

11-5-2018

An Instructional Strategy with Simulations Used to Increase Statistical Literacy among Students in a Hispanic Serving Institution

Eric O. Hernandez
Florida International University, ehern012@fiu.edu

Follow this and additional works at: <https://digitalcommons.fiu.edu/etd>



Part of the [Curriculum and Instruction Commons](#)

Recommended Citation

Hernandez, Eric O., "An Instructional Strategy with Simulations Used to Increase Statistical Literacy among Students in a Hispanic Serving Institution" (2018). *FIU Electronic Theses and Dissertations*. 3849.
<https://digitalcommons.fiu.edu/etd/3849>

This work is brought to you for free and open access by the University Graduate School at FIU Digital Commons. It has been accepted for inclusion in FIU Electronic Theses and Dissertations by an authorized administrator of FIU Digital Commons. For more information, please contact dcc@fiu.edu.

FLORIDA INTERNATIONAL UNIVERSITY

Miami, Florida

AN INSTRUCTIONAL STRATEGY WITH SIMULATIONS USED TO INCREASE
STATISTICAL LITERACY AMONG STUDENTS IN A HISPANIC SERVING
INSTITUTION

A dissertation submitted in partial fulfillment of the

requirements for the degree of

DOCTOR OF PHILOSOPHY

in

CURRICULUM AND INSTRUCTION

by

Eric O. Hernandez

2018

To: Dean Michael R. Heithaus
College of Arts, Sciences, and Education

This dissertation, written by Eric O. Hernandez and entitled *An Instructional Strategy with Simulations Used to Increase Statistical Literacy among Students in a Hispanic Serving Institutions*, having been approved in respect to style and intellectual content, is referred to you for judgment.

We have read this dissertation and recommend that it be approved.

Barbara King

B.M Golam Kibria

Leonard Bliss

Maria L. Fernandez, Major Professor

Date of Defense: November 5, 2018

The dissertation of Eric O. Hernandez is approved.

Dean Michael R. Heithaus
College of Arts, Sciences, and Education

Andrés G. Gil
Vice President for Research and Economic Development
and Dean of the University Graduate School

Florida International University, 2018

© Copyright 2018 by Eric O. Hernandez

All rights reserved.

DEDICATION

I dedicate this doctoral dissertation to my father Erih and my mother Marisabel for planting the seeds of education. I also dedicate this work to my grandfather Oscar and my uncle Vladimir from who I have learned the strength, discipline, dedication, and resilience to overcome obstacles in life.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to the committee. Thank you for providing me with the guidance I needed for the completion of this project. I am grateful to you for your consistent commitment and availability to clarify questions and assist me when I needed it. Your inspiration, motivation, strength, and persistence played a crucial role to finish this goal.

ABSTRACT OF THE DISSERTATION
AN INSTRUCTIONAL STRATEGY WITH SIMULATIONS USED TO INCREASE
STATISTICAL LITERACY AMONG STUDENTS IN A HISPANIC SERVING
INSTITUTION

by

Eric O Hernandez

Florida International University, 2018

Miami, Florida

Professor Maria L. Fernandez, Major Professor

This study analyzed the effects of a randomization-based inference teaching methodology on students' content mastery in an introductory statistics college course. The sample was 125 undergraduate students from Miami Dade College, a large Hispanic Serving Institution in the Southeast. A pretest-posttest nonequivalent group design was used for the study. Students in the randomization-based teaching modality received exposure to simulation activities, specifically bootstrap confidence intervals and randomization test, that aim to enhance conceptual understanding of inferential statistics, an important component of statistical literacy. The instructional strategy was designed to trigger critical reflection that confronted students with their thinking and lead them through a process of reorganization, restructure, and improvement of their concepts. The 40 item Comprehensive Assessment of Outcomes in a first Statistics course (CAOS) instrument was used to measure students' conceptual understanding of important statistical ideas along with a demographic and academic survey that collected data on student characteristics.

A stepwise linear regression method was used to look at the effects of group membership while controlling for Pre-CAOS scores, age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, Hispanic, highest math course taken in high school, and GPA. The full model showed that only Group and Pre-CAOS score were the only significant predictors of Post CAOS scores. None of the other variables were significant. The model was a significant predictor of Post CAOS score, $F(2, 121) = 16.96, p < .001$. The coefficient of determination, R^2 , showed the model accounted for 22.2% of the variance in the Post CAOS scores. For group membership, simulation-based teaching methodology had higher scores than theory-based teaching methodology, even when controlling for Pre CAOS test score, $B=11.22, t(123)= 4.70, p < .001$. For every one unit increase in pre CAOS test score, the post CAOS score increased by .46 units, $B=0.46, t(123)= 3.07, p = .003$.

The results supported the claim that the randomization-based teaching modality for inferential statistics help Hispanic students to achieve a better understanding of the learning outcomes associated with an undergraduate introductory statistics course.

TABLE OF CONTENTS

CHAPTER	PAGE
CHAPTER 1: INTRODUCTION	1
Background of the Problem	1
Statement of the Problem.....	18
Purpose of the Study	19
Research Questions.....	20
Nature of the Study	20
Contributions to the Field	22
Definitions of Terms	23
Summary	25
CHAPTER 2 LITERATURE REVIEW	27
Introduction.....	27
Statistical Literacy	28
Teaching Statistical Literacy.....	35
Barriers to Statistical Literacy	47
Student Characteristics as Obstacles to Statistical Literacy	54
Best Practices for Statistical Literacy	63
The Gaps in the Research	69
This Project.....	70
Summary	74
CHAPTER 3: METHODOLOGY	75
Methods and Research Design.....	75
Population and Sampling	77
Instruments.....	81
Instructional Strategy	85
Data Analysis	91
Summary	95
CHAPTER 4: RESULTS.....	96
Data Cleaning.....	97
Scale Creation	98
Descriptive Statistics.....	99
Race, Language and Highest Math Taken in HS Recoding.....	102
Prior-Statistical Experience	104
Research Question 1	108
Research Question 2	112
Research Question 3	114
CHAPTER 5 DISCUSSION.....	120

Summary of the Findings.....	120
Implications.....	126
Limitations of the Study.....	129
Recommendations.....	131
Conclusion	133
REFERENCES	135
APPENDICES	150
VITA	183

LIST OF TABLES

TABLE	PAGE
Table 1 Technology for Randomization-Based Methods	46
Table 2 Curricula Design	73
Table 3 CAOS Item Numbers by Topic	82
Table 4 Traditional Curricula – Theory-Based Inference Expected Time: Spring 2018..	87
Table 5 Simulation-Based Group Expected Time: Spring 2018	89
Table 6 Descriptive Statistics for Continuous Variables	99
Table 7 Descriptive Statistics for Categorical Variables by Class	100
Table 8 Frequency of Hispanic and Highest Math Taken in HS	103
Table 9 Descriptive Statistics for Hispanic and Highest Math Taken by Class	104
Table 10 Frequency of Prior Statistical Experience by Class.....	105
Table 11 CAOS Score Descriptive Statistics.....	106
Table 12 CAOS Pre- and Post-Test Descriptive Statistics by Class.....	107
Table 13 Regression of Group, Pre-CAOS test, on Post CAOS Percent.....	112
Table 14 Pearson Correlations	114
Table 15 Stepwise Regression of Group, Pre CAOS test, Gender, Language and Control Variables on Post CAOS.....	119
Table E1 Full Table Regression for Regression 1	180
Table E2 Full Table for Stepwise Regression.....	181
Table E3 Full Table for Enter Method Regression.....	182

LIST OF FIGURES

FIGURE	PAGE
Figure 1: Residuals for the Regression of Predictors on Total CAOS scores	109
Figure 2: P-Plot of Post CAOS Residuals	111
Figure 3: Residuals for the Regression of Predictors on total Post CAOS scores.....	116
Figure 4: P-Plot of the Residuals	117
Figure 5: Histogram of the Residuals	118
Figure E1: Histogram of Pre-Total CAOS Percent Score	178
Figure E2: Histogram of Post-Total CAOS Percent Score.....	179

CHAPTER 1: INTRODUCTION

In introductory statistics courses, statistical literacy is an established critical learning outcome (Garfield, delMas, & Zieffler, 2010). Statistical literacy is key to applying statistics beyond the classroom and pivotal to many career fields. One development in statistics education has been advances in the pedagogical approach to introductory statistics courses. The innovation embraces moving away from procedural knowledge to helping students' foster conceptual knowledge. But the majority of research on statistical literacy and teaching statistics has focused primarily on White students at traditional colleges and universities. While this knowledge is important, Hispanic students represent one of the fastest growing populations in higher education (Krogstad, 2016). There is a dearth of research about statistical literacy in Hispanic undergraduate students, even more so among Hispanic students attending Hispanic Serving Institutions. Thus, the present dissertation looked at whether changes in pedagogy in teaching statistics increased the statistical literacy among students attending Hispanic Serving Institutions.

