difficulty	0.13095	0.1190
Pre-course Generalize *Pre-course interest	0.08576	0.3068
Pre-course Generalize *Pre-course effort	-0.05727	0.4985
Pre-course Lurking variable*Pre-course affect	0.05393	0.5299
Pre-course Lurking variable *Pre-course cognitive competence	0.01453	0.8628
Pre-course Lurking variable *Pre-course value	0.04639	0.5822
Pre-course Lurking variable *Pre-course difficulty	0.11282	0.1797
Pre-course Lurking variable *Pre-course interest	0.0200	0.8119
Pre-course Lurking variable *Pre-course effort	0.07814	0.3553
Pre-course Method*Pre-course affect	-0.01888	0.8261
Pre-course Method *Pre-course cognitive competence	-0.04921	0.5581
Pre-course Method *Pre-course value	0.03823	0.6503
Pre-course Method *Pre-course difficulty	-0.10304	0.2207
Pre-course Method *Pre-course interest	-0.04295	0.6092
Pre-course Method *Pre-course effort	0.00574	0.9459
Pre-course Reported statistic*Pre-course affect	-0.05315	0.5359
Pre-course Reported statistic *Pre-course cognitive competence	-0.04428	0.5982
Pre-course Reported statistic *Pre-course value	0.09503	0.2589
Pre-course Reported statistic	-0.09894	0.2398

*Pre-course difficulty		
Pre-course Reported statistic *Pre-course interest	0.07876	0.3481
Pre-course Reported statistic *Pre-course effort	0.17006	0.0430
Pre-course Variation*Pre-course affect	0.11972	0.1619
Pre-course Variation *Pre-course cognitive competence	0.11914	0.1549
Pre-course Variation *Pre-course value	0.21918	0.0085
Pre-course Variation *Pre-course difficulty	0.04227	0.6162
Pre-course Variation *Pre-course interest	0.12666	0.1303
Pre-course Variation *Pre-course effort	-0.13256	0.1158

Table 5: Correlations and p-values between pre-course average statistical literacy components from Statistical Literacy Components Rubric and post-course attitude components from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Pre-course Bias*Post-course affect	0.00346	0.9673
Pre-course Bias* Post-course cognitive competence	-0.00893	0.9163
Pre-course Bias*Post-course value	-0.12111	0.1482
Pre-course Bias* Post-course difficulty	0.08626	0.3074
Pre-course Bias* Post-course interest	0.00286	0.9730
Pre-course Bias* Post-course effort	-0.09043	0.2845
Pre-course Causality* Post-course affect	0.13047	0.1204
Pre-course Causality * Post-course cognitive competence	0.08106	0.3393
Pre-course Causality * Post-course value	0.05977	0.4767
Pre-course Causality * Post-course difficulty	-0.0412	0.6264
Pre-course Causality * Post-course interest	0.04665	0.5814
Pre-course Causality * Post-course effort	0.04980	0.5561
Pre-course Definition* Post-course affect	0.07018	0.4049
Pre-course Definition * Post-course cognitive competence	0.11383	0.1790
Pre-course Definition * Post-course value	0.20234	0.015
Pre-course Definition * Post-course difficulty	0.09119	0.2804
Pre-course Definition * Post-course interest	0.13080	0.1208
Pre-course Definition * Post-course effort	-0.04571	0.5891
Pre-course Generalize* Post-course affect	0.1775	0.0339

Pre-course Generalize * Post-course cognitive competence	0.24207	0.0038
Pre-course Generalize * Post-course value	0.07074	0.3995
Pre-course Generalize * Post-course difficulty	0.20137	0.0163
Pre-course Generalize * Post-course interest	-0.00817	0.9231
Pre-course Generalize * Post-course effort	-0.14375	0.0879
Pre-course Lurking variable* Post-course affect	0.08476	0.3142
Pre-course Lurking variable * Post-course cognitive competence	0.01408	0.8684
Pre-course Lurking variable * Post-course value	0.05623	0.5032
Pre-course Lurking variable * Post-course difficulty	0.05861	0.4884
Pre-course Lurking variable * Post-course interest	0.09280	0.2720
Pre-course Lurking variable * Post-course effort	0.0607	0.4730
Pre-course Method* Post-course affect	0.004	0.9622
Pre-course Method * Post-course cognitive competence	0.00849	0.9204
Pre-course Method * Post-course value	0.01346	0.8728
Pre-course Method * Post-course difficulty	0.01503	0.8591
Pre-course Method * Post-course interest	-0.04752	0.5744
Pre-course Method * Post-course effort	-0.07390	0.3821
Pre-course Reported statistic* Post-course affect	-0.06029	0.4745
Pre-course Reported statistic * Post-course cognitive competence	-0.10059	0.2353
Pre-course Reported statistic * Post-course value	-0.1148	0.1706
Pre-course Reported statistic *	-0.12603	0.1350

Post-course difficulty		
Pre-course Reported statistic * Post-course interest	0.07963	0.3462
Pre-course Reported statistic * Post-course effort	0.09824	0.2448
Pre-course Variation* Post-course affect	0.07025	0.4044
Pre-course Variation * Post-course cognitive competence	0.07156	0.3991
Pre-course Variation * Post-course value	0.04656	0.5795
Pre-course Variation * Post-course difficulty	-0.00376	0.9646
Pre-course Variation * Post-course interest	0.14836	0.0781
Pre-course Variation * Post-course effort	-0.17470	0.0376

Table 6: Correlations between demographic variables with post-course attitude component scores from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Gender*Affect	-0.18297	0.0287
Gender*Cognitive competence	-0.16719	0.0475
Gender*Value	-0.08166	0.3305
Gender*Difficulty	-0.11391	0.1771
Gender*Interest	-0.1221	0.1478
Gender*Effort	0.23538	0.0048
Reported Math SAT score*Affect	0.33459	0.0004
Reported Math SAT score*Cognitive competence	0.39806	<0.0001
Reported Math SAT score*Value	0.16462	0.0902
Reported Math SAT score*Difficulty	0.26414	0.0062
Reported Math SAT score*Interest	0.18292	0.0605
Reported Math SAT score*Effort	-0.22927	0.0175
Reported Verbal SAT score*Affect	0.0218	0.8237
Reported Verbal SAT* Cognitive competence	-0.01006	0.9189
Reported Verbal SAT *Value	0.08221	0.3999
Reported Verbal SAT *Difficulty	-0.05752	0.5581
Reported Verbal SAT *Interest	0.06472	0.5098
Reported Verbal SAT *Effort	-0.07132	0.4654
Reported Total SAT score * Affect	0.0512	0.5835
Reported Total SAT score *Cognitive competence	0.0826	0.3823
Reported Total SAT score *Value	-0.0294	0.7531
Reported Total SAT score *Difficulty	-0.02988	0.7512
Reported Total SAT score *Interest	-0.01776	0.8506
Reported Total SAT score *Effort	-0.12428	0.1819
Number of high school math courses taken *Affect	0.03265	0.6997
Number of high school math courses taken *Cognitive competence	-0.01206	0.8876
Number of high school math	0.1813	0.8298

courses taken *Value		
Number of high school math courses taken *Difficulty	0.00834	0.9218
Number of high school math courses taken *Interest	0.05923	0.4854
Number of high school math courses taken *Effort	-0.00538	0.9495
Number of high school statistics courses taken *Affect	-0.00235	0.9778
Number of high school statistics courses taken *Cognitive competence	-0.0781	0.359
Number of high school statistics courses taken *Value	0.08508	0.3124
Number of high school statistics courses taken *Difficulty	-0.07507	0.3763
Number of high school math courses taken *Interest	0.09473	0.2638
Number of high school statistics courses taken *Effort	-0.00695	0.9348
Number of college math courses taken*Affect	0.21101	0.0114
Number of college math courses taken *Cognitive competence	0.23297	0.0054
Number of college math courses taken *Value	0.20882	0.012
Number of college math courses taken *Difficulty	0.1924	0.0218
Number of college math courses taken *Interest	0.23444	0.005
Number of college math courses taken *Effort	-0.08836	0.2957
Number of college statistics courses taken *Affect	-0.0135	0.8728
Number of college statistics courses taken *Cognitive competence	-0.00748	0.9299
Number of college statistics courses taken *Value	0.19567	0.0188
Number of college statistics courses taken *Difficulty	-0.07999	0.344
Number of college statistics courses taken *Interest	0.19036	0.0233

Number of college statistics courses taken *Effort	-0.0493	0.5602
Cumulative attempted points*Affect	-0.06451	0.444
Cumulative attempted points*Cognitive competence	-0.04403	0.6041
Cumulative attempted points*Value	0.06442	0.443
Cumulative attempted points*Difficulty	-0.00931	0.9125
Cumulative attempted points*Interest	0.15472	0.066
Cumulative attempted points*Effort	0.01407	0.868
Cumulative earned credits *Affect	-0.04589	0.5863
Cumulative earned credits *Cognitive competence	-0.00558	0.9477
Cumulative earned credits *Value	0.16124	0.0535
Cumulative earned credits *Difficulty	-0.0276	0.7444
Cumulative earned credits *Interest	0.16454	0.0504
Cumulative earned credits *Effort	-0.05685	0.5016
Cumulative points*Affect	0.0073	0.9311
Cumulative points *Cognitive competence	0.04088	0.6303
Cumulative points *Value	0.11671	0.1636
Cumulative points *Difficulty	0.05255	0.5345
Cumulative points *Interest	0.18277	0.0295
Cumulative points *Effort	0.01273	0.8804
Cumulative GPA*Affect	0.15143	0.071
Cumulative GPA*Cognitive competence	0.23872	0.0044
Cumulative GPA*Value	0.07986	0.3414
Cumulative GPA*Difficulty	0.17395	0.0384
Cumulative GPA*Interest	0.04583	0.5881
Cumulative GPA*Effort	-0.02532	0.7649
Course grade*Affect	0.43058	<0.0001
Course grade*Cognitive competence	0.51016	<0.0001
Course grade*Value	0.35759	<0.0001
Course grade*Difficulty	0.37855	<0.0001
Course grade*Interest	0.21088	0.0118

Course grade*Effort	0.0475	0.5746
Course Effort*Affect	0.1007	0.2315
Course Effort*Cognitive competence	0.06145	0.4691
Course Effort*Value	0.12821	0.1257
Course Effort*Difficulty	-0.01166	0.8905
Course Effort*Interest	0.14369	0.088
Course Effort*Effort	0.38672	<0.0001
Reported GPA*Affect	0.0942	0.2683
Reported GPA*Cognitive competence	0.16822	0.0486
Reported GPA*Value	0.15014	0.0756
Reported GPA*Difficulty	0.08976	0.2933
Reported GPA*Interest	0.09529	0.2645
Reported GPA*Effort	0.01131	0.8949

Table 7: Correlations and p-values between demographic variables and post-course average statistical literacy component scores from the Statistical Literacy Components Rubric

Variables	Pearson Correlation Coefficient	p-value
Post-course Bias*Gender	0.07404	0.3778
Post-course Bias*Reported Math SAT score	0.06856	0.4829
Post-course Bias*Reported Verbal SAT score	0.13756	0.1577
Post-course Bias *Reported Total SAT score	0.03690	0.6928
Post-course Bias*Number of high school math courses taken	0.05957	0.4797
Post-course Bias*Number of high school statistics courses taken	0.04680	0.5789
Post-course Bias*Number of college math courses taken	-0.11023	0.1884
Post-course Bias*Number of college statistics courses taken	-0.06312	0.4523
Post-course Bias* Cumulative attempted credits	-0.06707	0.4245
Post-course Bias* Cumulative earned credits	-0.05270	0.5305
Post-course Bias* Cumulative points	-0.05491	0.5133
Post-course Bias* Cumulative GPA	0.07269	0.3866
Post-course Bias*Course grade	0.05038	0.5487
Post-course Bias*Course Effort	-0.08477	0.3124
Post-course Bias*Reported GPA	0.12463	0.1409
Post-course Causality*Gender	-0.03949	0.6384
Post-course Causality *Reported Math SAT score	-0.01928	0.8437
Post-course Causality *Reported Verbal SAT score	-0.0055	0.9552
Post-course Causality *Reported Total SAT score	0.00710	0.9395

Post-course Causality *Number of high school math courses taken	-0.01292	0.8783
Post-course Causality *Number of high school statistics courses taken	0.14928	0.0752
Post-course Causality *Number of college math courses taken	0.01949	0.8167
Post-course Causality *Number of college statistics courses taken	0.11698	0.1626
Post-course Causality * Cumulative attempted credits	0.03528	0.6747
Post-course Causality * Cumulative earned credits	0.05832	0.4875
Post-course Causality * Cumulative points	0.03248	0.6992
Post-course Causality * Cumulative GPA	0.06790	0.4187
Post-course Causality *Course grade	0.05402	0.5202
Post-course Causality *Course Effort	0.04202	0.6170
Post-course Causality *Reported GPA	0.06758	0.4259
Post-course Definition*Gender	0.19198	0.0212
Post-course Definition *Reported Math SAT score	0.20910	0.0307
Post-course Definition *Reported Verbal SAT score	0.13844	0.1550
Post-course Definition *Reported Total SAT score	0.23130	0.0121
Post-course Definition *Number of high school math courses taken	-0.09620	0.2531
Post-course Definition *Number of high school statistics courses taken	-0.02968	0.7249
Post-course Definition *Number of college math courses taken	0.01113	0.8946

Post-course Definition *Number of college statistics courses taken	-0.06805	0.4177
Post-course Definition * Cumulative attempted credits	-0.00168	0.9841
Post-course Definition * Cumulative earned credits	-0.00905	0.9142
Post-course Definition * Cumulative points	0.03385	0.6871
Post-course Definition * Cumulative GPA	0.12551	0.1339
Post-course Definition *Course grade	0.19146	0.0215
Post-course Definition *Course Effort	0.06852	0.4145
Post-course Definition *Reported GPA	0.19872	0.0182
Post-course Generalize*Gender	0.00690	0.9346
Post-course Generalize *Reported Math SAT score	-0.05487	0.5745
Post-course Generalize *Reported Verbal SAT score	-0.06772	0.4883
Post-course Generalize *Reported Total SAT score	-0.10969	0.2391
Post-course Generalize *Number of high school math courses taken	-0.00292	0.9723
Post-course Generalize *Number of high school statistics courses taken	0.03930	0.6412
Post-course Generalize *Number of college math courses taken	-0.00964	0.9087
Post-course Generalize *Number of college statistics courses taken	0.15897	0.0570
Post-course Generalize * Cumulative attempted credits	-0.03281	0.6963
Post-course Generalize * Cumulative earned credits	-0.04244	0.6135
Post-course Generalize * Cumulative points	-0.00590	0.9440

Post-course Generalize * Cumulative GPA	-0.05899	0.4825
Post-course Generalize *Course grade	0.16879	0.0431
Post-course Generalize *Course Effort	0.11827	0.1580
Post-course Generalize *Reported GPA	0.07952	0.3486
Post-course Lurking variable*Gender	0.07546	0.3687
Post-course Lurking variable *Reported Math SAT score	-0.10017	0.3046
Post-course Lurking variable *Reported Verbal SAT score	0.03304	0.7355
Post-course Lurking variable *Reported Total SAT score	-0.03045	0.7445
Post-course Lurking variable *Number of high school math courses taken	-0.06312	0.4539
Post-course Lurking variable *Number of high school statistics courses taken	-0.00744	0.9297
Post-course Lurking variable *Number of college math courses taken	-0.06731	0.4228
Post-course Lurking variable *Number of college statistics courses taken	-0.05436	0.5175
Post-course Lurking variable * Cumulative attempted credits	0.00565	0.9464
Post-course Lurking variable * Cumulative earned credits	-0.00345	0.9673
Post-course Lurking variable * Cumulative points	-0.00367	0.9652
Post-course Lurking variable * Cumulative GPA	-0.05514	0.5116
Post-course Lurking variable *Course grade	-0.04399	0.6006
Post-course Lurking variable *Course Effort	0.00724	0.9314
Post-course Lurking variable *Reported GPA	0.00592	0.9445

Post-course Method*Gender	0.19160	0.0214
Post-course Method *Reported Math SAT score	0.08208	0.4006
Post-course Method *Reported Verbal SAT score	0.28497	0.0029
Post-course Method *Reported Total SAT score	0.23356	0.0113
Post-course Method *Number of high school math courses taken	-0.05062	0.5483
Post-course Method *Number of high school statistics courses taken	0.02168	0.7972
Post-course Method *Number of college math courses taken	-0.02836	0.7358
Post-course Method *Number of college statistics courses taken	-0.05867	0.4848
Post-course Method * Cumulative attempted credits	0.04788	0.5688
Post-course Method * Cumulative earned credits	0.05624	0.5031
Post-course Method * Cumulative points	0.09594	0.2527
Post-course Method * Cumulative GPR	0.23409	0.0047
Post-course Method *Course grade	0.21822	0.0086
Post-course Method *Course Effort	-0.13289	0.1123
Post-course Method *Reported GPA	0.21316	0.0112
Post-course Reported statistic*Gender	0.18866	0.0235
Post-course Reported statistic *Reported Math SAT score	0.03723	0.7034
Post-course Reported statistic *Reported Verbal SAT score	-0.01532	0.8756
Post-course Reported statistic *Reported Total SAT score	-0.14523	0.1182
Post-course Reported statistic *Number of high school math courses taken	0.09929	0.2381