Background of the Problem

The current job market desires human capital competent in comprehending and manipulating data to make compelling choices (Cobb, 2015). Such decisions impact society at large, making it the responsibility of higher education institutions to promote and encourage the right set of heuristics associated with good judgment, especially under

uncertainty. Individuals would be more proficient in their disciplines if they comprehend at an early stage how to make trustworthy statistical inferences. Individuals, in return, can yield the right motivation to advance their academic participation in scientific fields (Garfield & Ben-Zvi, 2005). Given the critical aspect statistics education plays in fulfilling the rising job market, demonstrating how early use of dynamic visualization tools can boost statistical literacy is important, as this is the pedagogy trending in the field (Maurer & Lock, 2016; Chance, Wong, & Tintle, 2016; Lane-Getaz, 2017; Ridgway & Nicholson, 2017).

Herk (2016) noted that the National Federation of Independent Business reported that 47% of small businesses failed to fill positions because they could not find qualified applicants. The main reason was the skill gaps among applicants, and statistical skills was listed just after teamwork and communication in importance (Herk, 2016). Furthermore, the non-STEM careers that need statistical literacy, like psychology, are expected to grow 19% by 2024, much faster than the average of 7% in other fields, according to the Bureau of Labor Statistics (2016). Despite a great effort by higher education institutions to promote and invest in STEM careers, including Hispanic Serving Institutions, the majority of undergraduate students enroll in non-STEM careers, and their only contact point between quantitative literacy skills and their workplace has been their statistics course. Therefore, statistical literacy is crucial in the academic formation of undergraduate students and the ability to promote efficient pedagogical approaches should not be neglected.

There is a plethora of research that reflects the concerns about the absence of statistical literacy in the social sciences (Carter, Brown, & Simpson, 2017; Ridgway &

Nicholson, 2017; Lane-Getaz, 2017). The research argues that both students and teachers need to be empowered to use statistics, and one way is to tell sociological stories using data (Scott Jones & Goldring, 2017). However, the ability to unlock data to tell a story goes in hand with the adoption of simulation-based methods to address the alignment problem between the intended curriculum, the implemented curriculum, and the attained curriculum (Budgett & Wild, 2014; Chance et al., 2016; Ridgway & Nicholson, 2017). The vast majority of effort has been centered in mapping out the intended curriculum while attained curriculum has just recently begun to gain attention.

The term ‘statistical literacy’ had been defined as “the ability to understand and critically evaluate statistical results that permeate our daily life, coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions” (Wallman, 1993, p. 1). Researchers and educators know that statistical literacy is important, and yet, statistical literacy rates, as measured by the Comprehensive Assessment of Outcomes in Statistics (CAOS), have remained around 50% (delMas, 2014). The CAOS test measures a student’s understanding of statistics concepts covered in an undergraduate introductory statistics course, with satisfactory internal Cronbach’s alpha reliability coefficient of 0.78 (delMas, 2014). The gap between the curriculum and actual statistical literacy has led researchers to ask why it is so difficult to teach students to think statistically.

People tend to evaluate situations by the ease with which they are retrieved from memory. This mental process is fairly useful, but it can lead to severe systematic errors. Mathematics has played an important role in shaping scientific reasoning. It has worked as a process to provide systematic explanations with strong foundations (Kahneman,

2013). When students are reflecting upon the question ‘why?’ there are different pathways that an answer can take, but only probabilistic explanations can successfully answer the question when deductive reasoning fails because the premises do not apply to the context of the problem (Kahneman, 2013).

Furthermore, people tend to default to intuitive heuristics when the question is challenging and a skilled explanation is not available (Tversky & Kahneman, 1974). Statistics and other STEM fields as well, have offered a process to answer these questions logically. The developing of a successful heuristic is the reason why statistics education is a fundamental pillar in the curriculum of every discipline (Ridgway & Nicholson, 2017); thus, the attainment of statistical literacy becomes paramount to statistics educators. Statistics instructors should become aware of the common misconceptions that plague students in order to address them effectively.

One concern among statistics educators is students’ strong bias towards having confidence that small samples closely duplicate the population from which they are drawn (Reaburn, 2014). Students do not understand why variance matters. Rather it seems to be perceived as a nuisance; however, it is important to understand and recognize sampling variability because significant decisions, like the effectiveness of a treatment, depend on measurements of variability through probability theory. Additionally, there is widespread misunderstanding of randomness, which can have severe consequences— for example, when generalizing to a broader population when most null hypothesis significant test studies do not provide evidence that validate or falsify a hypothesis (Lane-Getaz, 2017). People are pattern seekers and, for the most part, believe in a coherent world in which commonalities occur not by accident but as the product of mechanical

causality or of someone's intention (Kahneman, 2013). The associative brain machinery seeks causes; however, statistical thinking focuses on what could have happened within a world of possible outcomes. Therefore, if nothing in particular caused the event, chance is selected as an answer. But the illusion of patterns can affect our reasoning. Even after formal training, the affinity to see patterns in randomness is overwhelming, as is evident by the results of standardized statistics assessments in an undergraduate introductory course (delMas, 2014; Kahneman, 2013). For this reason, it is pivotal to understand the impact of simulation activities—namely, bootstrapping confidence interval and randomization tests—on students' cognitive effect in developing statistical literacy.

Given that the illusion of patterns affects our reasoning, statistics instructors should explore pedagogical alternatives (Aquilonius & Brenner, 2015) to remove the anchoring effects of people's perceptions of a causal world because it constrains students from becoming statistically literate (Kahneman, 2013). The current methodology used to teach statistics depends on mathematical theorems about distributions (Tintle, VanderStoep, Holmes, Quisenberry, & Swanson, 2011). In general, statistics has been perceived by students as an instrument to generate hypothesis employing statistical techniques to frame the uncertainty of the world in which we live. Limiting inferential reasoning to p-value judgments minimized the complexities of statistical inferences. Consequently, instructors must teach students how to move from probability theory to evidence while considering all possible scenarios of data representation (Ben-Zvi & Arcavi, 2001). Still, students find this shift challenging, especially the language used to produce the reasoning about distributions (Aquilonius & Brenner, 2015). Therefore, the reasoning from probability theory to evidence leaves students with a poor understanding

of statistical inference and some may even end up being averse to future statistical knowledge. Consequently, it is important to address the root problem with teaching and learning statistics across all disciplines and explore alternative pedagogies from those traditionally used to teach introductory level statistics classes.

The challenges in teaching and learning inferential statistics include the lack of basic mathematical understanding and proper prerequisites for the course, leaving students lacking the required skills to understand the nature of the stochastic process taught under the traditional method (Lane-Getaz, 2017; Reaburn, 2014). In addition, the current teaching approach had been behavioral which focuses the attention on students' procedural knowledge that sometimes impairs students' ability to develop the statistical reasoning in order to solve the problem in a particular context (Hassad, 2011).

Additionally, there is an inability to communicate statistically, especially in making sense of the decisions from reports and studies in the media or in the context of work (Ridgway & Nicholson, 2017). Likewise, there is an incorrect representation of the aspects related to uncertainty on behalf of the media, and the world at large, because of the lack of training in statistics (Kahneman, 2013). Finally, there is a lack of professionals that understand both the pedagogical and cognitive process required for optimal teaching of an undergraduate introductory course in statistics (Haines, 2015; Hassad, 2011). The majority of educators teach the subject using the textbooks that lean toward the mechanical aspects and recitation of statistical formulas rather than reflecting the best sequence of learning that should be used in order to understand the underlying concepts (Haines, 2015). In light of emergent educational technology that helps understand the

statistical reasoning with the use of real data, statistics educators should investigate the effects of such technology with undergraduate students.

Teaching statistics in alternative ways could provide an easy and fluent retrieval for the judgment and decision-making process. Dynamic visualization statistical educational tools such as StatKey, S-Plus, ThinkerPlots, and applets have improved the ability to impart statistical knowledge (Budgett, Pfannkuch, Regan, & Wild, 2013) . These tools have broadened the competencies and expected learning outcomes of introductory statistics courses (Chance et al., 2016; Garfield & Ben-Zvi, 2008; Maurer & Lock, 2016). Furthermore, statistics education exposes students with the heuristic associated with the possibility of establish generalizations that can either be validated or debunked (Reaburn, 2014).

The use of educational software in the statistics classroom has promoted different cognitive abilities that require mathematical and verbal reasoning by recognizing the omnipresence of uncertainty (Aquilonius & Brenner, 2015; Budgett, Pfannkuch, Regan, & Wild, 2013; Budgett & Wild, 2014). As a result, students understood that the sample provides information that enhances their knowledge about the population. As a result, the cognitive mapping opens the portal for two complementary ideas: sample representativeness and sample variability, which promotes inferential reasoning. Thus, the process of thinking requires that the student successfully coordinate the sample, the space where the sample was taken, and all possible samples from such space.

In the light of dynamic visualization tools, statistics educators have the ability to stimulate the understanding of concepts through graphical representations that were left before to the formal mathematical approach (Pfannkuch & Budgett, 2014). Introducing

statistical concepts with the aid of simulation tools, followed up with the formal mathematical system that supports such reasoning, enables the instructor to trigger cognitive conflict in students' mental schema with the appropriate pedagogical approach (Aquilonius & Brenner, 2015). According to Festinger (1957), cognitive dissonance "is the existence of unfitting relations among cognitions (p. 3). Cognition is "any knowledge, opinion, or belief about the environment, about oneself, or about one's behavior" (Festinger, 1957, p. 3). An examination of knowledge, opinions, or belief arising from the mind is considered a cognitive process (Tuan & Jay, 2016). Thus, dissonance originates from logical inconsistency among knowledge, opinions, or belief.

The hypothesis underpinning the dissonance theory affirmed that the presence of dissonance, being psychologically distressed, "will motivate the person to try to reduce the dissonance and achieve consonance where beliefs, opinions or knowledge are consistent" (Festinger, 1957, p. 3). The cognitive dissonance pedagogical framework has the potential to increase statistical literacy because the mind is not supposed to display conflicting knowledge (Tuan & Jay, 2016). Harmony needs to be restored, conforming to the concepts of inferential statistics. Once some cognitive challenge has been created with the randomization-based methods followed up by the traditional-based methods, the instructor helps the students recreate equilibrium, and learning takes place.

While there is a dearth of research under this theoretical framework, there is currently no research that explores the effect of cognitive dissonance approach on undergraduate students who attend a Minority Serving Institution. The aim of the present study was to determine whether the application of cognitive conflict in the learning of

inferential statistics led to the enhancement of student's conceptual knowledge about statistical literacy.

Simulation-based inference software tools have redefined the way in which statistics instructors can teach an undergraduate introductory course (Budgett et al., 2013; Pfannkuch & Budgett, 2014). These tools stimulate the progress of statistical thinking by assisting students' understanding of the association between sampling and the uncertainty of an event. In a simulation environment, students are able to change parameters of the simulation in order to generate changes in numerical and graphical indicators associated with stochastic distributions (Budgett et al., 2013; Budgett & Wild, 2014; Pfannkuch & Budgett, 2014). The interaction helps students to anchor their judgment of the probability of an outcome on a plausible experience and provides a stimulus to encourage inquiry about the diagnostic of the evidence (Khazanov & Prado, 2010; Tintle et al., 2014)

The new way of teaching statistics is grounded under the logic of inference, which can be accomplished with randomization methods (Tintle, Topliff, VanderStoep, Homes, & Swanson, 2012). Such approaches operate under the premise that the definite distribution of any test statistics specified under a proposition about a population parameter can be achieved by reshuffling the original data several times. The premise stands in opposition to the underpinning logic of normal distributions, which gravitate towards an exact solution under the condition that the sample size grows to the infinite. The teaching approach can shorten the gap between statistical education and statistical practice since not only are students able to experience the multiple mental images through the dynamic software (Budgett et al., 2013), but they can also think more about the different taxonomy levels of inferential statistics (Lane-Getaz, 2017).

The randomization-based inference methodology, also known as simulation-based inference, can provide the instructor with the opportunity to emphasize the following ideas: the p-value as an integral part of a larger statistical process, the concept of the distribution under the null hypothesis, the variance within groups and between groups, confidence intervals and two-sided tests to assess statistical significance, and the differentiation of sample statistics from population parameters (Lane-Getaz, 2017). These ideas are pillars in an undergraduate introductory statistics course because they are the basis for deciding on the basis of an observed sample in order to make a judgment about the event investigated. If these ideas are highlighted appropriately, they can overturn many of the documented misconceptions and misinterpretations associated with the null hypothesis significant test (Aquilonius & Brenner, 2015; Cumming, 2010; Reaburn, 2014) and increase statistical literacy among college students.

What Statistical Literacy Should Feature

The capability of an individual to debate concerns about conclusions and resolutions from the data analysis is what statistical literacy should feature. In addition, the individual should be able to communicate statistical information in a robust way (Garfield, 2002). This section discusses what an ideal statistically literate person must know and be able to express. In our current world, the ability to communicate statistical findings is critical to many fields. For this reason, several statistics educational groups and organizations advocate that students must comprehend basic statistics in order to detect any perverted use of data encountered in broadcast and social media (Ridgway & Nicholson, 2017).

Statistical literacy should enable students to synchronize three mechanisms (Garfield & Ben-Zvi, 2008): the analysis of graphical distributions, the evaluation of statistical test results, and the ability to carry out decisions in context. Ideally, students must be able to recognize the influence of sampling on the process of generating a distribution and on the outcome of performing asymptotic tests. Inferential analysis commands students to understand how the measures of central tendency and variability impact the interpretation of statistical outputs. Furthermore, students should be able to state conclusions that support a graphical representation and outcomes in the context of the research. Individuals with a dependable foundation must be able to move from outcomes and graphical representations to solid judgment about the information encoded in the data (Ben-Zvi, 2004; Saldanha & Thompson, 2003). This shift also requires being able to decode the details into meaningful conclusions, knowing that core statistical ideas, such as confounding and causality, should receive more attention in the conclusion (Garfield & Ben-Zvi, 2008; Nicholson & Ridgway, 2017).

At the same time, students must be able to comprehend that a p-value has been calculated from the desire to estimate the likelihood that a sample was drawn from a population with a specified parameter value (Reaburn, 2014). Such p-values produce a description on how samples should behave randomly (Reaburn, 2014). The concept that allows the systematic coordination between information from samples and the ability to measure uncertainty is variability. However, the nationally normed results from the CAOS test reveal that classes had not considerably enhanced students' understanding of hypothesis testing (delMas, 2014). According to delMas (2014), around half of the students recognized a correct interpretation of a p-value, thereby emphasizing "that

students show little improvement after an introductory undergraduate statistics course” (delMas, 2014, p. 5).

Students should be able to recognize statistically significant outputs because the data presents a small within-group variability and a large between-group variability where the model revealed the behavior of the data. Variability impacts the central tendency indicators. Statistically literate students should understand that the presence of extreme values or unusual distributions of data should caution them about the rigor of the conclusions derived from the analysis. Furthermore, students should be able to recognize that residuals represent how imprecise the model is. Consequently, the goal is to decipher variation by looking for the systematic effects of randomness and their influence in the results of asymptotic tests. Comprehending variability provides students with the leading mechanism to successfully perform in statistics (Garfield, 2002; Garfield & Ben-Zvi, 2005).