Post-course Reported statistic *Number of high school statistics courses taken	-0.09929	0.2381
Post-course Reported statistic *Number of college math courses taken	0.02190	0.7944
Post-course Reported statistic *Number of college statistics courses taken	0.02417	0.7737
Post-course Reported statistic * Cumulative attempted credits	0.13069	0.1185
Post-course Reported statistic * Cumulative earned credits	0.11250	0.1794
Post-course Reported statistic * Cumulative points	0.16859	0.0434
Post-course Reported statistic * Cumulative GPA	0.09740	0.2455
Post-course Reported statistic *Course grade	0.13333	0.1111
Post-course Reported statistic *Course Effort	0.05302	0.5279
Post-course Reported statistic *Reported GPA	0.13893	0.1004
Post-course Variation*Gender	0.00437	0.9585
Post-course Variation *Reported Math SAT score	0.11397	0.2425
Post-course Variation *Reported Verbal SAT score	-0.00353	0.9712
Post-course Variation *Reported Total SAT score	-0.06035	0.5180
Post-course Variation *Number of high school math courses taken	-0.04436	0.5988
Post-course Variation *Number of high school statistics courses taken	-0.23254	0.0052
Post-course Variation *Number of college math courses taken	0.10462	0.2121
Post-course Variation *Number of college statistics courses taken	0.02142	0.7988

Post-course Variation * Cumulative attempted credits	0.11221	0.1806
Post-course Variation * Cumulative earned credits	0.08244	0.3259
Post-course Variation * Cumulative points	0.16120	0.0536
Post-course Variation * Cumulative GPA	0.11102	0.1853
Post-course Variation *Course grade	0.18242	0.0286
Post-course Variation *Course Effort	0.12784	0.1268
Post-course Variation *Reported GPA	0.14814	0.0796

Table 8: Correlations and p-values between post-course average statistical literacy components from Statistical Literacy Components Rubric and pre-course attitude components from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Post-course Bias*Pre-course affect	-0.07174	0.4031
Post-course Bias*Pre-course cognitive competence	0.00173	0.9836
Post-course Bias*Pre-course value	0.10504	0.2118
Post-course Bias*Pre-course difficulty	-0.06995	0.4065
Post-course Bias*Pre-course interest	-0.03427	0.6834
Post-course Bias*Pre-course effort	0.02537	0.7644
Post-course Causality*Pre-course affect	0.07221	0.4
Post-course Causality *Pre-course cognitive competence	0.02570	0.7598
Post-course Causality *Pre-course value	0.00666	0.9371
Post-course Causality *Pre-course difficulty	-0.13608	0.1051
Post-course Causality *Pre-course interest	0.11397	0.1738
Post-course Causality *Pre-course effort	0.06902	0.4144
Post-course Definition*Pre-course affect	-0.04278	0.6183
Post-course Definition *Pre-course cognitive competence	0.03149	0.7079
Post-course Definition *Pre-course value	-0.00699	0.9340
Post-course Definition *Pre-course difficulty	0.08672	0.3031
Post-course Definition *Pre-course interest	-0.09113	0.2774
Post-course Definition *Pre-course effort	-0.02906	0.7313
Post-course Generalize*Pre-course affect	-0.01866	0.8280

Post-course Generalize *Pre-course cognitive competence	0.07024	0.4028
Post-course Generalize *Pre-course value	0.07088	0.4002
Post-course Generalize *Pre-course difficulty	-0.01590	0.8505
Post-course Generalize *Pre-course interest	0.10729	0.2006
Post-course Generalize *Pre-course effort	0.17367	0.0387
Post-course Lurking variable*Pre-course affect	0.00580	0.9461
Post-course Lurking variable *Pre-course cognitive competence	-0.04547	0.5884
Post-course Lurking variable *Pre-course value	-0.09017	0.2842
Post-course Lurking variable *Pre-course difficulty	0.03726	0.6586
Post-course Lurking variable *Pre-course interest	-0.00146	0.9861
Post-course Lurking variable *Pre-course effort	-0.10614	0.2087
Post-course Method*Pre-course affect	-0.11475	0.1802
Post-course Method *Pre-course cognitive competence	-0.01185	0.8879
Post-course Method *Pre-course value	-0.04137	0.6237
Post-course Method *Pre-course difficulty	0.01709	0.8394
Post-course Method *Pre-course interest	-0.02344	0.7803
Post-course Method *Pre-course effort	0.03643	0.6669
Post-course Reported statistic*Pre-course affect	-0.05575	0.5161
Post-course Reported statistic *Pre-course cognitive competence	0.00809	0.9234
Post-course Reported statistic *Pre-course value	-0.00394	0.9627
Post-course Reported statistic	-0.02988	0.7231

*Pre-course difficulty		
Post-course Reported statistic *Pre-course interest	-0.01877	0.8233
Post-course Reported statistic *Pre-course effort	0.16609	0.0482
Post-course Variation*Pre-course affect	-0.03525	0.6815
Post-course Variation *Pre-course cognitive competence	0.10786	0.1981
Post-course Variation *Pre-course value	-0.02589	0.7589
Post-course Variation *Pre-course difficulty	0.04367	0.6045
Post-course Variation *Pre-course interest	0.02313	0.7832
Post-course Variation *Pre-course effort	-0.04773	0.5727

Table 9: Correlations and p-values between post-course average statistical literacy components from SLCR and post-course attitude components from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Post-course Bias*Post-course affect	0.00651	0.9384
Post-course Bias* Post-course cognitive competence	0.01553	0.8550
Post-course Bias* Post-course value	-0.05676	0.4992
Post-course Bias* Post-course difficulty	-0.02968	0.7259
Post-course Bias* Post-course interest	-0.04332	0.6088
Post-course Bias* Post-course effort	-0.04869	0.5650
Post-course Causality* Post-course affect	0.1001	0.2343
Post-course Causality * Post-course cognitive competence	0.07405	0.3829
Post-course Causality * Post-course value	0.11129	0.1842
Post-course Causality * Post-course difficulty	-0.05318	0.5297
Post-course Causality * Post-course interest	0.15317	0.0688
Post-course Causality * Post-course effort	0.02695	0.7502
Post-course Definition* Post-course affect	0.0075	0.9292
Post-course Definition * Post-course cognitive competence	0.07341	0.3870
Post-course Definition * Post-course value	0.03193	0.7040
Post-course Definition * Post-course difficulty	0.09062	0.2835
Post-course Definition * Post-course interest	-0.00934	0.9121
Post-course Definition * Post-course effort	-0.08787	0.2984
Post-course Generalize* Post-course affect	0.07712	0.3599
Post-course Generalize *	0.12373	0.1438

Post-course cognitive competence		
Post-course Generalize * Post-course value	0.17153	0.0398
Post-course Generalize * Post-course difficulty	0.01422	0.8666
Post-course Generalize * Post-course interest	0.127	0.1320
Post-course Generalize * Post-course effort	-0.00334	0.9685
Post-course Lurking variable* Post-course affect	0.04073	0.6291
Post-course Lurking variable * Post-course cognitive competence	-0.05901	0.4870
Post-course Lurking variable * Post-course value	0.01894	0.8217
Post-course Lurking variable * Post-course difficulty	-0.05007	0.5540
Post-course Lurking variable * Post-course interest	0.09013	0.2861
Post-course Lurking variable * Post-course effort	0.00776	0.9270
Post-course Method* Post-course affect	0.01253	0.8819
Post-course Method * Post-course cognitive competence	0.11879	0.1606
Post-course Method * Post-course value	0.03421	0.6840
Post-course Method *Pre-course difficulty	0.11899	0.1584
Post-course Method * Post-course interest	-0.00366	0.9655
Post-course Method * Post-course effort	-0.06482	0.4435
Post-course Reported statistic* Post-course affect	-0.05840	0.4884
Post-course Reported statistic * Post-course cognitive competence	-0.04251	0.6167
Post-course Reported statistic * Post-course value	0.06811	0.4173
Post-course Reported statistic	-0.06983	0.4089

* Post-course difficulty		
Post-course Reported statistic * Post-course interest	0.10623	0.2083
Post-course Reported statistic * Post-course effort	0.10626	0.2082
Post-course Variation* Post-course affect	0.03186	0.7056
Post-course Variation * Post-course cognitive competence	0.08529	0.3146
Post-course Variation * Post-course value	0.09318	0.2666
Post-course Variation * Post-course difficulty	0.07655	0.3652
Post-course Variation * Post-course interest	0.11645	0.1675
Post-course Variation * Post-course effort	0.07960	0.3464

Table 10: Correlations and p-values for pre-course and post-course attitude scores from SATS-36©

Variables	Pearson Correlation Coefficient	p-value
Pre-course affect*Pre-course cognitive competence	0.84855	<0.0001
Pre-course affect*Pre-course value	0.46400	<0.0001
Pre-course affect*Pre-course difficulty	0.63219	<0.0001
Pre-course affect*Pre-course interest	0.53703	<0.0001
Pre-course affect*Pre-course effort	-0.09038	0.2936
Pre-course affect*Post-course affect	0.56233	<0.0001
Pre-course affect*Post-course cognitive competence	0.53980	<0.0001
Pre-course affect*Post-course value	0.34079	<0.0001
Pre-course affect*Post-course difficulty	0.36163	<0.0001
Pre-course affect*Post-course interest	0.42570	<0.0001
Pre-course affect*Post-course effort	-0.14527	0.0903
Pre-course cognitive competence*Pre-course value	0.46095	<0.0001
Pre-course cognitive competence *Pre-course difficulty	0.64047	<0.0001
Pre-course cognitive competence *Pre-course interest	0.42159	<0.0001
Pre-course cognitive competence *Pre-course effort	-0.09015	0.2860
Pre-course cognitive competence *Post-course affect	0.58197	<0.0001
Pre-course cognitive competence *Post-course cognitive competence	0.66310	<0.0001
Pre-course cognitive competence *Post-course value	0.31169	0.0001
Pre-course cognitive	0.47608	<0.0001

competence *Post-course difficulty		
Pre-course cognitive competence *Post-course interest	0.34375	<0.0001
Pre-course cognitive competence *Post-course effort	-0.14098	0.0942
Pre-course value*Pre-course difficulty	0.35326	<0.0001
Pre-course value *Pre-course interest	0.49591	<0.0001
Pre-course value *Pre-course effort	0.08925	0.2926
Pre-course value *Post-course affect	0.32241	<0.0001
Pre-course value *Post-course cognitive competence	0.31763	0.0001
Pre-course value *Post-course value	0.60013	<0.0001
Pre-course value *Post-course difficulty	0.17574	0.0371
Pre-course value *Post-course interest	0.44044	<0.0001
Pre-course value *Post-course effort	-0.09949	0.2405
Pre-course difficulty*Pre-course interest	0.24554	0.0031
Pre-course difficulty*Pre-course effort	-0.19183	0.0227
Pre-course difficulty*Post-course affect	0.47789	<0.0001
Pre-course difficulty*Post-course cognitive competence	0.46513	<0.0001
Pre-course difficulty*Post-course value	0.28944	0.0005
Pre-course difficulty*Post-course difficulty	0.61736	<0.0001
Pre-course difficulty*Post-course interest	0.19250	0.0222
Pre-course difficulty*Post-course effort	-0.17125	0.0423
Pre-course interest*Pre-course effort	0.17275	0.0398

Pre-course interest*Post-course affect	0.29342	0.0004
Pre-course interest*Post-course cognitive competence	0.22860	0.0064
Pre-course interest*Post-course value	0.38376	<0.0001
Pre-course interest*Post-course difficulty	0.05335	0.5283
Pre-course interest*Post-course interest	0.59477	<0.0001
Pre-course interest*Post-course effort	0.02737	0.7464
Pre-course effort*Post-course affect	-0.14369	0.0880
Pre-course effort*Post-course cognitive competence	-0.16415	0.0535
Pre-course effort*Post-course value	-0.05671	0.5026
Pre-course effort*Post-course difficulty	-0.10829	0.2028
Pre-course effort*Post-course interest	0.06182	0.4681
Pre-course effort*Post-course effort	0.52703	<0.0001
Post-course affect*Post course cognitive competence	0.88334	<0.0001
Post-course affect*Post-course value	0.48430	<0.0001
Post-course affect*Post-course difficulty	0.67236	<0.0001
Post-course affect*Post-course interest	0.50240	<0.0001
Post-course affect*Post-course effort	-0.13699	0.1040
Post-course cognitive competence*Post-course value	0.45710	<0.0001
Post-course cognitive competence*Post-course difficulty	0.73789	<0.0001
Post-course cognitive competence*Post-course interest	0.34251	<0.0001
Post-course cognitive	-0.16010	0.0598

competence*Post-course effort		
Post-course value*Post-course difficulty	0.23580	0.0047
Post-course value*Post-course interest	0.65287	<0.0001
Post-course value*Post-course effort	-0.02924	0.7298
Post-course difficulty*Post-course interest	0.09787	0.2482
Post-course difficulty*Post-course effort	-0.14388	0.0899
Post-course interest*Post-course effort	0.06533	0.4432

Table 11: Correlations and p-values for demographic variables and difference (post-course - pre-course) in average statistical literacy component scores from SLCR

Variables	Pearson Correlation Coefficient	p-value
Bias difference * Gender	0.14882	0.075
Bias difference * Math SAT score	-0.02788	0.7756
Bias difference * Verbal SAT score	0.10880	0.2646
Bias difference * Total SAT score	0.00448	0.9618
Bias difference * Number of high school math courses taken	-0.00496	0.9531
Bias difference * Number of high school statistics courses taken	0.04934	0.5584
Bias difference * Number of college math courses taken	-0.04428	0.5982
Bias difference * Number of college statistics courses taken	-0.04428	0.5982
Bias difference * Cumulative attempted credits	-0.09943	0.2357
Bias difference * Cumulative earned credits	-0.05676	0.4992
Bias difference * Cumulative points	-0.08326	0.3211
Bias difference * Cumulative GPA	0.02123	0.8006
Bias difference * Numerical grade	0.11408	0.1734
Bias difference * Course Effort	-0.04004	0.6337
Bias difference * Reported GPR	0.09552	0.2598
Causality difference * Gender	0.04954	0.5554
Causality difference * Math SAT score	-0.07875	0.4201

Causality difference * Verbal SAT score	0.0055	0.9552
Causality difference * Total SAT score	-0.01119	0.9047
Causality difference * Number of high school math courses taken	-0.02068	0.8063
Causality difference * Number of high school statistics courses taken	0.15709	0.061
Causality difference * Number of college math courses taken	0.19003	0.0225
Causality difference * Number of college statistics courses taken	0.19003	0.0225
Causality difference * Cumulative attempted credits	0.12920	0.1227
Causality difference * Cumulative earned credits	0.21095	0.0111
Causality difference * Cumulative points	0.12847	0.1249
Causality difference * Cumulative GPR	-0.00463	0.9561
Causality difference * Numerical grade	0.0089	0.9157
Causality difference * Course Effort	-0.09793	0.2429
Causality difference * Reported GPA	0.05392	0.5254
Definition difference * Gender	0.25685	0.0019
Definition difference * Math SAT score	0.0877	0.3691
Definition difference * Verbal SAT score	-0.11428	0.2412
Definition difference * Total SAT score	0.08801	0.3454
Definition difference * Number of high school math courses taken	-0.04773	0.5713
Definition difference *	-0.00082	0.9922

Number of high school statistics courses taken		
Definition difference * Number of college math courses taken	-0.13509	0.1064
Definition difference * Number of college statistics courses taken	-0.13509	0.1064
Definition difference * Cumulative attempted credits	-0.08472	0.3127
Definition difference * Cumulative earned credits	-0.15723	0.0598
Definition difference * Cumulative points	-0.09406	0.2621
Definition difference * Cumulative GPA	0.03121	0.7104
Definition difference * Numerical grade	-0.06118	0.4663
Definition difference * Course Effort	0.08113	0.3337
Definition difference * Reported GPR	-0.04371	0.6068
Generalize difference * Gender	-0.01615	0.8476
Generalize difference * Math SAT score	-0.17144	0.0775
Generalize difference * Verbal SAT score	-0.03121	0.7496
Generalize difference * Total SAT score	-0.11429	0.2198
Generalize difference * Number of high school math courses taken	-0.09487	0.2597
Generalize difference * Number of high school statistics courses taken	0.09889	0.24
Generalize difference * Number of college math courses taken	0.0088	0.9167
Generalize difference * Number of college statistics courses taken	0.0088	0.9167

Generalize difference * Cumulative attempted credits	-0.07806	0.3524
Generalize difference * Cumulative earned credits	-0.1287	0.1242
Generalize difference * Cumulative points	-0.06395	0.4464
Generalize difference * Cumulative GPA	-0.13178	0.1154
Generalize difference * Numerical grade	0.06196	0.4607
Generalize difference * Course Effort	0.11941	0.154
Generalize difference * Reported GPA	0.00944	0.9115
Lurking variable difference * Gender	-0.0957	0.2538
Lurking variable difference * Math SAT score	0.07846	0.4218
Lurking variable difference * Verbal SAT score	-0.01571	0.8724
Lurking variable difference * Total SAT score	-0.03145	0.7364
Lurking variable difference * Number of high school math courses taken	-0.00979	0.9076
Lurking variable difference * Number of high school statistics courses taken	-0.03128	0.7107
Lurking variable difference * Number of college math courses taken	-0.03934	0.6397
Lurking variable difference * Number of college statistics courses taken	-0.03934	0.6397
Lurking variable difference * Cumulative attempted credits	0.08099	0.3345
Lurking variable difference * Cumulative earned credits	0.06460	0.4417
Lurking variable difference * Cumulative points	0.06762	0.4206
Lurking variable difference	-0.01705	0.8393

* Cumulative GPA		
Lurking variable difference * Numerical grade	-0.10519	0.2095
Lurking variable difference * Course Effort	-0.08894	0.2891
Lurking variable difference * Reported GPA	0.00675	0.9367
Method difference * Gender	0.17158	0.0398
Method difference * Math SAT score	0.04831	0.6212
Method difference * Verbal SAT score	0.21919	0.0233
Method difference * Total SAT score	0.18813	0.0422
Method difference * Number of high school math courses taken	-0.06115	0.4682
Method difference * Number of high school statistics courses taken	-0.08339	0.3221
Method difference * Number of college math courses taken	-0.09897	0.2379
Method difference * Number of college statistics courses taken	-0.09897	0.2379
Method difference * Cumulative attempted credits	0.05797	0.4901
Method difference * Cumulative points	0.07819	0.3516
Method difference * Cumulative earned credits	0.04139	0.6223
Method difference * Cumulative GPA	0.18906	0.0232
Method difference * Numerical grade	0.09602	0.2523
Method difference * Course Effort	-0.06527	0.4370
Method difference * Reported GPA	0.15614	0.0645
Reported statistic difference * Gender	0.13881	0.0971

Reported statistic difference * Math SAT score	0.01184	0.9037
Reported statistic difference * Verbal SAT score	0.00865	0.9295
Reported statistic difference * Total SAT score	-0.08107	0.3849
Reported statistic difference * Number of high school math courses taken	0.01198	0.8871
Reported statistic difference * Number of high school statistics courses taken	-0.02897	0.7313
Reported statistic difference * Number of college math courses taken	0.09644	0.2502
Reported statistic difference * Number of college statistics courses taken	0.09644	0.2502
Reported statistic difference * Cumulative attempted credits	0.12322	0.1412
Reported statistic difference * Cumulative points	0.14247	0.0885
Reported statistic difference * Cumulative earned credits	0.12603	0.1323
Reported statistic difference * Cumulative GPA	0.08788	0.2949
Reported statistic difference * Numerical grade	0.16434	0.049
Reported statistic difference * Course Effort	0.08488	0.3118
Reported statistic difference * Reported GPA	0.15531	0.0659
Variation difference * Gender	0.00702	0.9335
Variation difference * Math SAT score	0.05408	0.5801
Variation difference * Verbal SAT score	-0.0009	0.9927
Variation difference * Total SAT score	-0.075	0.4216
Variation difference * Number of high school	-0.11618	0.1670

math courses taken		
Variation difference * Number of high school statistics courses taken	-0.27794	0.0008
Variation difference * Number of college math courses taken	-0.11941	0.154
Variation difference * Number of college statistics courses taken	-0.11041	0.1540
Variation difference * Cumulative attempted credits	0.14066	0.0926
Variation difference * Cumulative points	0.19160	0.0214
Variation difference * Cumulative earned credits	0.05616	0.5037
Variation difference * Cumulative GPA	0.13127	0.1168
Variation difference * Numerical grade	0.17861	0.0322
Variation difference * Course Effort	0.17792	0.0329
Variation difference * Reported GPA	0.15745	0.0622

Table 12: Correlations and p-values between pre-course attitude component scores and the difference (post-course – pre-course) in average statistical literacy component scores from SLCR

Variables	Pearson Correlation Coefficient	p-value
Bias difference * Pre-course Affect	-0.09493	0.2681
Bias difference * Pre-course Cognitive competence	-0.00827	0.9217
Bias difference * Pre-course Value	0.08363	0.3207
Bias difference* Pre-course Difficulty	-0.10524	0.2109
Bias difference * Pre-course Interest	-0.02473	0.7686
Bias difference * Pre-course Effort	0.11602	0.1691
Causality difference * Pre-course Affect	0.03163	0.7127
Causality difference * Pre-course Cognitive competence	-0.00716	0.9321
Causality difference * Pre-course Value	0.06047	0.4731
Causality difference * Pre-course Difficulty	-0.09507	0.2587
Causality difference * Pre-course Interest	0.13925	0.096
Causality difference * Pre-course Effort	0.12284	0.1453
Definition difference * Pre-course Affect	-0.05283	0.5383
Definition difference * Pre-course Cognitive competence	-0.03845	0.6473
Definition difference * Pre-course Value	-0.13103	0.1188
Definition difference * Pre-course Difficulty	-0.04776	0.5711
Definition difference * Pre-course Interest	-0.12302	0.1418
Definition difference * Pre-course Effort	-0.00352	0.9669

Generalize difference * Pre-course Affect	-0.15928	0.062
Generalize difference * Pre-course Cognitive competence	-0.09602	0.2523
Generalize difference * Pre-course Value	0.01528	0.8563
Generalize difference * Pre-course Difficulty	-0.10781	0.2
Generalize difference * Pre-course Interest	0.04212	0.6162
Generalize difference * Pre-course Effort	0.20634	0.0138
Lurking variable difference * Pre-course Affect	-0.0352	0.6819
Lurking variable difference * Pre-course Cognitive competence	-0.04489	0.5931
Lurking variable difference * Pre-course Value	-0.10132	0.2286
Lurking variable difference * Pre-course Difficulty	-0.0512	0.5437
Lurking variable difference * Pre-course Interest	-0.01519	0.8566
Lurking variable difference * Pre-course Effort	-0.13653	0.1052
Method difference * Pre-course Affect	-0.0865	0.3131
Method difference * Pre-course Cognitive competence	0.02564	0.7603
Method difference * Pre-course Value	-0.06364	0.4502
Method difference * Pre-course Difficulty	0.09021	0.2839
Method difference * Pre-course Interest	0.01101	0.8958
Method difference * Pre-course Effort	0.02715	0.7484
Reported statistic difference * Pre-course Affect	-0.01098	0.8983
Reported statistic difference * Pre-course Cognitive	0.03787	0.6522

competence		
Reported statistic difference * Pre-course Value	-0.06965	0.4085
Reported statistic difference * Pre-course Difficulty	0.04354	0.6056
Reported statistic difference * Pre-course Interest	-0.07114	0.3968
Reported statistic difference * Pre-course Effort	0.02456	0.7717
Variation difference * Pre-course Affect	-0.0907	0.2901
Variation difference * Pre-course Cognitive competence	0.04794	0.5683
Variation difference * Pre-course Value	-0.12791	0.1279
Variation difference * Pre-course Difficulty	0.02249	0.7898
Variation difference * Pre-course Interest	-0.03723	0.6577
Variation difference * Pre-course Effort	0.01635	0.8469

Table 13: Correlations and p-values between post-course attitude component scores and the difference (post-course – pre-course) in average statistical literacy component scores from SLCR

Variables	Pearson Correlation Coefficient	p-value
Bias difference * Post-course Affect	0.00312	0.9705
Bias difference * Post-course Cognitive competence	0.01888	0.8242
Bias difference * Post-course Value	0.03372	0.6882
Bias difference * Post-course Difficulty	-0.0832	0.3249
Bias difference * Post-course Interest	-0.03808	0.6528
Bias difference * Post-course Effort	0.02006	0.8127
Causality difference * Post-course Affect	-0.00775	0.9268
Causality difference * Post-course Cognitive competence	0.0062	0.9418
Causality difference * Post-course Value	0.05725	0.4955
Causality difference * Post-course Difficulty	-0.01745	0.8367
Causality difference * Post-course Interest	0.1062	0.2084
Causality difference * Post-course Effort	-0.0134	0.8742
Definition difference * Post-course Affect	-0.05313	0.5285
Definition difference * Post-course Cognitive competence	-0.03325	0.6955
Definition difference * Post-course Value	-0.14457	0.0838
Definition difference * Post-course Difficulty	0.00092	0.9913
Definition difference * Post-course Interest	-0.12022	0.1541
Definition difference *	-0.03723	0.66

Post-course Effort		
Generalize difference * Post-course Affect	-0.05174	0.5394
Generalize difference * Post-course Cognitive competence	-0.0491	0.5631
Generalize difference * Post-course Value	0.11432	0.1725
Generalize difference * Post-course Difficulty	-0.12881	0.1266
Generalize difference * Post-course Interest	0.12708	0.1318
Generalize difference * Post-course Effort	0.09837	0.2441
Lurking variable difference * Post-course Affect	-0.02934	0.7280
Lurking variable difference * Post-course Cognitive competence	-0.05456	0.5205
Lurking variable difference * Post-course Value	-0.02512	0.7650
Lurking variable difference * Post-course Difficulty	-0.07959	0.3464
Lurking variable difference * Post-course Interest	0.00337	0.9682
Lurking variable difference * Post-course Effort	-0.03718	0.6604
Method difference * Post-course Affect	0.0079	0.9254
Method difference * Post-course Cognitive competence	0.09696	0.2527
Method difference * Post-course Value	0.01988	0.8131
Method difference * Post-course Difficulty	0.09349	0.2684
Method difference * Post-course Interest	0.03213	0.7042
Method difference * Post-course Effort	-0.00119	0.9888
Reported statistic difference * Post-course Affect	-0.00821	0.9224
Reported statistic difference	0.03323	0.6956

* Post-course Cognitive competence		
Reported statistic difference * Post-course Value	0.13885	0.0970
Reported statistic difference * Post-course Difficulty	0.02759	0.7445
Reported statistic difference * Post-course Interest	0.03563	0.6737
Reported statistic difference * Post-course Effort	0.02332	0.7829
Variation difference * Post-course Affect	-0.00233	0.9780
Variation difference * Post-course Cognitive competence	0.04823	0.5701
Variation difference * Post-course Value	0.06790	0.4187
Variation difference * Post-course Difficulty	0.0755	0.3715
Variation difference * Post-course Interest	0.04246	0.6159
Variation difference * Post-course Effort	0.15869	0.0593

Appendix E

Tables for Results for Research Question 2

Table 1: Results from McNemar’s tests for topic categories for the Allstate advertisement

Topic Category	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Accuracy or Reliability of 6 million?	116	1	18	12.50	9	6.25	0.1221
Accurate to count if not causes injury or damages?	138	0	1	0.69	5	3.47	0.2188
Accurate to compare to other countries?	141	0	0	0.00	3	2.08	0.25
Accurate to count if not driver's fault?	143	0	1	0.69	0	0.00	1
Agenda for Allstate?	112	2	17	11.81	13	9.03	0.5847
Better place definition	126	2	5	3.47	11	7.64	0.2101
Biased?	127	1	11	7.64	5	3.47	0.2101
Cause of accident?	129	0	3	2.08	12	8.33	0.0352
Correlation?	143	1	0	0.00	0	0.00	.
Coverage error?	142	0	2	1.39	0	0.00	0.5
Dark figure? (unreported cases)	136	0	7	4.86	1	0.69	0.0703
Definition of America?	127	5	9	6.25	3	2.08	0.146
Definition of car?	130	0	12	8.33	2	1.39	0.0129
Definition or type of accident?	14	76	46	31.94	8	5.56	<0.0001
Effect of definition of America on 6 million	135	0	6	4.17	3	2.08	0.5078
Effect of definition of car accident on 6 million	90	13	31	21.53	10	6.94	0.0015
Effect of definition of car on 6 million	140	0	3	2.08	1	0.69	0.625
Effect of having insurance on driving	142	0	0	0.00	2	1.39	0.5
Effect of unreported accidents on 6 million	125	0	12	8.33	7	4.86	0.3593
Exact number?	120	2	16	11.11	6	4.17	0.0525
Factors affecting 6 million?	142	0	1	0.69	1	0.69	1
Factors affecting	107	7	11	7.64	19	13.19	0.2005

accidents?							
Generalizable?	143	0	1	0.69	0	0.00	1
Have Allstate insurance?	124	1	4	2.78	15	10.42	0.0192
Have insurance and reduce accidents?	126	5	6	4.17	7	4.86	1
How was information obtained?	48	20	67	46.53	9	6.25	<0.0001
Includes insured and uninsured drivers?	139	0	4	2.78	1	0.69	0.375
Location of accidents?	115	4	18	12.50	7	4.86	0.0433
Lurking variable?	138	0	6	4.17	0	0.00	0.0313
Measurement error?	139	0	4	2.78	1	0.69	0.375
Misconception	139	0	4	2.78	1	0.69	0.375
MOE or CI included?	129	0	15	10.42	0	0.00	0.0000
Number of accidents varies	125	2	10	6.94	7	4.86	0.6291
Number of cars or number of accidents?	131	1	5	3.47	7	4.86	0.7744
Number of cars, drivers, or accidents?	128	4	7	4.86	5	3.47	0.7744
Number of reported accidents?	143	0	0	0.00	1	0.69	1
Other ways to report?	137	0	2	1.39	5	3.47	0.4531
Outliers?	142	0	2	1.39	0	0.00	0.5
Parameter	143	0	1	0.69	0	0.00	1
Parameter or statistic?	142	0	2	1.39	0	0.00	0.5
Percent of cars or drivers on road	142	0	1	0.69	1	0.69	1
Percent of drivers not in accident	83	0	0	0.00	1	0.69	1
Picture in background	143	0	0	0.00	1	0.69	1
Population or population size?	136	0	7	4.86	1	0.69	0.0703
Probability of being in accident?	143	0	1	0.69	0	0.00	1
Question if could count all accidents	136	0	7	4.86	1	0.69	0.0703
Question if number of accidents really is 6 million	127	1	9	6.25	7	4.86	0.8036
Question wording	140	0	4	2.78	0	0.00	0.125
Reducing number of accidents	143	0	0	0.00	1	0.69	1

Relativeness of 6 million	123	5	9	6.25	7	4.86	0.8036
Relevance of 6 million	141	0	3	2.08	0	0.00	0.25
Representative?	139	0	2	1.39	3	2.08	1
Round figure	117	4	16	11.11	7	4.86	0.0931
Sample size?	136	0	6	4.17	2	1.39	0.2891
Six million is an average	124	5	8	5.56	7	4.86	1
Six million is an estimate	127	1	11	7.64	5	3.47	0.2101
Six million is correct	141	0	0	0.00	3	2.08	0.25
Source of 6 million?	71	26	31	21.53	16	11.11	0.04
Too many definition	135	0	8	5.56	1	0.69	0.0391
True for all insurance companies?	143	0	0	0.00	1	0.69	1
Unreported accidents included?	90	13	31	21.53	10	6.94	0.0015
Untrustworthy	134	0	7	4.86	3	2.08	0.3438
Vague claim	143	0	1	0.69	0	0.00	1
Validity	129	0	9	6.25	6	4.17	0.6072
Which years was study conducted?	113	2	20	13.89	9	6.25	0.0614
World vs. America	0	0	0	0.00	0	0.00	.

* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-values for comparing p_{01} and p_{10}

Table 2: Pre-course percentages for topic categories by sex for the Allstate advertisement

Topic Category	Percent female	Percent male	p-value for difference*
Accuracy or Reliability of 6 million?	7.14	6.67	0.5408
Accurate to count if not causes injury or damages?	3.57	3.33	0.6935 F
Accurate to compare to other countries?	2.38	1.67	0.0702 F
Accurate to count if not driver's fault?	0	0	1 F
Agenda for Allstate?	8.33	13.33	1 F
Better place definition	14.29	1.67	1 F
Biased?	4.76	3.33	1F
Cause of accident?	11.9	3.33	0.3329
Correlation?	1.19	0	0.0092
Coverage error?	0	0	1 F
Dark figure? (unreported cases)	0	1.67	0.0665
Definition of America?	4.76	6.67	1 F
Definition of car?	1.19	1.67	1F
Definition or type of accident?	58.33	58.33	0.4167 F
Effect of definition of America on 6 million	1.19	3.33	0.7192 F
Effect of definition of car accident on 6 million	15.48	16.67	1 F
Effect of definition of car on 6 million	0	1.67	1
Effect of having insurance on driving	2.38	0	0.5708 F
Effect of unreported accidents on 6 million	2.38	8.33	0.8475
Exact number?	7.14	3.33	0.4167 F
Factors affecting 6 million?	1.19	0	0.5105 F
Factors affecting accidents?	21.43	13.33	0.1282 F
Generalizable?	0	0	0.4691 F
Have Allstate insurance?	13.1	8.33	1 F
Have insurance and reduce accidents?	7.14	10	0.2131
How was information obtained?	17.86	23.33	0.4192
Includes insured and uninsured	0	1.67	0.37

drivers?			
Location of accidents?	9.52	5	0.5408
Lurking variable?	0	0	0.4192
Measurement error?	1.19	0	0.4167 F
Misconception	1.19	0	0.3610 F
MOE or CI included?	0	0	.
Number of accidents varies	5.95	6.67	1 F
Number of cars or number of accidents?	5.95	5	1 F
Number of cars, drivers, or accidents?	7.14	5	0.7352F
Number of reported accidents?	0	1.67	1 F
Other ways to report?	4.76	1.67	1 F
Outliers?	0	0	0.7352 F
Parameter	0	0	0.4167 F
Parameter or statistic?	0	0	0.4014 F
Percent of cars or drivers on road	0	1.67	0.4167F
Percent of drivers not in accident	1.19	0	1F
Picture in background?	0	1.67	0.4167F
Population or population size?	0	1.67	0.4167 F
Probability of being in accident?	0	0	.
Question if could count all accidents	0	1.67	0.4167 F
Question if number of accidents really is 6 million	5.95	5	0.4167 F
Question wording	0	0	.
Reducing number of accidents	0	1.67	0.4167 F
Relativeness of 6 million	9.52	6.67	1 F
Relevance of 6 million	0	0	.
Representative?	3.57	0	0.4167
Round figure	5.95	10	0.5408
Sample size?	2.38	0	0.5105F
Six million is an average	9.52	6.67	0.2657 F
Six million is an estimate	3.57	5	0.5263 F
Six million is correct	0	5	0.5105 F
Source of 6 million?	29.76	28.33	0.8525
Too many definition	0	1.67	0.4167 F
True for all insurance companies?	1.19	0	1 F

Unreported accidents included?	16.67	15	0.7878
Untrustworthy	1.19	3.33	0.5708 F
Vague claim	0	0	.
Validity	2.38	6.67	0.2347 F
Which years was study conducted?	10.71	3.33	0.1217 F
World vs. America	.	.	.