The Current State of Statistical Literacy

Despite the effort of the Consortium for the Advancement of Undergraduate Statistics Education (CAUSE) and the American Statistical Association (ASA), through their Guidelines for Assessment and Instruction in Statistics Education (GAISE), the levels of statistical literacy among undergraduate students fell short as reported by delMas (2014) in the evaluation of nine years of data from the CAOS test. Data from the College Board’s Advanced Placement (AP) Statistics courses revealed that student enrollment increased from 8,000 registered exams in 1997 to nearly 200,000 exams in 2015 (Haines, 2015). However, the number of capable instructors to incorporate new

pedagogical approaches has not kept the pace, compounding the low results in the AP scores (Haines, 2015).

Like the CAOS test, the AP test was designed to measure statistical literacy and a working knowledge of statistics that can be applied to a particular context (Haines, 2015). The 2017 results revealed that only 28.8 % of the national sample were well qualified, in contrast with 46.2 % who were not qualified and the remaining 25 % just scored acceptable (The College Board, 2017). Furthermore, AP Calculus AB and AP Calculus BC have considerably better scores for well-qualified students with 36.7 % and 60.7% respectively (The College Board, 2017). Therefore, both instruments (CAOS and AP Statistics) exposed the low level of statistical literacy and exploring the effects of a randomized inference methodology in an introductory statistics course should not be neglected from the literature, even more so in Minority Serving Institutions.

The current state of statistical literacy has permeated into the professional environment of the social sciences where professionals in the field do not identify quantitative work as a fundamental aspect of their discipline (Lane-Getaz, 2017). Undoubtedly, there is more to address in statistical literacy in higher education institutions, but there is no magic bullet to address the gap of statistical literacy. However, because it is a pre-requisite for an informed democracy, increasing this literacy is a key element in securing the health of our nation (Ridgway & Nicholson, 2017).

Conceptual change as a technique to increase statistical literacy. Conceptual change is fundamental for serious learning, and cognitive conflict is a well-thought-out premise for conceptual change (Liu, 2010). Cognitive conflict takes place when an individual cannot apply their existing concepts to solve a problem and is challenged with

a situation that prompts the learning of new concepts (Limon, 2001). Cognitive conflict strategies are used in many practical applications to attain conceptual change (Limon, 2001). As reported in the literature, applications of cognitive conflict strategies for attaining conceptual change can be characterized by the following key elements: making students aware of their existing concepts before the instructional intervention, confronting them with contradictory information, and measuring the resulting conceptual change (Limon, 2001).

The cognitive conflict approach has been used to advance a learning model with simulation based-inference (Liu, 2010). Liu (2010) established four phases in order to attain conceptual change: externalization, reflection, construction, and application. In her research, statistical simulations were the main learning tool in the reflection and construction phases and in part of the application phase. According to Liu (2010) in the externalization phase, students were led by the instructors in order to make them aware of their preconceptions before the instructional intervention. Hence, the purpose of this stage is to enable students to be aware of their own ideas about the statistics concepts they are learning.

In the reflection phase, Liu (2010) maneuver students to reflect on their own ideas that were externalized in the previous phase. To achieve such reflection, she exhibited a series of questions during the statistical simulations related to the concepts cultivated in the externalization stage in order to guide students to manipulate the simulations and to observe the results gradually. The students critically reflect on their ideas during the externalization and after the simulations (Liu, 2010). In the construction phase, Liu (2010) introduced exploratory activities to help students construct their concepts about

correlation. Students who did not understand the concepts after following specific activities were asked to repeat the entire construction phase. Finally, in the application phase, Liu (2010) provided students with two situations in different contexts and with different solution paths. The goal was to allow students to develop their newly constructed concepts by applying them to solve novel problems.

It is noteworthy that Liu (2010) evidenced cognitive conflict, using a case study approach, which was congruent to exploring students' learning processes in learning with technology (Liu, 2010). In Liu's study, interviews were conducted; therefore, it cannot be guaranteed that cognitive conflict occurred. However, Liu's research was used as a stepping stone to map the perspective on students' learning using statistical simulations. In addition, it was not necessary to measure cognitive conflict in order to study statistical literacy.

The conceptual change model was used in the current research, as it has been shown to be effective in teaching new math concepts. The statistical simulations used in this study were bootstrapping for statistical estimation of parameters and randomization test for statistical significance. A randomization test involves simulating what types of statistics would be observed if the null hypothesis were true, and seeing how extreme the observed statistic is compared with the simulated statistics thus the reason for the name randomization-based approach.

Hispanic Students and Statistical Literacy

Hispanics represented 24% of Florida's population in 2013 (Pew Research Center, 2014), and the projections for 2050 are that Hispanics will represent approximately 30% of the U.S. population, with a growth of 86% between 2015 to 2050

(Krogstad, 2014). Surprisingly, there is an absence of research on Hispanics students' statistical literacy. Such a lack of research represents a significant gap in the literature given the fact that Hispanics are the largest minority group in the U.S. and their educational attainment has larger implications in the progress of their communities and the economic development of the country.

The race gap has narrowed in college enrollment but not in college graduation, as reflected by the data from the National Center for Education Statistics (NCES, 2016). The total college undergraduate enrollment rate for Hispanics was 34% for ages 18-24 years old but only 53% graduated within 4 to 6 years (NCES, 2016). Hispanics were enrolled in a higher proportion of non-STEM fields: 19% of Hispanics were awarded a Bachelor's degree in business, 8% in health professions and related programs, 11% in the social sciences and history, 8% in psychology, and 4% in education (NCES, 2016). The data showed that Hispanics major, more than any other racial group, in the social sciences, psychology, and business. These majors required statistics as part of their curricula. Therefore, it is important for Hispanics students, in particular, to become statistically literate.

In addition to the projected growth, Hispanic students tend to attend Hispanic Serving Institutions (HSIs) because of cultural continuity (Corral, Gasman, Nguyen, & Samayoa, 2014). Therefore, it is imperative that these institutions start to conduct research on statistical literacy attainment in undergraduate introductory courses. Researchers should also conduct investigations on randomization-based inference methods since this course is the only contact point between a systematic scientific reasoning for non-STEM Hispanic students and the real world. Hispanic Serving

Institutions have recently focused effort, energy, and money to address their students' needs in terms of social and academic integration (Nunez, Crisp, & Elizondo, 2016). Many HSI students are first generation, non-native speakers, and non-traditional age students (Arbelo-Marrero & Milacci, 2016). Nevertheless, little to no attention has been given to curricula and pedagogical changes in statistics on a large scale (Cobb, 2015), a problematic situation since the majority of social sciences, business, and the medical field require statistics (Nicholson & Ridgway, 2017; Scott-Jones & Goldring, 2017). These students require proper training in order to reconstruct their intuitive thinking with the right set of heuristics to evaluate uncertainty—for example, how correlation does not imply causality.

Furthermore, Hispanic students pursue more non-STEM majors because of their shortcomings of a proper mathematical foundation, as reflected by the U.S. Census Bureau (2015). Hispanics with a bachelor's degree make up to only 10% of the STEM fields. Still, being statistically literate is important for non-STEM Hispanic students and is the only framework against students' strong willingness to predict rare events from weak evidence. In fact, the idea of regression to the mean is unfamiliar and difficult to comprehend and communicate (Kahneman, 2013).

Statistical literacy remains uncommon among students and more attention is needed to focus on early undergraduate statistical education. Randomization-based methods provide an alternative for students' diverse learning styles that may help students at HSIs in both degree retention and completion and support students to comprehend statistical concepts.

Statement of the Problem

As the U.S. Hispanic population increases, ensuring students' statistical literacy increases in importance. The Hispanic population has certainly made big inroads in college education, with 35% Hispanics ages 18 to 24 enrolled in either a two- or four-year college institution by 2014 (Krogstad, 2016). However, Hispanics still lag behind other groups in obtaining a four-year degree (Krogstad, 2016). Only 15% of Hispanics ages 25 to 29 have a bachelor's degree or higher, the lowest percentage compared to any other racial group (Krogstad, 2016; Hispanic Association of Colleges and Universities, 2017). Introductory statistics courses play a key role in college completion since it is a gatekeeper course for more advanced curricula (Cobb, 2015).

In light of the changes in teaching statistics, there was no research testing the new methods on Hispanic undergraduate statistical literacy. At the time of this research, there was a lack of studies in Hispanic statistical literacy, studies that explored whether the new teaching models overcame the barriers of statistical reasoning, and studies that examined how individual barriers, like language, might impact Hispanic students' statistical literacy. The language barrier that many Hispanic students experience in their academic journey is a strong predictor of college attrition, including in math and statistical courses (Garcia, 2000; Schmid, 2001). Language barriers were examined as one factor that the new statistical teaching strategies should overcome.