*p-values from chi-square tests were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

Table 3: Post-course percentages for topic categories by sex for the Allstate advertisement

Topic Category	Percent female	Percent male	p-value for difference*
Accuracy or Reliability of 6 million?	17.86	6.67	0.0504
Accurate to count if not causes injury or damages?	0	1.67	0.4167 F
Accurate to compare to other countries?	0	0	.
Accurate to count if not driver's fault?	0	1.67	0.4167 F
Agenda for Allstate?	16.67	8.33	0.1452
Better place definition	5.95	3.33	0.6994 F
Biased?	5.95	11.67	0.2213
Cause of accident?	1.19	3.33	0.5708 F
Correlation?	1.19	0	1 F
Coverage error?	1.19	1.67	1 F
Dark figure? (unreported cases)	7.14	1.67	0.2392 F
Definition of America?	4.76	16.67	0.0174
Definition of car?	8.33	8.33	1
Definition or type of accident?	85.71	83.33	0.6954
Effect of definition of America on 6 million	4.76	3.33	1 F
Effect of definition of car accident on 6 million	33.33	26.67	0.3919
Effect of definition of car on 6 million	1.19	3.33	0.5708 F
Effect of having insurance on driving	0	0	.
Effect of unreported accidents on 6 million	9.52	6.67	0.5408
Exact number?	13.1	11.67	0.7983
Factors affecting 6 million?	0	1.67	0.4167 F
Factors affecting accidents?	13.1	11.67	0.7983
Generalizable?	0	1.67	0.4167 F
Have Allstate insurance?	4.76	1.67	0.4014 F
Have insurance and reduce accidents?	7.14	8.33	1 F
How was information obtained?	63.1	56.67	0.4367

Includes insured and uninsured drivers?	2.38	3.33	1 F
Location of accidents?	14.29	16.67	0.6954
Lurking variable?	2.38	6.67	0.2347 F
Measurement error?	1.19	5	0.3080 F
Misconception	2.38	3.33	1 F
MOE or CI included?	11.9	8.33	0.4891
Number of accidents varies	5.95	11.67	0.2213
Number of cars or number of accidents?	4.76	3.33	1 F
Number of cars, drivers, or accidents?	8.33	6.67	0.7625 F
Number of reported accidents?	0	0	.
Other ways to report?	0	3.33	0.1719 F
Outliers?	1.19	1.67	1 F
Parameter	1.19	0	1 F
Parameter or statistic?	2.38	0	0.5105 F
Percent of cars or drivers on road	1.19	0	1 F
Percent of drivers not in accident	0	0	.
Picture in background	0	0	.
Population or population size?	7.14	1.67	0.2392 F
Probability of being in accident?	0	1.67	0.4167 F
Question if could count all accidents	8.33	0	0.0415 F
Question if number of accidents really is 6 million	5.95	8.33	0.7417 F
Question wording	2.38	3.33	1 F
Reducing number of accidents	0	0	.
Relativeness of 6 million	10.71	8.33	0.6345
Relevance of 6 million	2.38	1.67	1 F
Representative?	2.38	0	0.5105 F
Round figure	14.29	13.33	0.8706
Sample size?	5.95	1.67	0.4009 F
Six million is an average	10.71	6.67	0.4034
Six million is an estimate	10.71	5	0.2213
Six million is correct	0	0	.
Source of 6 million?	44.05	33.33	0.1949

Too many definition	7.14	3.33	0.4691 F
True for all insurance companies?	0	0	.
Unreported accidents included?	36.9	21.67	0.0503
Untrustworthy	5.95	3.33	0.6994 F
Vague claim	1.19	0	1 F
Validity	7.14	5	0.7352 F
Which years was study conducted?	17.86	11.67	0.3087
World vs America	2.38	5	0.6494 F

*p-values from chi-square tests were reported unless an “F” appeared indicating that the p-value resulted from Fisher’s exact test

Table 4: Results from McNemar's tests by sex for topic categories for the Allstate advertisement

Topic Category	Sex	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Accuracy or Reliability of 6 million?	Female	64	1	14	16.67	5	5.95	0.0636
	Male	52	0	4	6.67	4	6.67	1
Accurate to count if not causes injury or damages?	Female	81	0	0	0.00	3	3.57	0.25
	Male	57	0	1	1.67	2	3.33	1
Accurate to compare to other countries?	Female	82	0	0	0.00	2	2.38	0.5
	Male	59	0	0	0.00	1	1.67	1
Accurate to count if not driver's fault?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	1	1.67	0	0.00	1
Agenda for Allstate?	Female	64	1	13	15.48	6	7.14	0.1671
	Male	48	1	4	6.67	7	11.67	0.5488
Better place definition	Female	69	2	3	3.57	10	11.90	0.0923
	Male	57	0	2	3.33	1	1.67	1
Biased?	Female	76	1	4	4.76	3	3.57	1
	Male	51	0	7	11.67	2	3.33	0.1797
Cause of accident?	Female	73	0	1	1.19	10	11.90	0.0117
	Male	56	0	2	3.33	2	3.33	1
Correlation?	Female	83	1	0	0.00	0	0.00	.
	Male	60	0	0	0.00	0	0.00	.
Coverage error?	Female	83	0	1	1.19	0	0.00	1
	Male	59	0	1	1.67	0	0.00	1
Dark figure? (unreported cases)	Female	78	0	6	7.14	0	0.00	0.0313
	Male	58	0	1	1.67	1	1.67	1
Definition of America?	Female	77	1	3	3.57	3	3.57	1
	Male	50	4	6	10.00	0	0.00	0.0313
Definition of car?	Female	76	0	7	8.33	1	1.19	0.0703
	Male	54	0	5	8.33	1	1.67	0.2188
Definition or type of accident?	Female	7	44	28	33.33	5	5.95	<0.0001
	Male	7	32	18	30.00	3	5.00	<0.0001
Effect of definition of America on 6 million	Female	79	0	4	4.76	1	1.19	0.375

	Male	56	0	2	3.33	2	3.33	1
Effect of definition of car accident on 6 million	Female	51	8	20	23.81	5	5.95	0.0041
	Male	39	5	11	18.33	5	8.33	0.2101
Effect of definition of car on 6 million	Female	83	0	1	1.19	0	0.00	1
	Male	57	0	2	3.33	1	1.67	1
Effect of having insurance on driving	Female	82	0	0	0.00	2	2.38	0.5
	Male	60	0	0	0.00	0	0.00	.
Effect of unreported accidents on 6 million	Female	74	0	8	9.52	2	2.38	0.1094
	Male	51	0	4	6.67	5	8.33	1
Exact number?	Female	69	2	9	10.71	4	4.76	0.2668
	Male	51	0	7	11.67	2	3.33	0.1797
Factors affecting 6 million?	Female	83	0	0	0.00	1	1.19	1
	Male	59	0	1	1.67	0	0.00	1
Factors affecting accidents?	Female	60	5	6	7.14	13	15.48	0.1671
	Male	47	2	5	8.33	6	10.00	1
Generalizable?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	1	1.67	0	0.00	1
Have Allstate insurance?	Female	70	1	3	3.57	10	11.90	0.0192
	Male	70	1	3	5.00	10	16.67	0.0923
Have insurance and reduce accidents?	Female	74	2	4	4.76	4	4.76	1
	Male	52	3	2	3.33	3	5.00	1
How was information obtained?	Female	28	12	41	48.81	3	3.57	<0.0001
	Male	20	8	26	43.33	6	10.00	<0.0001
Includes insured and uninsured drivers?	Female	82	0	2	2.38	0	0.00	0.5
	Male	57	0	2	3.33	1	1.67	1
Location of accidents?	Female	66	2	10	11.90	6	7.14	0.4545
	Male	49	2	8	13.33	1	1.67	0.0391
Lurking variable?	Female	82	0	2	2.38	0	0.00	0.5
	Male	56	0	4	6.67	0	0.00	0.125
Measurement error?	Female	82	0	1	1.19	1	1.19	1
	Male	57	0	3	5.00	0	0.00	0.25
Misconception	Female	81	0	2	2.38	1	1.19	1

	Male	58	0	2	3.33	0	0.00	0.5
MOE or CI included?	Female	74	0	10	11.90	0	0.00	0.002
	Male	55	0	5	8.33	0	0.00	<0.0001
Number of accidents varies	Female	74	0	5	5.95	5	5.95	1
	Male	51	2	5	8.33	2	3.33	0.4531
Number of cars or number of accidents?	Female	76	1	3	3.57	4	4.76	1
	Male	55	0	2	3.33	3	5.00	1
Number of cars, drivers or accidents?	Female	74	3	4	4.76	3	3.57	1
	Male	54	1	3	5.00	2	3.33	1
Number of reported accidents?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	0	0.00	1	1.67	1
Other ways to report?	Female	80	0	0	0.00	4	4.76	0.125
	Male	57	0	2	3.33	1	1.67	1
Outliers?	Female	83	0	1	1.19	0	0.00	1
	Male	59	0	1	1.67	0	0.00	1
Parameter	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Parameter or statistic?	Female	82	0	2	2.38	0	0.00	0.5
	Male	60	0	0	0.00	0	0.00	.
Percent of cars or drivers on road	Female	83	0	1	1.19	0	0.00	1
	Male	59	0	0	0.00	1	1.67	1
Percent of drivers not in accident	Female	83	0	0	0.00	1	1.19	1
	Male	60	0	0	0.00	0	0.00	.
Picture in background?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	0	0.00	1	1.67	1
Population or population size?	Female	78	0	6	7.14	0	0.00	0.0313
	Male	58	0	1	1.67	1	1.67	1
Probability of being in accident?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	1	1.67	0	0.00	1
Question if could count all accidents?	Female	77	0	7	8.33	0	0.00	0.0156
	Male	59	0	0	0.00	1	1.67	1
Question if number of accidents really is 6 million	Female	75	1	4	4.76	4	4.76	1

	Male	52	0	5	8.33	3	5.00	0.7266
Question wording	Female	82	0	2	2.38	0	0.00	0.5
	Male	58	0	2	3.33	0	0.00	0.5
Reducing number of accidents	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	0	0.00	1	1.67	1
Relativeness of 6 million	Female	70	3	6	7.14	5	5.95	1
	Male	53	2	3	5.00	2	3.33	1
Relevance of 6 million	Female	82	0	2	2.38	0	0.00	0.5
	Male	59	0	1	1.67	0	0.00	1
Representative?	Female	79	0	2	2.38	3	3.57	1
	Male	60	0	0	0.00	0	0.00	.
Round figure	Female	69	2	10	11.90	3	3.57	0.0923
	Male	48	2	6	10.00	4	6.67	0.7539
Sample size?	Female	77	0	5	5.95	2	2.38	0.4531
	Male	59	0	1	1.67	0	0.00	1
Six million is an average	Female	71	4	5	5.95	4	4.76	1
	Male	53	1	3	5.00	3	5.00	1
Six million is an estimate	Female	73	1	8	9.52	2	2.38	0.1094
	Male	54	0	3	5.00	3	5.00	1
Six million is correct	Female	84	0	0	0.00	0	0.00	.
	Male	57	0	0	0.00	3	5.00	0.25
Source of 6 million?	Female	40	18	19	22.62	7	8.33	0.029
	Male	31	8	12	20.00	9	15.00	0.6636
Too many definition	Female	78	0	6	7.14	0	0.00	0.0313
	Male	57	0	2	3.33	1	1.67	1
True for all insurance companies?	Female	83	0	0	0.00	1	1.19	1
	Male	60	0	0	0.00	0	0.00	.
Unreported accidents included?	Female	50	11	20	23.81	3	3.57	<0.0001
	Male	40	2	11	18.33	7	11.67	0.4807
Untrustworthy	Female	78	0	5	5.95	1	1.19	0.2188
	Male	56	0	2	3.33	2	3.33	1
Vague claim	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Validity	Female	76	0	6	7.14	2	2.38	0.2891
	Male	53	0	3	5.00	4	6.67	1

Which years was study conducted?	Female	62	2	13	15.48	7	8.33	0.2632
	Male	51	0	7	11.67	2	3.33	0.1797

* n_{00} represents the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represents the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represents the number of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
n_{10} represents the number of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represents the percentage of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represents the percentage of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

*** Exact p-values for comparing p_{01} and p_{10}

Table 5: Results from McNemar's tests for topic categories for the LiveActive advertisement

Topic Category	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Accuracy and Reliability?	108	4	15	10.42	17	11.81	0.8601
Agenda?	116	2	15	10.42	11	7.64	0.5572
Amount of cottage cheese to eat?	143	0	1	0.69	0	0.00	1
Be that other woman definition	143	0	1	0.69	0	0.00	1
Bias	116	2	18	12.50	8	5.56	0.0755
Causality	120	3	8	5.56	13	9.03	0.3833
Effect of definition of occasional irregularity on 4 of 5	113	5	23	15.97	3	2.08	<0.0001
Effect of definition of occasional on 4 of 5	130	2	8	5.56	4	2.78	0.3877
Effect of factors on 4 of 5	104	13	14	9.72	13	9.03	1
Effect of how information obtained on 4 of 5	100	7	32	22.22	5	3.47	<0.0001
Effect of location on irregularity	133	2	6	4.17	3	2.08	0.5078
Effect of lurking variable on 4 of 5	140	0	4	2.78	0	0.00	0.125
Factors affecting irregularity	56	36	23	15.97	29	20.14	0.4885
Generalize?	125	2	8	5.56	9	6.25	1
How did they get information?	20	50	63	43.75	11	7.64	<0.0001
Length of study?	143	0	1	0.69	0	0.00	1
Location?	111	6	17	11.81	10	6.94	0.2478
Lurking variable	123	0	20	13.89	1	0.69	<0.0001
Measurement error	126	2	13	9.03	3	2.08	0.0213
Men not included	141	0	2	1.39	1	0.69	1
Misconception	123	2	12	8.33	7	4.86	0.3593
Misrepresenting	143	0	0	0.00	1	0.69	1
Misunderstand ad	122	3	11	7.64	8	5.56	0.6476
MOE/CI included?	127	1	16	11.11	0	0.00	<0.0001
Non-response	133	0	11	7.64	0	0.00	<0.0001
Other woman question	137	0	4	2.78	3	2.08	1
Population	132	1	8	5.56	3	2.08	0.2266
Question premise of ad	88	15	17	11.81	24	16.67	0.3489

Question wording	120	0	23	15.97	1	0.69	<0.0001
Relativeness of 4 of 5	142	0	0	0.00	2	1.39	0.5
Relevance of 4 of 5	143	0	1	0.69	0	0.00	1
Representative?	113	4	20	13.89	7	4.86	0.0192
Round number	141	0	3	2.08	0	0.00	0.25
Sample size?	61	28	34	23.61	21	14.58	0.1048
Sensitive topic	134	0	10	6.94	0	0.00	0.002
Source of 4 of 5	93	11	16	11.11	24	16.67	0.2682
Spoon size	142	0	2	1.39	0	0.00	0.5
Statistic and parameter	140	0	4	2.78	0	0.00	0.125
Untrustworthy	141	0	2	1.39	1	0.69	1
Vague claim	125	0	13	9.03	6	4.17	0.1671
Validity	120	0	10	6.94	14	9.72	0.5413
Variation	141	1	1	0.69	1	0.69	1
When study conducted?	139	0	4	2.78	1	0.69	0.375
Which yogurt for comparison?	135	1	7	4.86	1	0.69	0.0703
Why study conducted?	141	0	2	1.39	1	0.69	1

* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-values for comparing p_{01} and p_{10}

Table 6: Pre-course percentages for topic categories by sex for the LiveActive advertisement

Topic Category	Percent female	Percent male	p-value for differences*
Accuracy and Reliability?	13.1	16.67	0.5494
Agenda?	5.95	13.33	0.1276
Amount of cottage cheese to eat?	0	0	.
Be that other woman definition	0	0	.
Bias	7.14	6.67	1 F
Causality	5.95	18.33	0.0198
Effect of definition of occasional irregularity on 4 of 5	4.76	6.67	0.7192 F
Effect of definition of occasional on 4 of 5	2.38	6.67	0.2347
Effect of factors on 4 of 5	23.81	10	0.0337
Effect of how information obtained on 4 of 5	7.14	10	0.5408
Effect of location on irregularity	2.38	5	0.6494 F
Effect of lurking variable on 4 of 5	0	0	.
Factors affecting irregularity	55.95	30	0.002
Generalize?	10.71	3.33	0.1217
How did they get information?	45.24	38.33	0.4084
Length of study?	0	0	.
Location?	10.71	11.67	0.8577
Lurking variable	1.19	0	1 F
Measurement error	3.57	3.33	1 F
Men not included	0	1.67	0.4167 F
Misconception	7.14	5	0.7352 F
Misrepresenting	1.19	0	1 F
Misunderstand ad	8.33	6.67	0.7625 F
MOE/CI included?	0	1.67	0.4167 F
Non-response	0	0	.
Other woman question	2.38	1.67	1 F
Population	3.57	1.67	0.6408 F
Question premise of ad	25	30	0.5056
Question wording	0	1.67	0.4167 F
Relativeness of 4 of 5	1.19	1.67	1 F
Relevance of 4 of 5	0	0	.
Representative?	5.95	10	0.5263 F
Round number	0	0	.
Sample size?	33.33	35	0.8351
Sensitive topic	0	0	.
Source of 4 of 5	23.81	25	0.8696
Spoon size	0	0	.