Consequently, there was a need to assure that the low level of statistical literacy attainment among undergraduate students was addressed in order to advance Hispanic students towards college completion. Additionally, language barriers for non-native speakers hinder students' ability to learn inferential statistics, and the randomization-

based inference methodology provides an alternative pathway for the development of statistical knowledge with the potential to bring them up to the level of native speakers in their understanding of statistical concepts.

Purpose of the Study

The purpose of this study was to explore the effects of a randomization-based teaching methodology in introductory undergraduate statistics courses in the light of promising advantages with the use of simulations on the level of statistical literacy among students at a Hispanic Serving Institution. The present project focused on two simulations: bootstrapping confidence interval method and the randomization hypothesis test method. The appeal of the bootstrapping confidence interval method and the randomization method was that they were logical and accessible, which could support students with the comprehension of statistical inference. Hispanic students have a unique set of needs in order to become statistically literate. Additionally, there was an absence of research that explored how student characteristics, such as non-native speaker, non-traditional student, and first generation, interact in the achievement of statistical knowledge.

The present study looked at the learning gains in the nine topics that comprise statistical literacy as measured by the CAOS. These topics were data collection and design, descriptive statistics, graphical representations, boxplots, bivariate data, probability, sampling variability, confidence intervals, and test of significance in order to assess how well simulations aids in the understanding of key statically ideas. The goal was to test whether the randomization-based teaching methodology increased statistical literacy, even when accounting for student characteristics, over traditional methods.

Research Questions

Several unanswered questions prevail related to assessing and understanding statistical literacy among undergraduate Hispanic students. In light of advances to the introductory undergraduate statistics course, statistical literacy should be assessed among Hispanics students and it needs to be considered essential to instructors who impart the course.

The following research questions were examined in this study:

1. Do students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group?
2. Are students' demographic and academic characteristics—age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, highest math class taken, and GPA—associated with statistical literacy scores?
3. Do students in a randomization-based inference teaching modality group have higher statistical literacy scores than students in the traditional theory-based inference teaching methodology group when controlling for demographic and academic characteristics?

Nature of the Study

To address these research questions, the study had one traditional theory-based comparison group and the simulation-based group. The simulation group had an instructional strategy with the intention to trigger conceptual change grounded on Liu's

(2010) learning model in order to stimulate different representations of the concepts involving inferential statistics. The research questions evaluated the real potential of a randomization-base teaching strategy in achieving the learning outcomes associated with an introductory undergraduate statistics course. In accordance with the literature, simulations should trigger distinctive response on students' insight and image of statistical activity while using it fosters the awareness of statistics as a tool to organize information (Budgett & Wild, 2014).

The study examined the most pervasive misconceptions in statistical thinking related to sampling variability, parameter estimation, and hypothesis testing by triggering critical reflection with the use of simulations and then exposing students to traditional theory-based methods, since misconceptions can peacefully coexist with the right statistical concepts (Khazanov & Prado, 2010). However, when multiple cognitions are stimulated, they have to be consistent with one another, thus assisting students in constructing a satisfactory abstract representation (Bernstein, Clarke-Stewart, Roy, Srull, & Wickens, 1994; Khazanov & Prado, 2010).

The students' concept images are generated from the simulations. Tall and Vinner (1981) defined the concept image as "the total cognitive structure that is associated with the concept, which includes all mental pictures and associated properties and process" (Tall & Vinner, 1981, p. 152), whereas a concept definition is the words by which a mathematical term is defined (Aquilonius & Brenner, 2015). As mathematics comes to be "more abstract students' concept images will have to be consistent with formal definitions" (Aquilonius & Brenner, 2015, p. 11). In the present study the concepts that were triggered with the randomization-based approach were definitions associated with

the process of inferential reasoning—for example, sampling distributions of the mean, margin of error, confidence interval, p-value, alpha value, and so forth.

The study measured the effects of an instructional strategy using statistical simulations with the CAOS instrument. The instrument targeted “statistical literacy and conceptual understanding, with an emphasis on reasoning about variability” (delMas et al., 2007, p. 31). It consisted of 40 multiple-choice items that cover the six major topics of any undergraduate introductory statistics course. The topics included data collection and data design, graphical representations, variability, sampling variability, test of significance, and bivariate data. It was important to test all topics across the introductory undergraduate statistics curricula because there was no reported data in the literature about Hispanics statistics literacy rates and because all the topics were either to a higher or lesser degree correlated with inferential analysis. By 2014, over 30,000 undergraduate students had taken the CAOS test at the end of their respective courses (delMas, 2014); the test is nationally normed and backed by a reliability study yielding a Cronbach’s alpha coefficient of 0.78.

Contributions to the Field

There is a dearth of research about the effects of a randomization base-inference teaching methodology within the Hispanic undergraduate student body. The effectiveness of the randomization-based teaching methods needs to be appreciated in relationship with the instructional contexts, particularly HSI and with the demographic and academic characteristics of students within this higher educational context.

The present study provided insights into the achievement level of Hispanics students in undergraduate introductory statistical courses, since there was no baseline

reported in the literature for this ethnic group. The instrument used for the study, CAOS, captured and measured the required learning outcomes in a statistics course thus providing a baseline to compare it with the results of the national sample. The majority of students in the national sample have been taught under the traditional theory-based approach (Tintle et al., 2012). Randomization based approaches have been conducted in Predominantly White Institutions (PWI) where a White enrollment rate ranges from 56% to 80% (McDonald & Vrana, 2007). Therefore, the CAOS national sample did not capture the achievement level of Hispanics students. The present study added new knowledge about Hispanics students' statistical literacy to the literature.

The current study supported the use of the new randomization-based method approach to the teaching and learning of undergraduate introductory statistics courses by measuring the effectiveness of simulations followed up with the theory base methods in students' statistical achievement. Moreover, it encouraged Minority Serving Institutions, specifically HSIs, to revisit and perhaps overhaul the traditional theory-based teaching methodology in order to level off with PWIs with the current pedagogical trends in teaching statistics.

Definitions of Terms

Key terms that were frequently used throughout the study are defined to assure coherent information.

Cognitive Conflict: it is the existence of logical inconsistency among knowledge, opinions or belief arising from the mind (Tuan & Jay, 2016).

Concept Image: it “refers to the total cognitive structure that is associated with the concept, which includes all the mental pictures and associated properties and processes” (Tall & Vinner, 1981, p. 152).

Concept Definition: the “words by which a mathematical term is described and accepted by the mathematical community at large” (Tall & Vinner, 1981, p. 152).

Conceptual Knowledge: “the interrelationships among the basic elements within a larger structure that enable them to function together” (Krathwohl, 2002, p. 4).

Hispanic Serving Institution (HSI): it “is an institution of higher education that is an eligible institution and has an enrollment of undergraduate full-time equivalent students that is at least 25% Hispanic students at the end of the award year immediately preceding the date of application” (The United States Department of Education, 2013).

Minority Serving Institution (MSI): any institution that is defined as a Historically Black College or University (HBCU) or a Hispanic Serving Institution (HSI) (Hubbard & Stage, 2009).

Predominantly White Institution (PWI): an institution with a White enrollment rate that range from 56% to 80% (McDonald & Vrana, 2007).

Procedural Knowledge: “how to do something; methods of inquiry, and criteria for using skills, algorithms, techniques, and methods” (Krathwohl, 2002, p. 4).

Statistical Literacy: this definition differed in the spectrum of basic skills to critical thinking. The most recent definition is by Jane Watson (2011) which refer to: “understanding of basic statistical problems and terminology, being able to use the basics in the real world, and questioning statistical conclusions and results” (Watson, 2011, p. 198).

Traditional Statistics Curricula: “Introductory statistics curricula which use theory-based approaches, usually start with descriptive statistics, then transition to probability and sampling distributions, ending with statistical inference” (Tintle et al., 2011, p. 2).

Summary

This chapter introduced the nature of the problem and the current state of research in regards to statistical literacy among undergraduate college students with an emphasis on Hispanics students attending HSI. The fact that 60.8% of all Hispanics students attend HSI (Hispanic Associations of Colleges & Universities, 2017) deserves attention in light of emergent pedagogical approaches for the teaching and learning of inferential statistics, specifically randomization-based methods. Also, this chapter introduced the research questions that were addressed in the study.

Chapter 2 presents a comprehensive review of the literature in the field of statistics education related to the traditional theory-based teaching methodology, current instruments that measured statistical literacy, the randomization-based teaching methodology, a summary of courses and textbooks designed to teach statistical literacy with the randomization base-methods, students' barriers to achieving statistical literacy, a detailed description of the undergraduate Hispanics students' characteristics, and journal articles of students' understanding of statistical literacy are critiqued to justify this project.

Chapter 3 presents the research questions and the methods used to conduct this study. To answer the research questions, a non-equivalent group design (NEGD) was used since the class selected for the simulation-based teaching and for the traditional

theory-based teaching lacked random assignment of students. The CAOS test was selected as the instrument along with a demographic and academic survey that focuses on the students' characteristics that attend a HSI in order to conduct a stepwise linear multiple regression analysis.

Chapter 4 reports the outcomes of the study. This includes the descriptive statistics and inferential statistics to test the research questions. Finally, chapter 5 provides a discussion and the conclusions of all research questions based on the outcomes of this study. Furthermore, this chapter presents the limitations of the study, the implications for the practice, and for future research.

CHAPTER 2

LITERATURE REVIEW

Introduction

In 2002, Jane Watson, in her speech to the American Statistical Association, said that statistics was the most important quantitative subject in a college curriculum. She foresaw the changes in the world ahead. Teaching statistical courses are critical and challenging in a world in need of fast and effective decisions because they serve a student body with a different set of backgrounds and skills, many of whom have had negative experiences in quantitative courses (Garfield, 1995). These experiences influence future students' beliefs, perceptions, and attitudes towards statistics, and for this reason the society at large (Garfield, 1995).

The statistics education community still perceives statistics as a field with a need for considerable innovation in how students are educated, even with the recent attention given to the teaching and learning of statistics and the surge in statistics education research (Garfield & Ben-Zvi, 2008; Lane-Getaz, 2017). Recently, Lane-Getaz (2017) has been trying to comprehend the complexities in teaching and learning inferential statistics to establish desirable curriculum advances for the formation of future college professionals in response to current forces trying to ban inferential procedures from social sciences courses, as suggested by Trafimov and Marks (2015).

Statistical literacy is important in the workplace and everyday life. Recently, statistics educators have given priority to statistical literacy in statistics education reforms (Ben-Zvi, 2004; Cumming, 2010; Pfannkuch et al., 2011). One of the implications is the

movement towards socially-based curriculum frameworks and application-based approaches that teach students to think and reason critically about social circumstances through data, usually denoted as applying statistical literacy.

Since the arrival of undergraduate introductory statistics courses, the core of statistical education has centered around inference and is based on the Central Limit Theorem and the sampling distributions of estimators (Budgett et al., 2013). The logic of inferential statistics is grounded on normal-based theory and it has long been well-known that students have challenges in understanding how asymptotic testing operates (Cumming, 2010). Furthermore, Liu and Thompson (2009) reported that many instructors are challenged to fully comprehend the details involved in the methods of making inferential reasoning. Nonetheless, with developments in computing power, we can rely on alternative methods to foster students' understanding of inferential statistics. Several statistics education researchers (Budgett et al., 2013; Chance et al., 2016; Maurer & Lock, 2016) have introduced the bootstrapping and randomization methods as integral components in the teaching and learning of inferential statistics. The allure of these techniques is that they are logical, are reachable, and effortlessly lend themselves to visual systems, which could help students with their comprehension of the inference process (Budgett et al., 2013).

Statistical Literacy

There is no unanimously accepted definition of statistical literacy. However, the statistics education research community has shaped an expanding view of what comprises this construct (Watson, 1997). One of these definitions is "People's ability to interpret and critically evaluate statistical information and data-based arguments appearing in

diverse media channels, and their ability to discuss their opinions regarding such statistical information” (Gal, 2002, p.2). Traditionally, researchers have two main positions regarding the nature of statistical literacy. One position considers that statistical literacy is an unnatural act (Garfield, 2002). Students think about probabilistic concepts in ways that produce cognitive conflict with the predictions of probability theory. From this point of view, understanding probability does not emerge instinctively, and training has a simple effect in improving students’ misconceptions (Cobb, 2007; Knold, 1995). The alternative position affirms that students cultivate insight notions of probabilistic concepts and reasoning that allow them to deal with basic statistical activities (Aquilonius & Brenner, 2015). These insight notions originate from the experience with circumstances that display random behavior (Nisbett, 1993), and they are used in a number of reasoning processes like induction.

According to Watson (2011), statistical literacy includes an “understanding of basic statistical problems and terminology, being able to use the basics in the real world, and questioning statistical findings and results” (Watson, 2011, p. 198). Therefore, important foundational blocks of statistical literacy are confidence intervals, null hypothesis significant test procedures, and p-values (Reaburn, 2014). Null hypothesis significance tests emerged as one alternative around the problem of uncertainty; a proposal is made about a population parameter (the null hypothesis), then a sample is collected, and the appropriate sample statistics are calculated in order to decide as to how likely the sample statistic would be if the proposal about the parameter were true. This decision is performed by calculating the probability of observing the sample with the given or more extreme statistic (Reaburn, 2014). This conditional probability is known as

the p-value (Reaburn, 2014). That is, the p-value is the probability of making the wrong decision if the null hypothesis is true.

Previous studies show that students may think that the null hypothesis speaks of both the sample and the population (Aquilonius & Brenner, 2015; Cumming, 2010; Lane-Getaz, 2017). Consequently, they formed the wrong concepts about a null hypothesis significant test since the beginning of the process, and this erroneous misconception is compounded by the students' belief that the p-value refers to the probability that the null hypothesis is true and to the students' belief that 1 minus the p-value is the probability that the alternative hypothesis is true, thereby leaving students with low statistical literacy achievement levels (Reaburn, 2014). As a result, it is important to revise the current conditions of undergraduate students about the teaching and learning of statistics, especially inferential statistics since it is the contact point between judgment about uncertainty and generalizations.

Statistical Literacy among College Students

Statistical literacy among college students reflects the historical development of statistics. Most college courses introduce descriptive statistics with basic probabilities. Then, these two ideas converge and upgrade when estimations about parameters and hypothesis testing are introduced (Cobb, 2007). This sequence of events leaves many students with a vague conceptual understanding that is reflected by not mastering the required learning outcomes for introductory statistics courses at the college level (Chance et al., 2007). To understand statistics, learners must comprehend the essence of sampling and the manifestation of variance, as well as the relationship between statistics and parameter (Garfield & Ben-Zvi, 2007). Students need to understand probability as a

measure of uncertainty. They need to know how to develop and use models that simulate random phenomena and how to use data to estimate probabilities (Garfield, 1995). More importantly, they need to adjust and apply the probabilistic knowledge to the data they collected to the preliminary conclusions they are working on.

An undergraduate introductory statistics course should provide a stable viewpoint between descriptive and inferential statistics that build students' abilities from sampling awareness towards sample collection, description, analysis, and inference (Carver & Stephens, 2014; Tittle et al., 2011). Several statistics educators agree that the leading goal for teaching introductory statistics courses is the development of statistical literacy (delMas et al., 2007; Garfield & Ben-Zvi, 2007; Rumsey, 2002). A big component of statistical literacy is statistical inference which assesses the extent to which the pattern in sampling could be attributed to an underlying probability distribution (Cobb, 2007). In other words, statistical inference evaluates how much variation in the estimators that describe the sample can be expected by randomness, and to what extent students can be sure that their conclusions represent the underlying behavior of data. By comprehending that the sample is a subset of the population, and statistics are measurements from the sample, not measurements from the population (parameters), the students can make statistical decisions without having observed all instances of a phenomenon. By addressing and measuring variation through probability theory, statistical inferences permit the students to reason about the expected behavior in samples (Lock, Lock, Morgan, Lock, & Lock, 2017).