Statistic and parameter	0	0	.
Untrustworthy	1.19	0	1 F
Vague claim	3.57	5	0.6935 F
Validity	5.95	15	0.0708
Variation	1.19	1.67	1 F
When study conducted?	0	1.67	0.4167 F
Which yogurt for comparison?	1.19	1.67	1 F
Why study conducted?	0	1.67	0.4167 F

*Exact p-values were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

Table 7: Post-course percentages for topic categories by sex for the LiveActive advertisement

Topic Category	Percent female	Percent male	p-value for differences*
Accuracy and Reliability?	17.86	6.67	0.0504
Agenda?	15.48	6.67	0.1063
Amount of cottage cheese to eat?	0	1.67	0.4167 F
Be that other woman definition	1.19	0	1 F
Bias	16.67	10	0.2541
Causality	8.33	6.67	0.7625 F
Effect of definition of occasional irregularity on 4 of 5	19.05	20	0.8868
Effect of definition of occasional on 4 of 5	8.33	5	0.522 F
Effect of factors on 4 of 5	26.19	8.33	0.0068
Effect of how information obtained on 4 of 5	32.14	20	0.106
Effect of location on irregularity	5.95	5	1 F
Effect of lurking variable on 4 of 5	3.57	1.67	0.6408 F
Factors affecting irregularity	45.24	35	0.2181
Generalize?	8.33	5	0.522 F
How did they get information?	84.52	70	0.0366
Length of study?	1.19	0	1 F
Location?	15.48	16.67	0.8475
Lurking variable	16.67	10	0.2541
Measurement error	13.1	6.67	0.2131
Men not included	2.38	0	0.5105 F
Misconception	9.52	10	0.9242
Misrepresenting	0	0	.
Misunderstand ad	9.52	10	0.9242
MOE/CI included?	11.9	11.67	0.9652
Non-response	9.52	5	0.361 F
Other woman question	2.38	3.33	1 F

Population	8.33	3.33	0.3051 F
Question premise of ad	20.24	25	0.498
Question wording	19.05	11.67	0.2333
Relativeness of 4 of 5	0	0	.
Relevance of 4 of 5	1.19	0	1 F
Representative?	22.62	8.33	0.0233
Round number	1.19	3.33	0.5708 F
Sample size?	44.05	41.67	0.776
Sensitive topic	8.33	5	0.522 F
Source of 4 of 5	21.43	15	0.3299
Spoon size	2.38	0	0.5105 F
Statistic and parameter	4.76	0	0.1407 F
Untrustworthy	1.19	1.67	1 F
Vague claim	10.71	6.67	0.4034
Validity	5.95	8.33	0.7417 F
Variation	0	3.33	0.1719 F
When study conducted?	1.19	5	0.308 F
Which yogurt for comparison?	7.14	3.33	0.4691 F
Why study conducted?	1.19	1.67	1 F

*Exact p-values were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

Table 8: Results from McNemar's tests by sex for topic categories for the LiveActive advertisement

Topic Category	Sex	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Accuracy and Reliability?	Female	59	1	14	16.67	10	11.90	0.5413
	Male	49	3	1	1.67	7	11.67	0.0703
Agenda?	Female	68	2	11	13.10	3	3.57	0.0574
	Male	48	0	4	6.67	8	13.33	0.3877
Amount of cottage cheese to eat?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	1	1.67	0	0.00	1
Be that other woman definition	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Bias	Female	66	2	12	14.29	4	4.76	0.0768
	Male	50	0	6	10.00	4	6.67	0.7539
Causality	Female	74	2	5	5.95	3	3.57	0.7266
	Male	46	1	3	5.00	10	16.67	0.0923
Effect of definition of occasional irregularity on 4 of 5	Female	66	2	14	16.67	2	2.38	0.0042
	Male	47	3	9	15.00	1	1.67	0.0215
Effect of definition of occasional on 4 of 5	Female	75	0	7	8.33	2	2.38	0.1797
	Male	55	2	1	1.67	2	3.33	1
Effect of factors on 4 of 5	Female	54	12	10	11.90	8	9.52	0.8145
	Male	50	1	4	6.67	5	8.33	1
Effect of how information obtained on 4 of 5	Female	54	3	24	28.57	3	3.57	<0.0001
	Male	46	4	8	13.33	2	3.33	0.1094
Effect of location on irregularity	Female	77	0	5	5.95	2	2.38	0.4531
	Male	56	2	1	1.67	1	1.67	1
Effect of lurking variable on 4 of 5	Female	81	0	3	3.57	0	0.00	0.25
	Male	59	0	1	1.67	0	0.00	1
Factors affecting	Female	27	28	10	11.90	19	22.62	0.136

irregularity								
	Male	29	8	13	21.67	10	16.67	0.6776
Generalize?	Female	70	2	5	5.95	7	8.33	0.7744
	Male	55	0	3	5.00	2	3.33	1
How did they get information?	Female	9	34	37	44.05	4	4.76	<0.0001
	Male	11	16	26	43.33	7	11.67	0.0013
Length of study?	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Location?	Female	64	2	11	13.10	7	8.33	0.4807
	Male	47	4	6	10.00	3	5.00	0.5078
Lurking variable	Female	69	0	14	16.67	1	1.19	0.001
	Male	54	0	6	10.00	0	0.00	0.0313
Measurement error	Female	72	2	9	10.71	1	1.19	0.0215
	Male	54	0	4	6.67	2	3.33	0.6875
Men not included	Female	82	0	2	2.38	0	0.00	0.5
	Male	59	0	0	0.00	1	1.67	1
Misconception	Female	71	1	7	8.33	5	5.95	0.7744
	Male	52	1	5	8.33	2	3.33	0.4531
Misrepresenting	Female	83	0	0	0.00	1	1.19	1
	Male	60	0	0	0.00	0	0.00	.
Misunderstand ad	Female	71	2	6	7.14	5	5.95	1
	Male	51	1	5	8.33	3	5.00	0.7266
MOE/CI included?	Female	74	0	10	11.90	0	0.00	0.002
	Male	53	1	6	10.00	0	0.00	0.0313
Non-response	Female	76	0	8	9.52	0	0.00	0.0078
	Male	57	0	3	5.00	0	0.00	0.25
Other woman question	Female	80	0	2	2.38	2	2.38	1
	Male	57	0	2	3.33	1	1.67	1
Population	Female	74	0	7	8.33	3	3.57	0.3438
	Male	58	1	1	1.67	0	0.00	1
Question premise of ad	Female	53	7	10	11.90	14	16.67	0.5413
	Male	35	8	7	11.67	10	16.67	0.6291
Question wording	Female	68	0	16	19.05	0	0.00	<0.0001
	Male	52	0	7	11.67	1	1.67	0.0703

Relativeness of 4 of 5	Female	83	0	0	0.00	1	1.19	1
	Male	59	0	0	0.00	1	1.67	1
Relevance of 4 of 5	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Representative?	Female	63	3	16	19.05	2	2.38	0.0013
	Male	50	1	4	6.67	5	8.33	1
Round number	Female	83	0	1	1.19	0	0.00	1
	Male	58	0	2	3.33	0	0.00	0.5
Sample size?	Female	36	17	20	23.81	11	13.10	0.1496
	Male	25	10	14	23.33	11	18.33	0.5413
Sensitive topic	Female	77	0	7	8.33	0	0.00	0.0156
	Male	57	0	3	5.00	0	0.00	0.25
Source of 4 of 5	Female	51	5	13	15.48	15	17.86	0.8506
	Male	42	6	3	5.00	9	15.00	0.146
Spoon size	Female	82	0	2	2.38	0	0.00	0.5
	Male	60	0	0	0.00	0	0.00	.
Statistic and parameter	Female	80	0	4	4.76	0	0.00	0.125
	Male	60	0	0	0.00	0	0.00	.
Untrustworthy	Female	82	0	1	1.19	1	1.19	1
	Male	59	0	1	1.67	0	0.00	1
Vague claim	Female	72	0	9	10.71	3	3.57	0.146
	Male	53	0	4	6.67	3	5.00	1
Validity	Female	74	0	5	5.95	5	5.95	1
	Male	46	0	5	8.33	9	15.00	0.424
Variation	Female	83	0	0	0.00	1	1.19	1
	Male	58	1	1	1.67	0	0.00	1
When study conducted?	Female	83	0	1	1.19	0	0.00	1
	Male	56	0	3	5.00	1	1.67	0.625
Which yogurt for comparison?	Female	77	0	6	7.14	1	1.19	0.125
	Male	58	1	1	1.67	0	0.00	1
Why study conducted?	Female	83	0	1	1.19	0	0.00	1
	Male	58	0	1	1.67	1	1.67	1

* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-value for comparing p_{01} and p_{10}

Table 9: Results from McNemar's tests for topic categories for the depression article

Topic Category	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Accuracy/Reliability	102	4	16	11.11	22	15.28	0.4177
Agenda	137	1	5	3.47	1	0.69	0.2188
At start of the study	142	0	1	0.69	1	0.69	1
Average of 4 years	143	0	0	0.00	1	0.69	1
Bias	117	3	18	12.50	6	4.17	0.0227
Boys depression rate	136	1	4	2.78	3	2.08	1
Correlation - cause and effect	141	0	3	2.08	0	0.00	0.25
Coverage error	144	0	0	0.00	0	0.00	.
Dark figure	141	0	3	2.08	0	0.00	0.25
Definition of adolescence	141	0	3	2.08	0	0.00	0.25
Definition of anxiety	143	0	0	0.00	1	0.69	1
Definition of major	143	0	1	0.69	0	0.00	1
Definition of major depressive episode	26	56	62	43.06	0	0.00	<0.0001
Definition of substantial	143	1	0	0.00	0	0.00	.
Effect of definition of depression	72	17	48	33.33	7	4.86	<0.0001
Effect of factors on results	87	10	23	15.97	24	16.67	1
Effect of girls admit more than boys	136	1	2	1.39	5	3.47	0.4531
Effect of how information obtained on statistics	70	17	40	27.78	17	11.81	0.0032
Effect of location on results	96	8	21	14.58	19	13.19	0.8746
Effect of lurking variables	135	0	9	6.25	0	0.00	0.0039
Effect of number of boys and girls	97	11	20	13.89	16	11.11	0.6177
Effect of non-teen years included	136	0	5	3.47	3	2.08	0.7266
Effect of non-response	142	0	2	1.39	0	0.00	0.5
Effect of question asked on results	110	3	27	18.75	4	2.78	<0.0001
Effect of same number	132	3	4	2.78	5	3.47	1

of gender by age							
Effect of when study done on results	137	0	7	4.86	0	0.00	0.0156
Errors	137	0	7	4.86	0	0.00	0.0156
Exact number or percentage?	139	0	5	3.47	0	0.00	0.0625
Extraneous variable	141	0	3	2.08	0	0.00	0.25
Factors affecting depression	61	31	22	15.28	30	20.83	0.3317
Generalize	108	7	18	12.50	11	7.64	0.2649
Girls more likely to admit depression	98	12	25	17.36	9	6.25	0.009
Girls vs. boys social situation	142	0	0	0.00	2	1.39	0.5
How was information obtained?	25	62	44	30.56	13	9.03	<0.0001
Includes non-teen years	133	1	7	4.86	3	2.08	0.3438
Inconsistent	143	0	0	0.00	1	0.69	1
Journal credibility	143	0	1	0.69	0	0.00	1
Last statement concerns	137	0	4	2.78	3	2.08	1
Length of study?	139	0	3	2.08	2	1.39	1
Location of study?	79	17	26	18.06	22	15.28	0.6655
Lurking variable	124	0	19	13.19	1	0.69	<0.0001
Measurement error	89	9	37	25.69	9	6.25	0.0000
Misconception	126	0	10	6.94	8	5.56	0.8145
MOE/CI included?	132	1	10	6.94	1	0.69	0.0117
Multiple depressive episodes definition	143	0	0	0.00	1	0.69	1
Non-response error	136	0	8	5.56	0	0.00	0.0078
Not understand statistics given	140	0	2	1.39	2	1.39	1
Number of boys and girls?	66	22	41	28.47	15	10.42	0.0007
One in 10 each year during study?	143	0	0	0.00	1	0.69	1
Population or population size?	136	0	7	4.86	1	0.69	0.0703
p-value and alpha	143	0	1	0.69	0	0.00	1
Qualification of researcher?	141	1	2	1.39	0	0.00	0.5
Question 2 times	136	0	6	4.17	2	1.39	0.2891

statistic							
Question premise of article	104	4	6	4.17	30	20.83	<0.0001
Questions asked?	99	6	35	24.31	4	2.78	<0.0001
Relevance?	142	0	1	0.69	1	0.69	1
Representative?	113	5	21	14.58	5	3.47	0.0025
Results for different age groups?	143	0	1	0.69	0	0.00	1
Same no. of boys and girls in each age	130	3	5	3.47	6	4.17	1
Sample size?	122	1	13	9.03	8	5.56	0.3833
Sensitive subject	142	0	2	1.39	0	0.00	0.5
Simpson's Paradox	143	0	1	0.69	0	0.00	1
"Some time in her life" concern	143	0	1	0.69	0	0.00	1
Statistic vs. Parameter	143	0	1	0.69	0	0.00	1
Trustworthy?	139	0	3	2.08	2	1.39	1
US vs. Canada depression rate	140	0	2	1.39	2	1.39	1
Vague claim	139	1	4	2.78	0	0.00	0.125
Validity	123	2	13	9.03	6	4.17	0.1671
When did study occur?	134	0	8	5.56	2	1.39	0.1094
Who funded/conducted study?	137	2	4	2.78	1	0.69	0.375

* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-value for comparing p_{01} and p_{10}

Table 10: Pre-course percentages for topic categories by sex for the depression article

Topic Category	Percent female	Percent male	p-value for differences*
Accuracy/Reliability	19.05	16.67	0.7142
Agenda	1.19	1.67	1 F
At start of the study	1.19	0	1 F
Average of 4 years	0	1.67	0.4167
Bias	7.14	5	0.7352 F
Boys depression rate	2.38	3.33	1 F
Correlation - cause and effect	0	0	.
Coverage error	0	0	.
Dark figure	0	0	.
Definition of adolescence	0	0	.
Definition of anxiety	0	1.67	0.4167 F
Definition of major	0	0	.
Definition of major depressive episode	39.29	38.33	0.908
Definition of substantial	1.19	0	1 F
Effect of definition of depression	17.86	15	0.6501
Effect of factors on results	23.81	23.33	0.9471
Effect of girls admit more than boys	3.57	5	0.6935 F
Effect of how information obtained on statistics	28.57	16.67	0.0972
Effect of location on results	20.24	16.67	0.5883
Effect of lurking variables	0	0	.
Effect of number of boys and girls	21.43	15	0.3299
Effect of non-teen years included	2.38	1.67	1 F
Effect of non-response	0	0	.
Effect of question asked on results	3.57	6.67	0.4507 F
Effect of same number of gender by age	7.14	3.33	0.4691 F
Effect of when study done on results	0	0	.
Errors	0	0	.
Exact number or	0	0	.

percentage			
Extraneous variable	0	0	.
Factors affecting depression	48.81	33.33	0.0639
Generalize	9.52	16.67	0.2013
Girls more likely to admit depression	15.48	13.33	0.7195
Girls vs. boys social situation	2.38	0	0.5105 F
How was information obtained?	54.76	48.33	0.4465
Includes non-teen years	2.38	3.33	1 F
Inconsistent	1.19	0	1 F
Journal credibility	0	0	.
Last statement concerns	1.19	3.33	0.5708 F
Length of study?	1.19	1.67	1 F
Location of study?	30.95	21.67	0.2164
Lurking variable	1.19	0	1 F
Measurement error	8.33	18.33	0.0758
Misconception	2.38	10	0.0673 F
MOE/CI included?	1.19	1.67	1 F
Multiple depressive episodes definition	0	1.67	0.4167 F
Non-response error	0	0	.
Not understand statistics given	2.38	0	0.5105 F
Number of boys and girls?	29.76	20	0.1863
One in 10 each year during study?	0	1.67	0.4167 F
Population or population size?	1.19	0	1 F
p-value and alpha	0	0	.
Qualification of researcher?	0	1.67	0.4167 F
Question 2 times statistic	1.19	1.67	1 F
Question premise of article	19.05	30	0.1271
Questions asked?	5.95	8.33	0.7417 F
Relevance?	1.19	0	1 F
Representative?	8.33	5	0.5220 F
Results for different age groups?	0	0	.
Same number of boys and girls in each age	8.33	3.33	0.3051 F

Sample size?	3.57	10	0.1641 F
Sensitive subject	0	0	.
Simpson's Paradox	0	0	.
"Some time in her life" concern	0	0	.
Statistic vs. Parameter	0	0	.
Trustworthy?	1.19	1.67	1 F
US vs. Canada depression rate	2.38	0	0.5105 F
Vague claim	1.19	0	1 F
Validity	3.57	8.33	0.2778 F
When did study occur?	2.38	0	0.5105 F
Who funded/conducted study?	1.19	3.33	0.5708 F

*Exact p-values were reported unless an "F" appeared indicating that the p-value resulted from Fisher's exact test

Table 11: Post-course percentages for topic categories by sex for the depression article