The American Statistical Association endorsed the Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report, originally created in 2005

(Aliaga et al., 2005). The guidelines provided six recommendations which are: 1. Accentuate statistical literacy and statistical reasoning, 2. The use of real data, 3. Emphasize conceptual understanding instead of procedural knowledge, 4. Promote student centered activities in the classroom, 5. The use of technology gear towards statistical inference and data analysis, and 6. The use of assessment to measure students' gains. However, these guidelines do not propose in-depth changes to the content and pedagogy, which might be one of the reasons for poor performance on the Comprehensive Assessment of Outcomes in Statistics test (CAOS) national data (De Veaux, 2014; Horton, 2015; Tintle et al., 2011). The 2016 GAISE College Report recommended two new aims for teaching statistical literacy that reflect current practices: "Teach statistics as an investigation process of problem-solving and decision-making and give students experience with multivariable thinking" (GAISE, 2016, p. 6).

Measuring Statistical Literacy

Assessing statistical literacy in introductory statistics courses, in the light of an emergent trend of randomization-based methods, would allow comparisons with the traditional theory-based method that teaches asymptotic testing for inferential statistics. The assessments that have been created, up to now, to measure outcomes for college students enrolled in an introductory statistics course seem to do a satisfactory job of assessing students' statistical ability in making appropriate connections related to the interpretation of the findings from the data collected (Garfield et al., 2012). These results apply to both instructional approaches: the traditional-theory based (consensus curriculum) and the randomization-based (simulation-based) curriculum. The goal of this project is to teach the randomization-based inference approach as an introduction to the

traditional-theory based methods and determine if this system results in students' learning gains.

There are four instruments found in the Assessment Resource Tool for Improving Statistical Thinking (ARTIST) website that have been developed and used with undergraduate students at the college level. The instruments are ARTIST Comprehensive Assessment of Outcomes in a First Statistics Course (CAOS), ARTIST Topic Scale tests, Statistics Concept Inventory (SCI), and Statistical Reasoning Assessment (SRA).

The CAOS test was designed to measure students' overall statistical knowledge (delMas et al., 2007). "The goal was to measure students' conceptual understanding of the big ideas in statistics, including variability" (delMas et al., 2007, p. 31). The ARTIST Topic Scale tests was used to build some of the items on the CAOS test. The test consists of 40 selected-response items which were developed over a period of three and a half years and have been tested and revised several times. The test was designed to measure statistical literacy and reasoning. It is the only assessment that is nationally normed and backed up by a reliability study (Maurer & Lock, 2016). The reliability is high; coefficient alpha is .82. As of 2016 over 35,000 students have taken the CAOS test as a posttest in a college-level introductory statistics course, including students in AP statistics courses (delMas, 2014). The CAOS test is the only instrument that allows researchers to compare their findings against the national sample of students who have taken the test.

The ARTIST Topic Scales assessment consists of 11 selected-response tests designed to "assess the big ideas in statistics such as data collection, probability, and tests of significance" (Garfield et al., 2002). Each topic scale test consists of nine to 15 selected-response items. The items were developed to measure statistical literacy and

reasoning (delMas, Garfield, & Ooms, 2005). However, questions from the ARTIST test have been criticized as being increasingly outdated and lacking reliability and validity (Ziegler, 2014).

The SRA evaluates conceptual understanding of statistics (Reed-Rhoads, Murphy, & Terry, 2006). This instrument contains 25 items and it is a selected-response statistics test that is often used for STEM students. One downside of this instrument, is that non-STEM students have been shown to score significantly lower on it (Allen, 2006), thus the reason why it is not commonly used in statistics education research. The sections in the test are descriptive, probability, inferential, and graphical. The SRA was designed to determine the students' ability to reason about statistics and probability (Garfield, 2003). The SRA comprises 20 selected-response items. Each item consists of multiple correct answers, but only one contains the correct rationale. The remaining five questions are open-ended. Even though the SRA is an easy assessment to administer, it is a flawed assessment tool because the internal consistency reliability coefficients revealed that the intercorrelations between questions were low and the questions did not appear to be measuring one trait (Garfield, 2003). Furthermore, in the criterion-related validity study, the resulting correlations were all extremely low (Garfield, 2003). For this reason, this instrument was not considered.

For the purpose of this study, the CAOS test will be used to test changes in statistical knowledge. Specifically, two groups will be tested, one which received the randomization-based teaching methods and one which just received the traditional theory-based pedagogical approach in order to determine if there are differences between the groups at the end of the course statistical knowledge.

Teaching Statistical Literacy

The vast majority of introductory statistics courses adopt the traditional theory-based approach which highlights the mechanical computation and the understanding of statistics terminology. Such curriculum responds to the behavioral approach to teaching and learning statistical literacy (Hassad, 2011). In today's classroom, the traditional approach to inference embraces several sections in order to understand the central limit theorem so students can perform normal approximations once the assumptions are satisfied. These courses are designed to use basic algebra-based inference procedures that depend on tables, formulas, and/or scientific calculators based on assumptions of a particular distribution. There has been an ongoing effort to make the traditional theory-based inference approach meaningful through media examples in order to make it relevant to student's daily life, helping them to become informed citizens. However, Utts (2003) identified several topics that were usually misinterpreted by students under the traditional theory-based inference methodology such as conclusions about causation, coincidences are not unusual, statistical significance versus practical significance, comprehension about the nature of variability, and bias among others.

The teaching of statistical literacy has shifted greatly from a course dominated by the use of distributional assumptions to include data exploration, simulations, and model diagnostics (Roy Rossman, Chance, Cobb, VanderStoep, Tintle, & Swanson, 2014). Pedagogical advancements towards active learning is now combined with content change and the increased use of technology. The technological devices that dominate the classrooms are limited to the graphing calculator, applets, and multimedia materials (Maurer & Lock, 2016). There is a growing group of statistics educators who consider

that the curriculum approach has stopped short of the potentials that computation might offer (Cobb, 2007). Cobb (2007) discussed that teachers of inferential statistics have done well to embrace technology to displace dull calculations but have not successfully changed the approach to teaching statistical inference. He stated, “our curriculum is needlessly complicated because we put the normal distribution...at the center of our curriculum, instead of the core logic of inference at the center” (Cobb, 2007, p. 4). Either because of resistance or a lack of training on behalf of statistics instructors to embrace a simulation-based curriculum, social scientists, most recently being Trafimow and Marks (2015), have called for a ban of hypothesis testing methods from the *Journal of Basic and Applied Social Psychology* and any applied social science course. However, other education statistics’ researchers have incorporated the heart of inferential reasoning technology based-simulations to improve the conceptual understanding (e.g., Chance et al., 2016; Diez, Barr, & Cetinkaya-Rundel 2014; Forbes et al., 2014; Maurer & Lock, 2016; Tittle et al., 2015; Zieffler et al., 2015).

According to the Florida Department of Education (FLDOE, 2017), all introductory statistics courses at the 2000 level must include the following topics: “collecting, grouping and presenting measures of central tendency and dispersion; probability; hypothesis testing; confidence intervals; and correlations” (FLDOE, 2017). However, this description does not suggest the use of randomization-based inference, thus leaving college instructors unmotivated to explore the effects of simulations in students’ performance. The following sections describe each of these approaches and the strengths and weaknesses for creating student statistical literacy.

Traditional Theory-Based Inference Teaching Methodology

The traditional theory-based inference approach, also known as the algebra-based inference, requires comprehension of mathematical ideas that demand awareness of mathematical concepts like proofs to be understood (Tappin, 2000). Educators can use a mathematical mechanism to demonstrate why students should expect random sampling to behave in a particular way. However, teaching through mathematical proofs has some adversities. First, it requires advance prior mathematical background and skills from students, which many students lack or do not consider an essential part of their statistical training (Tappin, 2000). Second, focusing on formal proofs takes time away from instruction and more grounded engaging activities (Tappin, 2000). Formal mathematical structures are important for learners who have an inclination for proofs and logic, but they are not as relevant for students with more applied interests. The current tendency in statistics education is to highlight mathematical formalisms, but it does not carry from formal probability theory to the applied activities in statistics (De Veaux, 2014; Tintle et al., 2014).