Topic Category	Percent female	Percent male	p-value for differences*
Accuracy/Reliability	17.86	8.33	0.1033
Agenda	4.76	3.33	1 F
At start of the study	0	1.67	0.4167 F
Average of 4 years	0	0	.
Bias	17.86	10	0.1878
Boys depression rate	1.19	6.67	0.1607 F
Correlation - cause and effect	1.19	3.33	0.5708 F
Coverage error	4.76	1.67	0.4014 F
Dark figure	2.38	1.67	1 F
Definition of adolescence	3.57	0	0.2657 F
Definition of anxiety	0	0	.
Definition of major	1.19	0	1 F
Definition of major depressive episode	86.9	75	0.0671
Definition of substantial	1.19	0	1 F
Effect of definition of depression	47.62	41.67	0.4792
Effect of factors on results	29.76	13.33	0.0208
Effect of girls admit more than boys	2.38	1.67	1 F
Effect of how information obtained on statistics	51.19	23.33	0.0008
Effect of location on results	16.67	25	0.2189
Effect of lurking variables	5.95	6.67	1 F
Effect of number of boys and girls	22.62	20	0.7062
Effect of non-teen years included	3.57	3.33	1 F
Effect of non-response	0	3.33	0.1719 F
Effect of question asked on results	22.62	18.33	0.5324
Effect of same number of gender by age	5.95	3.33	0.6994 F
Effect of when study done on results	5.95	3.33	0.6994 F
Errors	4.76	5	1 F
Exact number or percentage	4.76	1.67	0.4014 F
Extraneous variable	1.19	3.33	0.5708 F

Factors affecting depression	39.29	33.33	0.4653
Generalize	14.29	21.67	0.249
Girls more likely to admit depression	25	26.67	0.8215
Girls vs. boys social situation	0	0	.
How was information obtained?	86.9	55	<0.0001
Includes non-teen years	5.95	5	1 F
Inconsistent	0	0	.
Journal credibility	1.19	0	1 F
Last statement concerns	4.76	0	0.1407 F
Length of study?	2.38	1.67	1 F
Location of study?	23.81	38.33	0.0604
Lurking variable	14.29	11.67	0.6471
Measurement error	27.38	38.33	0.1646
Misconception	9.52	3.33	0.1942 F
MOE/CI included?	7.14	8.33	1 F
Multiple depressive episodes definition	0	0	.
Non-response error	4.76	6.67	0.7192 F
Not understand statistics given	2.38	0	0.5105 F
Number of boys and girls?	44.05	43.33	0.9321
One in 10 each year during study?	0	0	.
Population or population size?	7.14	1.67	0.2392 F
p-value and alpha	1.19	0	1 F
Qualification of researcher?	1.19	3.33	0.5708 F
Question 2 times statistic	5.95	1.67	0.4009 F
Question premise of article	7.14	6.67	1 F
Questions asked?	28.57	28.33	0.9751
Relevance?	1.19	0	1 F
Representative?	17.86	18.33	0.9416
Results for different age groups?	0	1.67	0.4167 F
Same number of boys and girls in each age	5.95	5	1 F
Sample size?	13.1	5	0.106
Sensitive subject	0	3.33	0.1719 F
Simpson's Paradox	1.19	0	1 F
"Some time in her life"	1.19	0	1 F

concern			
Statistic vs. Parameter	1.19	0	1 F
Trustworthy?	1.19	3.33	0.5708 F
US vs. Canada depression rate	0	3.33	0.1719 F
Vague claim	3.57	3.33	1 F
Validity	11.9	8.33	0.4891
When did study occur?	7.14	3.33	0.4691 F
Who funded/conducted study?	2.38	6.67	0.2347 F

*Exact p-values were reported unless an “F” appeared indicating that the p-value resulted from Fisher’s exact test

Table 12: Results from McNemar's tests by sex for topic categories for the depression article

Topic Category	Sex	n_{00}*	n_{11}*	n_{01}*	$\hat{p}_{01}$ **	n_{10}*	$\hat{p}_{10}$ **	p-value***
Accuracy/Reliability	Female	56	3	12	14.29	13	15.48	1
	Male	46	1	4	6.67	9	15.00	0.2668
Agenda	Female	79	0	4	4.76	1	1.19	0.375
	Male	58	1	1	1.67	0	0.00	1
At start of the study	Female	83	0	0	0.00	1	1.19	1
	Male	59	0	1	1.67	0	0.00	1
Average of 4 years	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	0	0.00	1	1.67	1
Bias	Female	65	2	13	15.48	4	4.76	0.049
	Male	52	1	5	8.33	2	3.33	0.4531
Boys depression rate	Female	81	0	1	1.19	2	2.38	1
	Male	55	1	3	5.00	1	1.67	0.625
Correlation - cause and effect	Female	83	0	1	1.19	0	0.00	1
	Male	58	0	2	3.33	0	0.00	0.5
Coverage error	Female	84	0	0	0.00	0	0.00	.
	Male	60	0	0	0.00	0	0.00	.
Dark figure	Female	82	0	2	2.38	0	0.00	0.5
	Male	59	0	1	1.67	0	0.00	1
Definition of adolescence	Female	81	0	3	3.57	0	0.00	0.25
	Male	60	0	0	0.00	0	0.00	.
Definition of anxiety	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	0	0.00	1	1.67	1
Definition of major	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Definition of major depressive episode	Female	11	33	40	47.62	0	0.00	<0.0001
	Male	15	23	22	36.67	0	0.00	<0.0001
Definition of substantial	Female	83	1	0	0.00	0	0.00	.
	Male	60	0	0	0.00	0	0.00	.
Effect of definition of depression	Female	41	12	28	33.33	3	3.57	<0.0001
	Male	31	5	20	33.33	4	6.67	0.0015

Effect of factors on results	Female	46	7	18	21.43	13	15.48	0.4731
	Male	41	3	5	8.33	11	18.33	0.2101
Effect of girls admit more than boys	Female	80	1	1	1.19	2	2.38	1
	Male	56	0	1	1.67	3	5.00	0.625
Effect of how information obtained on statistics	Female	30	13	30	35.71	11	13.10	0.0043
	Male	40	4	10	16.67	6	10.00	0.4545
Effect of location on results	Female	57	4	10	11.90	13	15.48	0.6776
	Male	39	4	11	18.33	6	10.00	0.3323
Effect of lurking variables	Female	79	0	5	5.95	0	0.00	0.0625
	Male	56	0	4	6.67	0	0.00	0.125
Effect of number of boys and girls	Female	54	7	12	14.29	11	13.10	1
	Male	43	4	8	13.33	5	8.33	0.5811
Effect of non-teen years included	Female	79	0	3	3.57	2	2.38	1
	Male	57	0	2	3.33	1	1.67	1
Effect of non-response	Female	84	0	0	0.00	0	0.00	
	Male	58	0	2	3.33	0	0.00	0.5
Effect of question asked on results	Female	64	2	17	20.24	1	1.19	0.0001
	Male	46	1	10	16.67	3	5.00	0.0923
Effect of same number of gender by age	Female	76	3	2	2.38	3	3.57	1
	Male	56	0	2	3.33	2	3.33	1
Effect of when study done on results	Female	79	0	5	5.95	0	0.00	0.0625
	Male	58	0	2	3.33	0	0.00	0.5
Errors	Female	80	0	4	4.76	0	0.00	0.125
	Male	57	0	3	5.00	0	0.00	0.25
Exact number or percentage	Female	80	0	4	4.76	0	0.00	0.125
	Male	59	0	1	1.67	0	0.00	1
Extraneous variable	Female	83	0	1	1.19	0	0.00	1
	Male	58	0	2	3.33	0	0.00	0.5

Factors affecting depression	Female	33	23	10	11.90	18	21.43	0.1849
	Male	28	8	12	20.00	12	20.00	1
Generalize	Female	66	2	10	11.90	6	7.14	0.4545
	Male	42	5	8	13.33	5	8.33	0.5811
Girls more likely to admit depression	Female	58	8	13	15.48	5	5.95	0.0963
	Male	40	4	12	20.00	4	6.67	0.0768
Girls vs. boys social situation	Female	82	0	0	0.00	2	2.38	0.5
	Male	60	0	0	0.00	0	0.00	.
How was information obtained?	Female	5	40	33	39.29	6	7.14	<0.0001
	Male	20	22	11	18.33	7	11.67	0.4807
Includes non-teen years	Female	78	1	4	4.76	1	1.19	0.375
	Male	55	0	3	5.00	2	3.33	1
Inconsistent	Female	83	0	0	0.00	1	1.19	1
	Male	60	0	0	0.00	0	0.00	.
Journal credibility	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	
Last statement concerns	Female	79	0	4	4.76	1	1.19	0.375
	Male	58	0	0	0.00	2	3.33	0.5
Length of study?	Female	81	0	2	2.38	1	1.19	1
	Male	58	0	1	1.67	1	1.67	1
Location of study?	Female	47	9	11	13.10	17	20.24	0.3449
	Male	32	8	15	25.00	5	8.33	0.0414
Lurking variable	Female	71	0	12	14.29	1	1.19	0.0034
	Male	53	0	7	11.67	0	0.00	0.0156
Measurement error	Female	57	3	20	23.81	4	4.76	0.0015
	Male	32	6	17	28.33	5	8.33	0.0169
Misconception	Female	74	0	8	9.52	2	2.38	0.1094
	Male	52	0	2	3.33	6	10.00	0.2891
MOE/CI included?	Female	77	0	6	7.14	1	1.19	0.125
	Male	55	1	4	6.67	0	0.00	0.125
Multiple depressive episodes definition	Female	84	0	0	0.00	0	0.00	.

	Male	59	0	0	0.00	1	1.67	1
Non-response error	Female	80	0	4	4.76	0	0.00	0.125
	Male	56	0	4	6.67	0	0.00	0.125
Not understand statistics given	Female	80	0	2	2.38	2	2.38	1
	Male	60	0	0	0.00	0	0.00	.
Number of boys and girls?	Female	35	13	24	28.57	12	14.29	0.0652
	Male	31	9	17	28.33	3	5.00	0.0026
One in 10 each year during study?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	0	0.00	1	1.67	1
Population or population size?	Female	77	0	6	7.14	1	1.19	0.125
	Male	59	0	1	1.67	0	0.00	1
p-value and alpha	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Qualification of researcher?	Female	83	0	1	1.19	0	0.00	1
	Male	58	1	1	1.67	0	0.00	1
Question 2 times statistic	Female	78	0	5	5.95	1	1.19	0.2188
	Male	58	0	1	1.67	1	1.67	1
Question premise of article	Female	65	3	3	3.57	13	15.48	0.0213
	Male	39	1	3	5.00	17	28.33	0.0026
Questions asked?	Female	58	3	21	25.00	2	2.38	<0.0001
	Male	41	3	14	23.33	2	3.33	0.0042
Relevance?	Female	82	0	1	1.19	1	1.19	1
	Male	60	0	0	0.00	0	0.00	.
Representative?	Female	66	4	11	13.10	3	3.57	0.0574
	Male	47	1	10	16.67	2	3.33	0.0386
Results for different age groups?	Female	84	0	0	0.00	0	0.00	.
	Male	59	0	1	1.67	0	0.00	1
Same no. of boys and girls in each age	Female	75	3	2	2.38	4	4.76	0.6875
	Male	55	0	3	5.00	2	3.33	1
Sample size?	Female	71	1	10	11.90	2	2.38	0.0386
	Male	51	0	3	5.00	6	10.00	0.5078

Sensitive subject	Female	84	0	0	0.00	0	0.00	.
	Male	58	0	2	3.33	0	0.00	0.5
Simpson's Paradox	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
"Some time in her life" concern	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Statistic vs. Parameter	Female	83	0	1	1.19	0	0.00	1
	Male	60	0	0	0.00	0	0.00	.
Trustworthy?	Female	82	0	1	1.19	1	1.19	1
	Male	57	0	2	3.33	1	1.67	1
US vs. Canada depression rate	Female	82	0	0	0.00	2	2.38	0.5
	Male	58	0	2	3.33	0	0.00	0.5
Vague claim	Female	81	0	2	2.38	1	1.19	0.5
	Male	58	0	2	3.33	0	0.00	0.5
Validity	Female	72	1	9	10.71	2	2.38	0.0654
	Male	51	1	4	6.67	4	6.67	1
When did study occur?	Female	76	0	6	7.14	2	2.38	0.2891
	Male	58	0	2	3.33	0	0.00	0.5
Who funded/conducted study?	Female	82	1	1	1.19	0	0.00	1
	Male	55	1	3	5.00	1	1.67	0.625

* n_{00} represented the number of students who did not have a concern about the topic category prior to and after SIEL
n_{11} represented the number of students who had a concern about the topic category prior to and after SIEL
n_{01} represented the number of students who did not have a concern about the topic category prior to SIEL and did have a concern about the topic category after SIEL
n_{10} represented the number of students who did have a concern about the topic category prior to SIEL and did not have a concern about the topic category after SIEL

** $\hat{p}_{01}$ represented the percentage of students who did not have a concern about the topic category prior to SIEL but did have a concern about the topic category after SIEL
$\hat{p}_{10}$ represented the percentage of students who did have a concern about the topic category prior to SIEL but did not have a concern about the topic category after SIEL

*** Exact p-value for comparing p_{01} and p_{10}

Appendix F

Tables for Results for Research Question 3

Table 1: Attributes to be successful in SIEL

Attribute to be successful	Percentage
Accept past mistakes	1.67%
Analytical	5.00%
Apply definitions	3.33%
Apply yourself	1.67%
Appreciation for statistics	1.67%
Ask questions	8.33%
Attend class	35.00%
Be excited about course	1.67%
Be prepared or willing to learn new things	8.33%
Common sense	3.33%
Comprehend word problems	6.67%
Conscientious	1.67%
Creative	3.33%
Dedication	5.00%
Desire	1.67%
Determination/driven	3.33%
Different way of thinking	3.33%
Do basic math or math skills	23.33%
Do homework or practice	28.33%
Do not think analytically	1.67%
Do not underestimate class	1.67%
Do not need to be good or interested in statistics	1.67%
Effort	6.67%
Good listeners	8.33%
Good memory	3.33%
Intelligence	1.67%
Interest	1.67%
Learn definitions	6.67%
Learn or use formulas	13.33%
Like math	1.67%
Memorize	8.33%
Motivation	6.67%
Observant	1.67%
Open mind	25.00%
Organized	6.67%

Participate in activities or class	10.00%
Patience	3.33%
Pay attention in class	20.00%
Perceptive	1.67%
Print notes for class	6.67%
Problem solving or logical thinking	20.00%
Question statistics	5.00%
Responsibility	3.33%
Self-control	1.67%
Serious attitude	1.67%
Study	13.33%
Take good notes	6.67%
Think or think on feet	3.33%
Understand and apply concepts	11.67%
Understand basics	1.67%
Understand big picture	8.33%
Understand graphs	1.67%
Understand marketing and advertising	1.67%
Understand statistics before course	5.00%
Understand theories	1.67%
Visual learner	1.67%
Want to learn	1.67%
Work hard or work ethic	18.33%
Work well in groups	5.00%

Table 2: Overall percentages for adjectives to describe statistics

Adjective to describe statistics	Percentage
Analytical	5.00%
Applicable	1.67%
Biased	1.67%
Calculating	1.67%
Challenging	10.00%
Common	1.67%
Compelling	1.67%
Complex	10.00%
Complicated	6.67%
Conceptual	3.33%
Conditional	3.33%
Confusing	3.33%
Controversial	3.33%
Deceiving	1.67%
Deceptive	1.67%
Descriptive	1.67%
Different	1.67%
Difficult	5.00%
Dubious	1.67%
Doable	1.67%
Dynamic	1.67%
Easily manipulated	1.67%
Easy	3.33%
Engaging	1.67%
Effective	1.67%
Entertaining	1.67%
Essential	1.67%
Exciting	1.67%
False	1.67%
Fluctuating	1.67%
Formula related	1.67%
Fun	1.67%
Helpful	3.33%
Human	1.67%
Important	11.67%
In depth	3.33%
Inaccurate	1.67%
Incorrect	1.67%
Informative	10.00%

Insightful	6.67%
Interactive	1.67%
Interesting	23.33%
Interpretive	1.67%
Intricate	1.67%
Intriguing	3.33%
Logical	3.33%
Malleable	1.67%
Manipulated	6.67%
Measured	1.67%
Mathematical	5.00%
Methodical	3.33%
Meticulous	1.67%
Mind-altering	1.67%
Misleading	6.67%
Misunderstood	1.67%
Misused	3.33%
Necessary	3.33%
Numerical	3.33%
Numerical information	1.67%
Open mind	1.67%
Ordered	1.67%
Positive	1.67%
Persuasive	3.33%
Picky	1.67%
Practical	1.67%
Predictive	1.67%
Productive	1.67%
Quantitative data	1.67%
Questionable	1.67%
Rational	1.67%
Representative	3.33%
Revealing	1.67%
Rewarding	1.67%
Risky	1.67%
Sample	3.33%
Scholarly	1.67%
Shocking	1.67%
Subjective	5.00%
Supportive	1.67%
Surprising	1.67%
Team work	1.67%

Theoretical	1.67%
Thought provoking	5.00%
Time consuming	1.67%
Tricky	1.67%
Ubiquitous	3.33%
Unpredictable	1.67%
Useful	25.00%
Vague	5.00%
Vital	1.67%
Widespread	1.67%
Wordy	1.67%

Please describe past experiences in other mathematics or statistics courses that you have taken.

Table 3: Percentages based on student sex versus view of past experiences

	View of past experience			
Sex	Indifferent	Mixed	Negative	Positive
Female	8.57%	11.43%	40%	40%
Male	32%	4%	28%	36%

p-value for Fisher's Exact Test 0.1247

Explain how your experiences in SIEL were similar to your past experiences in mathematics or statistics courses.

Table 4: Percentages for similarities of past experiences in mathematics or statistics courses
with SIEL based on student sex

	Similarities with other courses				
Sex	Anxiety	Content	Course format	None	Problem solving
Female	5.71%	48.57%	34.29%	8.57%	2.86%
Male	8%	64%	24%	0%	4%

p-value for Fisher's Exact Test 0.5240

Explain how your experiences in SIEL were different from your past experiences in mathematics or statistics courses.

Table 5: Percentages for differences of past experiences in mathematics or statistics courses with SIEL based on student sex

	Differences with other courses			
Sex	Analysis	Application to everyday life	Course format	First Statistics course
Female	8.57%	31.43%	57.14%	2.86%
Male	36%	36%	24%	4%

p-value for Fisher's Exact Test 0.0145

Explain what "statistical literacy" means to you.

Table 6: Percentages of students' understanding of the concept of statistical literacy based on student sex

	Basic understanding of statistical literacy	
Sex	No	Yes
Female	25.71%	74.29%
Male	40%	60%

p-value for Chi-square test 0.2409

Do you feel that you are "statistically literate?"

Table 7: Percentages of students' who believe they are statistically literate based on student sex

	Statistically literate?	
Sex	No	Yes
Female	2.86%	97.14%
Male	8%	92%

p-value for Fisher's Exact test 0.5653

Has SIEL changed the way in which you look at the statistics you encounter in everyday life?

Table 8: Percentages of students' who say that the course has changed the way they look at statistics encountered in everyday life based on student sex

	Changed way look at everyday life statistics?	
Sex	No	Yes
Female	14.29%	85.17%
Male	12%	88%

p-value for Fisher's Exact test 1.00

Attributes to be successful: Abilities

Table 9: Percentages of students who said ability attribute was needed to be successful in course by sex

	Abilities?	
Sex	No	Yes
Female	45.71%	54.29%
Male	44%	56%

p-value for Chi-square test 0.8953

Attributes to be successful: Class behavior

Table 10: Percentages of students who said class behavior attribute was needed to be successful in course by sex

	Class behavior?	
Sex	No	Yes
Female	28.57%	71.43%
Male	60%	40%

p-value for Chi-square test 0.0149

Attributes to be successful: Student quality

Table 11: Percentages of students who said student quality attribute was needed to be successful in course by sex

	Student quality?	
Sex	No	Yes
Female	34.29%	65.71%
Male	40%	60%

p-value for Chi-square test 0.6507

Attributes to be successful: Thinking capabilities

Table 12: Percentages of students who said thinking capabilities attribute was needed to be successful in course by sex

	Thinking capabilities?	
Sex	No	Yes
Female	57.14%	42.86%
Male	60%	40%

p-value for Chi-square test 0.8248

During the course were you challenged with material that challenged your beliefs?

Table 13: Percentages of students' who say that during the course were you challenged with material that challenged your beliefs based on student sex

	Beliefs challenged?	
Sex	No	Yes
Female	91.43%	8.57%
Male	96%	4%

p-value for Fisher's Exact test 0.6339

How would you describe your attitude toward statistics – good, moderate, or poor?

Table 14: Percentages of students' who describe their attitude toward statistics as good, moderate or poor based on student sex

	Attitude		
Sex	Good	Moderate	Poor
Female	40%	60%	0%
Male	60%	40%	0%

p-value for Fisher's Exact test 0.1902 (compare good and moderate without poor attitude)

As a result of this class, has your attitude toward statistics changed?

Table 15: Percentages of students' who say that their attitude toward statistics has changed as a result of this course based on student sex

	Attitude changed?	
Sex	No	Yes
Female	25.71%	74.29%
Male	40%	60%

p-value for Chi-square test 0.2409

For those students who said that their attitude had not changed as a result of this course, how did they describe their attitude?

Table 16: Description of attitude as a result of course for those who said that their attitude was not changed based on student sex

	Attitude		
Sex	Negative	Neutral	Positive
Female	11.11%	33.33%	55.56%
Male	10%	40%	50%

p-value for Fisher's Exact test 1.00

If attitude had changed as a result of this course, in what way did it change?

Table 17: Percentages of direction of students' attitude change toward statistics as a result of this course based on student sex

	Attitude change	
Sex	Negative to Positive	Neutral to Positive
Female	15.38%	84.62%
Male	13.33%	86.67%

p-value for Fisher's Exact test 1.00

Table 18: Adjectives to describe statistics for female students

Adjective to describe statistics	Percentage
Analytical	8.57%
Applicable	2.86%
Biased	2.86%
Calculating	2.86%
Challenging	11.43%
Common	0%
Compelling	2.86%
Complex	5.71%
Complicated	8.57%
Conceptual	0%
Conditional	0%
Confusing	5.71%
Controversial	5.71%
Deceiving	2.86%
Deceptive	2.86%
Descriptive	0%
Different	2.86%
Difficult	2.86%
Dubious	2.86%
Doable	2.86%
Dynamic	2.86%
Easily manipulated	2.86%
Easy	2.86%
Engaging	0%
Effective	2.86%
Entertaining	2.86%
Essential	2.86%
Exciting	0%
False	2.86%
Fluctuating	2.86%
Formula related	0%
Fun	0%
Helpful	0%
Human	2.86%
Important	8.57%
In depth	0%
Inaccurate	0%
Incorrect	2.86%
Informative	11.43%
Insightful	5.71%

Interactive	0%
Interesting	28.57%
Interpretive	2.86%
Intricate	2.86%
Intriguing	5.71%
Logical	5.71%
Malleable	2.86%
Manipulated	8.57%
Measured	0%
Mathematical	2.86%
Methodical	5.71%
Meticulous	0%
Mind-altering	0%
Misleading	2.86%
Misunderstood	0%
Misused	5.71%
Necessary	2.86%
Numerical	2.86%
Numerical information	0%
Open mind	0%
Ordered	0%
Positive	0%
Persuasive	2.86%
Picky	2.86%
Practical	0%
Predictive	2.86%
Productive	0%
Quantitative data	0%
Questionable	2.86%
Rational	0%
Representative	0%
Revealing	2.86%
Rewarding	0%
Risky	0%
Sample	2.86%
Scholarly	2.86%
Shocking	2.86%
Subjective	5.71%
Supportive	2.86%
Surprising	2.86%
Team work	0%
Theoretical	2.86%

Thought provoking	2.86%
Time consuming	0%
Tricky	2.86%
Ubiquitous	2.86%
Unpredictable	2.86%
Useful	31.43%
Vague	2.86%
Vital	2.86%
Widespread	2.86%
Wordy	2.86%

Table 19: Adjectives to describe statistics for male students

Adjective to describe statistics	Percentage
Analytical	0%
Applicable	0%
Biased	0%
Calculating	0%
Challenging	8.00%
Common	4.00%
Compelling	0%
Complex	16.00%
Complicated	4.00%
Conceptual	8.00%
Conditional	4.00%
Confusing	0%
Controversial	0%
Deceiving	0%
Deceptive	0%
Descriptive	4.00%
Different	0%
Difficult	8.00%
Dubious	0%
Doable	0%
Dynamic	0%
Easily manipulated	0%
Easy	4.00%
Engaging	4.00%
Effective	0%
Entertaining	0%
Essential	0%

Exciting	4.00%
False	0%
Fluctuating	0%
Formula related	4.00%
Fun	4.00%
Helpful	8.00%
Human	0%
Important	16.00%
In depth	8.00%
Inaccurate	4.00%
Incorrect	0%
Informative	8.00%
Insightful	8.00%
Interactive	4.00%
Interesting	16.00%
Interpretive	0%
Intricate	0%
Intriguing	0%
Logical	0%
Malleable	0%
Manipulated	4.00%
Measured	4.00%
Mathematical	8.00%
Methodical	0%
Meticulous	4.00%
Mind-altering	4.00%
Misleading	12.00%
Misunderstood	4.00%
Misused	0%
Necessary	4.00%
Numerical	4.00%
Numerical information	4.00%
Open mind	4.00%
Ordered	4.00%
Positive	4.00%
Persuasive	4.00%
Picky	0%
Practical	4.00%
Predictive	0%
Productive	4.00%
Quantitative data	4.00%
Questionable	0%

Rational	4.00%
Representative	8.00%
Revealing	0%
Rewarding	4.00%
Risky	4.00%
Sample	4.00%
Scholarly	0%
Shocking	0%
Subjective	4.00%
Supportive	0%
Surprising	0%
Team work	4.00%
Theoretical	0%
Thought provoking	8.00%
Time consuming	4.00%
Tricky	0%
Ubiquitous	4.00%
Unpredictable	0%
Useful	16.00%
Vague	8.00%
Vital	0%
Widespread	0%
Wordy	0%

Please describe past experiences in other mathematics or statistics courses that you have taken.

Table 20: Percentages based on student effort level versus view of past experiences

	View of Past Experience			
Effort	Indifferent	Mixed	Negative	Positive
Low	33.33%	0%	33.33%	33.33%
Moderate	10.52%	15.79%	52.63%	21.05%
High	20%	5.71%	25.71%	48.57%

p-value for Fisher's Exact Test 0.1813

Explain how your experiences in SIEL were similar to your past experiences in mathematics or statistics courses.

Table 21: Percentages for similarities of past experiences in mathematics or statistics courses with SIEL based on student effort level

	Similarities with other courses				
Effort	Anxiety	Content	Course format	None	Problem solving
Low	16.67%	66.67%	16.67%	0%	0%
Moderate	10.56%	63.16%	21.05%	0%	1.67%
High	2.86%	48.57%	37.14%	8.57%	2.86%

p-value for Fisher's Exact Test 0.4966

Explain how your experiences in SIEL were different from your past experiences in mathematics or statistics courses.

Table 22: Percentages for similarities of past experiences in mathematics or statistics courses with SIEL based on student effort level

	Differences with other courses			
Effort	Analysis	Application to everyday life	Course format	First Statistics course
Low	16.67%	16.67%	66.67%	0%
Moderate	26.32%	47.37%	21.05%	5.26%
High	17.14%	28.57%	51.43%	2.86%

p-value for Fisher's Exact Test 0.2860

Explain what "statistical literacy" means to you.

Table 23: Percentages of students' understanding of the concept of statistical literacy based on student effort level

	Basic understanding of statistical literacy	
Effort	No	Yes
Low	50%	50%
Moderate	15.79%	84.21%
High	37.14%	62.86%

p-value for Fisher's Exact Test 0.1343

Do you feel that you are "statistically literate?"

Table 24: Percentages of students' who believe they are statistically literate based on student effort level

	Statistically literate?	
Effort	No	Yes
Low	0%	100%
Moderate	0%	100%
High	8.57%	91.43%

p-value for Fisher's Exact Test 0.6696

Has SIEL changed the way in which you look at the statistics you encounter in everyday life?

Table 25: Percentages of students' who say that the course has changed the way they look at statistics encountered in everyday life based on student effort level

	Changed way look at everyday life statistics?	
Effort	No	Yes
Low	16.67%	83.33%
Moderate	10.53%	89.47%
High	14.29%	85.71%

p-value for Fisher's Exact Test 1.00

Attributes to be successful: Abilities

Table 26: Percentages of students who said ability attribute was needed to be successful in course by effort

	Abilities?	
Effort	No	Yes
Low	16.67%	83.33%
Moderate	47.37%	52.63%
High	48.57%	52.63%

p-value for Fisher's Exact Test 0.4058

Attributes to be successful: Class behavior

Table 27: Percentages of students who said class behavior attribute was needed to be successful in course by effort

	Class behavior?	
Effort	No	Yes
Low	83.33%	16.67%
Moderate	42.11%	57.89%
High	34.29%	65.71%

p-value for Fisher's Exact Test 0.1163

Attributes to be successful: Student quality

Table 28: Percentages of students who said student quality attribute was needed to be successful in course by effort

	Student quality?	
Effort	No	Yes
Low	50%	50%
Moderate	31.58%	68.42%
High	37.14%	62.86%

p-value for Fisher's Exact Test 0.7282

Attributes to be successful: Thinking capabilities

Table 29: Percentages of students who said thinking capabilities attribute was needed to be successful in course by effort

	Thinking Capabilities?	
Effort	No	Yes
Low	66.67%	33.33%
Moderate	52.63%	47.37%
High	60%	40%

p-value for Fisher's Exact Test 0.7985

During the course were you challenged with material that challenged your beliefs?

Table 30: Percentages of students' who say that during the course were you challenged with material that challenged your beliefs based on student effort level

	Beliefs challenged?	
Effort	No	Yes
Low	100%	0%
Moderate	100%	0%
High	88.57%	11.43%

p-value for Fisher's Exact Test 0.3972

How would you describe your attitude toward statistics – good, moderate, or poor?

Table 31: Percentages of students' who describe their attitude toward statistics as good, moderate, or poor based on student effort level

	Attitude		
Effort	Good	Moderate	Poor
Low	33.33%	66.67%	0%
Moderate	47.37%	52.63%	0%
High	51.43%	48.57%	0%

p-value for Fisher's Exact Test 0.8013 (compare good and moderate without poor attitude)

As a result of this class, has your attitude toward statistics changed?

Table 32: Percentages of students' who say that their attitude toward statistics has changed as a result of this course based on student effort level

	Attitude changed?	
Effort	No	Yes
Low	66.67%	33.33%
Moderate	31.58%	68.42%
High	25.71%	74.29%

p-value for Fisher's Exact Test 0.1484

For those students who said that their attitude had not changed as a result of this course, how did they describe their attitude?

Table 33: Description of attitude as a result of course for those who said that their attitude was not changed based on student effort level

	Attitude		
Effort	Negative	Neutral	Positive
Low	0%	75%	25%
Moderate	16.67%	16.67%	66.67%
High	11.11%	33.33%	55.56%

p-value for Fisher's Exact Test 0.5628

If attitude had changed as a result of this course, in what way did it change?

Table 34: Percentages of direction of students' attitude change toward statistics as a result of this course based on student effort level

	Attitude changed?	
Effort	Negative to Positive	Neutral to Positive
Low	0%	100%
Moderate	7.69%	92.31%
High	19.23%	80.77%

p-value for Fisher's Exact Test 0.7407

Table 35: Adjectives to describe statistics for students with high effort

Adjective to describe statistics	Percentage
Analytical	2.86%
Applicable	2.86%
Biased	0%
Calculating	2.86%
Challenging	8.57%
Common	2.86%
Compelling	2.86%
Complex	5.71%
Complicated	5.71%
Conceptual	0%
Conditional	2.86%
Confusing	2.86%
Controversial	5.71%
Deceiving	0%
Deceptive	0%
Descriptive	2.86%
Different	2.86%
Difficult	2.86%
Dubious	2.86%
Doable	2.86%
Dynamic	2.86%
Easily manipulated	2.86%
Easy	2.86%
Engaging	2.86%
Effective	0%
Entertaining	0%
Essential	2.86%
Exciting	0%
False	0%
Fluctuating	0%
Formula related	2.86%
Fun	0%
Helpful	0%
Human	2.86%
Important	17.14%
In depth	2.86%
Inaccurate	0%
Incorrect	2.86%
Informative	11.43%
Insightful	2.86%

Interactive	2.86%
Interesting	22.86%
Interpretive	2.86%
Intricate	2.86%
Intriguing	5.71%
Logical	2.86%
Malleable	2.86%
Manipulated	2.86%
Measured	2.86%
Mathematical	8.57%
Methodical	2.86%
Meticulous	0%
Mind-altering	2.86%
Misleading	5.71%
Misunderstood	2.86%
Misused	2.86%
Necessary	2.86%
Numerical	2.86%
Numerical information	2.86%
Open mind	2.86%
Ordered	0%
Positive	2.86%
Persuasive	2.86%
Picky	2.86%
Practical	2.86%
Predictive	2.86%
Productive	2.86%
Quantitative data	2.86%
Questionable	2.86%
Rational	0%
Representative	5.71%
Revealing	2.86%
Rewarding	2.86%
Risky	0%
Sample	2.86%
Scholarly	2.86%
Shocking	0%
Subjective	2.86%
Supportive	2.86%
Surprising	2.86%
Team work	2.86%
Theoretical	2.86%

Thought provoking	5.71%
Time consuming	2.86%
Tricky	2.86%
Ubiquitous	0%
Unpredictable	0%
Useful	17.14%
Vague	5.71%
Vital	2.86%
Widespread	2.86%
Wordy	2.86%

Table 36: Adjectives to describe statistics from students with moderate effort

Adjective to describe statistics	Percentage
Analytical	5.26%
Applicable	0%
Biased	5.26%
Calculating	0%
Challenging	0%
Common	0%
Compelling	0%
Complex	15.79%
Complicated	10.53%
Conceptual	5.26%
Conditional	0%
Confusing	0%
Controversial	0%
Deceiving	5.26%
Deceptive	5.26%
Descriptive	0%
Different	0%
Difficult	10.53%
Dubious	0%
Doable	0%
Dynamic	0%
Easily manipulated	0%
Easy	5.26%
Engaging	0%
Effective	5.26%
Entertaining	0%
Essential	0%
Exciting	0%
False	5.26%
Fluctuating	5.26%
Formula related	0%
Fun	5.26%
Helpful	5.26%
Human	0%
Important	5.26%
In depth	5.26%
Inaccurate	5.26%
Incorrect	0%
Informative	5.26%
Insightful	10.53%

Interactive	0%
Interesting	15.79%
Interpretive	0%
Intricate	0%
Intriguing	0%
Logical	5.26%
Malleable	0%
Manipulated	15.79%
Measured	0%
Mathematical	0%
Methodical	5.26%
Meticulous	0%
Mind-altering	0%
Misleading	10.53%
Misunderstood	0%
Misused	5.26%
Necessary	5.26%
Numerical	5.26%
Numerical information	0%
Open mind	0%
Ordered	5.26%
Positive	0%
Persuasive	5.26%
Picky	0%
Practical	0%
Predictive	0%
Productive	0%
Quantitative data	0%
Questionable	0%
Rational	5.26%
Representative	0%
Revealing	0%
Rewarding	0%
Risky	5.26%
Sample	5.26%
Scholarly	0%
Shocking	5.26%
Subjective	10.53%
Supportive	0%
Surprising	0%
Team work	0%
Theoretical	0%

Thought provoking	5.26%
Time consuming	0%
Tricky	0%
Ubiquitous	10.53%
Unpredictable	5.26%
Useful	36.84%
Vague	5.26%
Vital	0%
Widespread	0%
Wordy	0%

Table 37: Adjectives to describe statistics from students with low effort

Adjective to describe statistics	Percentage
Analytical	16.67%
Applicable	0%
Biased	0%
Calculating	0%
Challenging	50.00%
Common	0%
Compelling	0%
Complex	16.67%
Complicated	0%
Conceptual	16.67%
Conditional	0%
Confusing	16.67%
Controversial	0%
Deceiving	0%
Deceptive	0%
Descriptive	0%
Different	0%
Difficult	0%
Dubious	0%
Doable	0%
Dynamic	0%
Easily manipulated	0%
Easy	0%
Engaging	0%
Effective	0%
Entertaining	16.67%
Essential	0%
Exciting	16.67%
False	0%
Fluctuating	0%
Formula related	0%
Fun	0%
Helpful	16.67%
Human	0%
Important	0%
In depth	0%
Inaccurate	0%
Incorrect	0%
Informative	16.67%
Insightful	16.67%

Interactive	0%
Interesting	50.00%
Interpretive	0%
Intricate	0%
Intriguing	0%
Logical	0%
Malleable	0%
Manipulated	0%
Measured	0%
Mathematical	0%
Methodical	0%
Meticulous	16.67%
Mind-altering	0%
Misleading	0%
Misunderstood	0%
Misused	0%
Necessary	0%
Numerical	0%
Numerical information	0%
Open mind	0%
Ordered	0%
Positive	0%
Persuasive	0%
Picky	0%
Practical	0%
Predictive	0%
Productive	0%
Quantitative data	0%
Questionable	0%
Rational	0%
Representative	0%
Revealing	0%
Rewarding	0%
Risky	0%
Sample	0%
Scholarly	0%
Shocking	0%
Subjective	0%
Supportive	0%
Surprising	0%
Team work	0%
Theoretical	0%

Thought provoking	0%
Time consuming	0%
Tricky	0%
Ubiquitous	0%
Unpredictable	0%
Useful	33.33%
Vague	0%
Vital	0%
Widespread	0%
Wordy	0%

REFERENCES

Adams, N.A. & Holcomb, W.R. (1986). Analysis of the relationship between anxiety about mathematics and performance. *Psychological Reports*, 59, 943-948.

Aiken, L.R. (1980). Attitudes measurement research. In D.A. Payne (Ed.), *Recent developments in affective measurement* (pp. 1-24). San Francisco: Josey-Bass.

Assessment Resource Tools for Improving Statistical Thinking Web site (September 25, 2006). Retrieved October 28, 2008, from ARTIST website: http://app.gen.umn.edu/artist/publications.html

Atkinson, J.W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review*, 64, 359-372.

Ausubel, D.P., Novak, J.D., & Hanesian, H. (1978). *Educational Psychology: A Cognitive View* (2nd Ed.). New York, NY: Holt, Rinehart, and Winston, Inc.

Auzmendi, E. (1991, April). Factors related to attitudes toward statistics: A study with a Spanish sample. Paper presented at the American Educational Research Association, Chicago, IL.

Baldi, B. & Moore, D.S. (2009). The chi-square test for goodness of fit. In *The Practice of Statistics in the Life Sciences,* New York, NY: W.H. Freeman and Company.

Barbieri, G.A. & Giacche′, P. (2006). The worth of data: the tale of an experience for promoting and improving statistical literacy. ICOTS-7, 1-6. Retrieved from www.redeabe.org.br/ICOTS7/Proceedings/PDFs/InvitedPapers/1A1_BARB.pdf

Benson, J. (1989). Structural components of statistical test anxiety in adults: An exploratory model. *Journal of Experimental Education*, 57(3), 247-261.

Bude′, L., Van de wiel, M.W.J., Imbos, T., Candel, M.J.J.M., Broers, N.J., & Berger, M.P.F. (2007). Students' achievement in a statistics course in relation to motivational aspects and study behaviour. *Statistics Education Research Journal*, 6(1), 5-21.

Callingham, R. & Watson, J.M. (2005). Measuring statistical literacy. *Journal of Applied Measurement*, 6, 19-47.

Cantor, A.B. (1996). Sample-size calculation for Cohen's kappa. *Psychological Methods*, 1, 150-153.

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

Cashin, S.E. & Elmore, P.B. (2005). The survey of attitudes toward statistics scale: A construct validity study. *Educational and Psychological Measurement*, 65, 509-524.

Cerrito, P.B. (1999). Teaching statistical literacy. *College Teaching*, 47(1), 9-18.

Cherian, V.I. & Glencross, M.J. (1997). Sex, socioeconomic status, and attitude toward applied statistics among postgraduate education students. *Journal of the Royal Anthropological Institute,* 3, 1385-1386.

Cochran, W.G. (1977). *Sampling Techniques*. New York , NY: John Wiley & Sons.

Cruise, R.J., Cash, R.W., & Bolton, D.L. (1985). Development and validation of an instrument to measure statistical anxiety. *American Statistical Association Proceedings of the Section on Statistics Education,* 92-97.

Delmas, R., Garfield, J., Ooms, A. & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6(2), 28-58.

Dempster, M. & McCorry, N.K. (2009). The role of previous experience and attitudes toward statistics in statistics assessment outcomes among undergraduate psychology students. *Journal of Statistics Education,* 17(2). Retrieved from www.amstat.org/publications/jse/v17n2/dempsterpdf.html

Eccles, J.S. & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task value and expectancy-related beliefs. *Personality and Social Psychology Bulletin,* 21, 215-225.

Elmore, P.B. & Vasu, E.S. (1980). Relationship between selected variables and statistics achievement: building a theoretical model. *Journal of Educational Psychology*, 72(4), 457-467.

Faghihi, F. & Rakow, E. (1995, November). The relationship of instructional methods with student responses to the survey of attitudes toward statistics. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Biloxi, MS.

Feinberg, L.B. & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. *Journal of Experimental Education*, 46, 11-18.

Fennema, E. & Sherman, J. (1977). Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American Educational Research Journal*, 14, 51-71.

Gal, I. (1997). Numeracy: reflections on imperatives of a forgotten goal. In L.A. Steen (Ed.), *Quantitative literacy* (pp. 36-44). Washington, DC: College Board.

Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review,* 70, 1-51.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking* (pp. 47-78). Dordrecht: Kluwer Academic Publishers.

Gal, I. & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, 2(2). Retrieved from www.amstat.org/publications/jse

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). Amsterdam, Holland: IOS Press.

Galagedera, D. (1998). Is remedial mathematics a real remedy? Evidence from learning statistics at tertiary level. *International Journal of Mathematics Education, Sciences, and Technology*, 29, 475-480.

Galagedera, D. Woodward, G., & Degamboda, S. (2000). An investigation of how perceptions of mathematics ability can affect elementary statistics performance. *International Journal of Mathematics Education, Sciences, and Technology,* 31, 679-689.

Galli, S., Ciancaleoni, M., Chiesi, F. & Primi, C. (2008) Who failed the introductory statistics examination? A study on a sample of psychology students. In *Proceedings of the 11th International Congress on Mathematical Education, Monterrey, Mexico,* Retrieved October 28, 2008 from http://tsg.icme11org/document/get/526

Gelman, A., Nolan, D., Men, A., Warmerdam, S. & Bautista, M. (1998). Student projects on statistical literacy and the media. *The American Statistician*, 52(2), 160-167.

George, D. & Mallory, P. (2003). *SPSS for Windows step by step: A single guide and reference.* 11th update (4th ed.) Boston: Allyn & Bacon.

Harlow, L.L., Burkholder, G.J. & Morrow, J.A. (2002). Evaluating attitudes, skill, ad performance in a learning-enhanced quantitative methods course: A structural modeling approach. *Structural Equation Modeling,* 9, 413-430.

Hill, H.H., Jr., Martinez-Dawson, R. & Bridges, W.C., Jr. (2006). Towards statistical literacy: statistics in everyday life. Abstract. Joint Statistical Meetings.

Hilton, S.C., Schau, C. & Olsen, J.A. (2004). Survey of attitudes toward statistics: factor structure invariance by gender and by administration time. *Structural Equation Modeling*, 11(1), 92-109.

Hoehler, F.K. (1999). Bias and prevalence effects on kappa viewed in terms of sensitivity and specificity. *Journal of Clinical Epidemiology,* 53, 499-503.

Johnson, M. & Kuennen, E. (2006). Basic math skills and performance in an introductory statistics course. *Journal of Statistics Education,* 14(2). Retrieved from www.amstat.org/publications/jse/v12n2/johnson.html

Lalonde, R.N. & Gardner, R.C. (1993). Statistics as a second languate? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science*, 25, 108-125.

Landis, J.R. & Koch, G.G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159-174.

Lane, M.J. (1990). Women and minorities in science and engineering, National Science Foundation Report 90-301, p. 8.

Levin, J.R. & Serlin, R.C. (2000). Changing students' perceptions of McNemar's test of change. *Journal of Statistics Education,* 8(2). Retrieved from http://www.amstat.org/PUBLICATIONS/JSE/secure/v8n2/levin.cfm

Luke, A. & Freebody, P. (1997). Shaping the social practices of reading. In S. Musprati, A. Luke & P. Freebody (Eds.), *Constructing critical literacies: Teaching and learning textual practice* (pp. 185-225). St. Leonards, NSW: Allen and Unwin.

Maehr, M. (1984). Meaning and motivation: Toward a theory of personal investment. In R. Ames & C. Ames (Eds.), *Research on motivation in education, Vol 1: Student motivation* (pp. 115-144). Orlando, FL: Academic.

McCall, C.H., Belli, G., & Madjidi, F. (1990). The complexities of teaching graduate students in educational administration introductory statistical concepts. Unpublished report.

Merriman, L. (2006). Using media reports to develop statistical literacy in year 10 students. Retrieved from www.stat.auckland.ac.nz/~iase/publications/17/8A3_MERR.pdf

Moore, D. (2001). Statistical literacy and statistical competence in the new century. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/7/Moore.pdf

Mueller, D.J. (1986). *Measuring Social Attitudes: A Handbook for Researchers and Practitioners.* New York, NY: Teachers College Press.

Nasser, F. (1999). Prediction of statistics achievement. *Proceedings of the International Statistical Institute 52nd Conference,* Helsinki, Finland, 3, 7-8.

Nasser, F. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education,* 12(1) Retrieved from www.amstat.org/publications/jse/v12n1/nasser.html

Olson, J.M. & Zanna, M.P. (1993). Attitudes and attitude change. *Annual Review of Psychology,* 44, 117-154.

Pugh, C.M. & Lowther, S. (2004). College math performance and last high school math course. Paper presented at the Annual Conference of the Southern Association for Institutional Research, Biloxi MS. Retrieved from http://oira.auburn.edu/about/publications/SAIRPaper2004.pdf

Reston, E.D. (2005). Assessing statistical literacy in graduate level statistics education. Paper presented at the 55th Session of the International Statistical Institute, Sydney, Australia. Retrieved from www.stat.auckland.ac.nz/~iase/publications.php?show=13

Roberts, D.M. & Bilderback, E.W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement*, 40, 235-238.

Roberts, D.M. & Saxe, J.E. (1982). Validity of a statistics attitude survey: A follow-up study. *Educational and Psychological Measurement*, 42, 907-912.

Rumsey, D.J. (2002) Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*, 10(3) Retrieved from www.amstat.org/publications/jse/v10n3/rumsey2.html

Sanchez, J. (2007). Building statistical literacy assessment tools with the IASE/ISLP. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/sat07/Sanchez.pdf

Schau, C. (2003). Students' attitudes: The "other" important outcome in statistics education. Retrieved from http://evaluationandstatistics.com/JSM2003.pdf

Schau, C. Scoring the SATS-36© Retrieved from www.evaluationandstatistics.com/Final36scoring.pdf

Schau, C.G., Dauphinee, T. & Del Vecchio, A. (1992, April). The development of the Survey of Attitudes Toward Statistics. Paper presented at the meeting of the American Educational Research Association, San Francisco, CA.

Schau, C., Stevens, J., Dauphinee, T.L., & Del Vecchio, A. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, 55(5), 868-875.

Schield, M. "Statistical literacy: Thinking critically about statistics" Retrieved from http://www.augsburg.edu/ppages/schield

Schield, M. (2004). Statistical literacy curriculum design. *Curricular Development in Statistics Education*, 54-74.

Schield, M. (2006). Statistical literacy survey results: Reading graphs and tables of rates and percentages. Retrieved from www.augsburg.edu/statlit/pdf/2006SchieldIASSIST5b.doc

Schram, C.M. (1996). A meta-analysis of gender differences in applied statistics achievement. *Journal of Educational and Behavioral Statistics,* 21, 55-70.

Shaughnessy, M.J. (1992). Research in probability and statistics: Reflections and directions. In D.A. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning* (pp. 465-494). New York, NY: Macmillan.

Sim, J. & Wright, C.C. (2005). The kappa statistic in reliability studies: Use, interpretation, and sample size requirements. *Physical Therapy*, 85(3), 257-268.

Simon, J.L. & Bruce, P. (1991). Resampling: A tool for everyday statistical work. *Chance*, 4, 22-32.

Snell, L. (1999). Using *Chance* media to promote statistical literacy. Paper presented at the 1999 Joint Statistical Meetings, Dallas, TX.

Statistical Literacy Assessment Scale (SLAS) Retrieved October 28, 2008, from http://course1.winona.edu/cblumberg/reston.pdf

Statistics Canada and Organisation for Economic Cooperation and Development (OECD) (1996). *Literacy, economy, and society: First results from the International Adult Literacy Survey*, Ottawa.

Stephens, L.J. (1982). Mathematical competency, as a predictor of success in college statistics courses. *International Journal of Mathematical Education in Science and Technology*, 13(4), 423-425.

Sutarso, T. (1992, November). Students' attitudes toward statistics (STATS). Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Knoxville, TN.

Tempelaar, D.T., Van Der Loeff, S.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.

Tognolini, J. (1996). Rasch modeling: Advantages and limitations. In *Session notes National Meeting on Assessment and Reporting* 25-26 November 1996. Manly: NSW Department of School Education.

Tomazic, T.J. & Katz, B.M. (1988). Statistical anxiety in introductory applied statistics. *American Statistical Association Proceedings of the Section on Statistical Education,* 239-244.

Townsend, M. (2006). Developing statistical literacy in youth: statistics Canada's education outreach program. Retrieved from www.statlit.org/PDF/2006TownsendStatLitInYouth.pdf

Utts, J. (2002). What educated citizens should know about statistics and probability. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/1/1b3_utts.pdf

Wallman, K.K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association*, 88, 1-8.

Waters, L.K., Martelli, T.A., Zakrajsek, T., & Popovich, P.M. (1988). Attitudes toward statistics: an evaluation of multiple measures. *Educational and Psychological Measurement*, 48, 513-516.

Watson, J.M. (1997). Assessing statistical thinking using the media. In I.Gal and J.Garfield (Eds.), *The Assessment Challenge in Statistics Education* (pp. 107-121). Amsterdam, The Netherlands: IOS Press.

Watson, J.M. (2002). Statistical literacy before adulthood. *International Statistical Review*, 70(1), 26-30.

Watson, J.M. (2006). Statistical literacy- A global goal. In A.H. Schoenfeld (Ed.), *Statistical Literacy at School: Growths and Goals* (pp. 247-272). Mahwah, New Jersey: Lawrence Erlbaum Associates.

Watson, J.M. & Callingham, R.A. (2003). Statistical literacy: a complex hierarchical construct. *Statistics Education Research Journal,* 2(2), 3-46.

Watson, J.M. & Kelly, B.A. (2003). The vocabulary of statistical literacy. In *Educational Research, Risks, & Dilemmas: Proceedings of the joint conference of the New Zealand Association for Research in Education and the Australian Association for Research in Education,* Auckland, New Zealand. Retrieved from http://aare.edu.au/03pap/wat03297.pdf

Watson, J.M. & Moritz, J.B. (2000). Development of understanding of sampling for statistical literacy. *Journal of Mathematical Behavior*, 19, 109-136.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology,* 71, 3-25.

Weisstein, E.W. *Bonferroni Correction*. Retrieved on February 4, 2010, from MathWorld A Wolfram Web Resource: http://mathworld.wolfram.com/BonferroniCorrection.html

Wild, C. & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review,* 67(3), 221-248.

Wise, S.L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, 45, 401-405.

Wisenbaker, J.M. & Scott, J.S. (1997, April). *Modeling aspects of students' attitudes and achievement in introductory statistics courses.* Paper presented at the meeting of the American Educational Research Association, Chicago, IL.

Wisenbaker, J.M., Scott, J.S. & Nasser, F. (2000). *Structural equation models relating attitudes about and achievement in introductory statistics courses: A comparison of results from the U.S. and Israel.* Paper presented at the Ninth International Congress on Mathematics Education, Tokyo, Japan.

Wooten, C.T. (1998). Factors influencing student learning in introductory accounting classes: A comparison of traditional and nontraditional students. *Issues in Accounting Education,* 13, 357-373.

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: some interesting parallels. *British Journal of Educational Psychology*, 61, 319-328.