The traditional theory-based methodology follows the consensus approach beginning with descriptive statistics, then move towards probability and sampling distributions, and finishing with confidence intervals and test of significance for one and two samples (De Veaux, 2014). The traditional theory-based approach introduces statistical inference in the final third of the course and around this time of the semester students may have lost sight of the big picture of the course and end up in survival mode (Tintle et al., 2011). Furthermore, by the end of the semester, students' stress and overload leads to a superficial level of inferential statistics understanding, which is "the

crux of the course” (Tintle et al., 2011). The courses end up emphasizing a series of asymptotic tests with complex conditions rather than the logic of statistical inference (Cobb, 2007). These arguments are partially justified by poor student performance, as reported by delMas (2014). The sub-topic of understanding sampling variability showed the lowest mean performance. Overall, students tend to score 50% correct on the CAOS test from 2005 to 2013 (delMas, 2014), suggesting that a more aggressive statistics reform movement should take place.

Impact of Traditional Practice to Student Statistical Literacy. According to Pfannkuch et al. (2011), most students leave the course not thinking and reasoning probabilistically. Students apply a deterministic form of reasoning to situations that involved uncertainty. Budgett et al. (2013) explained that the traditional practice of teaching hypothesis testing is, for the most part, confined to the Neyman-Pearson decision theoretic framework, which is one of the most challenging topics for students to comprehend. They documented persistent misconceptions related to inference and how these misconceptions hinder student advancement in reaching statistical literacy. The most common misconceptions are “considering a p-value as the probability that the null hypothesis is true rather than the probability of the data, regarding a p-value as the probability that the research results are due to chance, believing that the size of a p-value is an indicator of the size of any difference, accepting the null hypothesis if the p-value is considered to be large, and interpreting a statistically significant outcome as essentially important” (Lane-Getaz, 2017, p. 379).

Aquilonius and Brenner’s (2015) study about students’ inferential reasoning developed three problems (coin problem, home value problem, and tranquilizer problem)

to explore, through conversations and interviews, the traditional approach to hypothesis testing and the misconceptions about each step of the process. The authors provided to the 16 students from different community colleges in an introductory statistic course the seven steps to follow in order to provide a judgment about the uncertainty of the event. Those steps were “list the information given in the problem, set up the hypothesis, decide on a level of significance, sketch a graph showing the critical region, use the calculator to compute the p-value, decide to reject or fail to reject the null hypothesis, and answer the question asked in the problem” (Aquilonius & Brenner, 2015, p. 15). Even though the majority of students were able to make the correct decisions based on the p-values, the students could not explain to the researchers what a p-value was. Some students build the concept that a low p-value is a small likelihood for the null hypothesis to be true. For others students, the p-value did not have any importance if a non-significant level was specified. The transcribed data from the interviews reflect that students did not seem to consider the concept of sampling variability that is intimately connected with the concept of p-value. These findings support the quantitative analysis by delMas (2014) where the lowest mean score in the sub-topics of the CAOS test was understanding sampling variability.

Overall, the traditional-theory based teaching methodology did serve well for those students with a mathematical background and inclination for the subject. However, this pedagogical approach fell short in helping students with a non-mathematical background in developing statistical thinking and reasoning, both aspects of statistical literacy and desirable goals of any undergraduate introductory statistics course. In light of emergent changes in educational technology, it is the duty of statistics educators to

investigate another pedagogical approach that helps students to achieve the desired level of statistical knowledge. Such educational technologies go beyond tools that aid in the computation like calculators and/or spreadsheet; this is the case for the randomization-based inference approach.

Randomization-Based Inference Teaching Methodology

The randomization-based inference methodology, also known as the simulation-based teaching approach, is based on resampling techniques. This approach is only feasible with the use of current technology which provides instructors today with a unique opportunity to teach resampling methods in addition to asymptotic test such as student's t-test (Tintle et al., 2011). The most frequent methods used under this approach in an introductory statistics course are bootstraps for confidence intervals and randomization test for hypothesis testing (Lock et al., 2013). These two particular approaches aid the students in understanding the concept of variability from sample to sample while at the same time act as a stimulus to understand the idea behind estimation of parameters and the idea of testing propositions about a parameter, both key ideas of inferential statistics.

Efron (1979) introduced the Bootstrap method, which requires that the samples be drawn with replacement. The goal is to estimate the population parameter from the observed sample, which is considered to be representative of the population. In order to accomplish the goal, the bootstrap sample is taken multiple times from the observed sample until the bootstrap sampling distribution is generated. In other words, the bootstrap sample is to the sample like the sample is to the population. Once the distribution is generated the confidence intervals can be found by either using the

standard error of the bootstrap distribution or by looking at the percentiles in the bootstrap distribution (Lock et al., 2013).

The bootstrap method helps the student to understand the concept of the sampling distributions of the means. This concept cannot be visualized only imagined unless the students have access to a dynamic visualization tool. For the bootstrap confidence interval, the tool allows the students to generate the bootstrap sampling distribution of the means thus students can actually see the spread of the distribution around the mean and determine, with a margin of error, the location of the parameter with a particular certainty level.

Alternatively, Fisher (1936) described, for the first-time, randomization test by comparing two groups. In a randomization test for hypothesis testing, we have to sample without replacement in order to generate random samples where the null hypothesis is simulated by combining two groups of data and randomly assigned observations to the two groups. Several randomized samples are taken, and then the observed statistic is compared against the distribution in order to determine if the outcomes are significant (Lock et al., 2013).

The randomization test aids the students in visualizing the natural variability from sample to sample and the concept that, after selecting multiple samples, eventually we either accept or reject the proposition made about the parameter based on the observed statistic. This process of selecting thousands of samples can only be performed with simulations, otherwise, students have just to imagine the scenario and every student evoke their own particular image as explained by Tall and Vinner (1981). However, with

the use of simulations, the instructor has the opportunity to generate a more homogeneous concept image.

Each of these methods—bootstrap and randomization—recreated a representation of a statistical idea evoked by the student, thus helping them in the construction of statistical knowledge (Pfannkuch & Budgett, 2014). The focus of the randomization-based inference method is on conceptual knowledge, rather than procedural knowledge which is aligned with the theory-based methods.

Rationale for the Randomization-Based Inference Teaching Methodology.

One benefit of using randomization-based inference methods is that there are fewer assumptions associated with it than with the algebra-based method (Chance et al., 2016). In fact, there are circumstances where randomization-based procedures are more precise than those of the algebra-based method (Hesterberg, Monaghan, Moore, Clipson, & Epstein, 2003). Consequently, if randomization-based methods are instructed in addition to the algebra-based method, students are expected to gain deeper comprehension of the theoretical solutions because the affordance in producing graphical representations of abstract concepts associated with statistical reasoning, like the process associated with uncertainty (Tintle et al., 2012).

Blejec (2002) claimed that randomizations may, in fact, fulfill a role once reserved for mathematical proofs. He proposed that dynamic visualization tools that supported randomizations can be used as “proofs” of statistical concepts for students who do not have advance mathematical backgrounds. Therefore, it is possible to generate graphical representations of mathematical objects in which students can actively engage with the mathematical properties of those objects. Randomness is only visible when

students look at cumulative events over several trials and with simulation-supported tools, it is possible to generate random processes, capture events over time, and see how trends change as a function of different parameters (Budgett et al., 2013).

The scope of the randomization-based inference approach goes beyond replacing physical simulations and facilitating the presentation of randomness in the instructional process. Simulation-based inference permits students to connect the random process with different graphical representations. From this perspective, statistical simulations serve as a tool to observe phenomena that cannot be observed under normal conditions (Chance & Rossman, 2014). Several proponents of the randomization-based inference methodology dispute that this allows students to have contact to the core concepts of statistical inference without first needing the comprehension of theoretical probability distributions (Maurer & Lock, 2016). Randomization-based methods that have been taught in introductory statistics courses include bootstrap confidence intervals and randomization tests (Pfannkuch & Budgett, 2014).

Randomizations-based inference activities foster students' being involved in authentic statistical problems (Budgett & Wild, 2014; Pfannkuch & Budgett, 2014). Authentic problems mirror actual research quests and thus highlight essential phases of statistical inquiry, such as understanding the nature of the sampling distributions of the means. Randomization activities demand that learners engage in activities embedded in disciplinary contexts that push students to deal with aspects of learning that remain unclear when students are just presented with artificial reduced versions of statistical problems, with activities centered in procedural knowledge, and with mathematical formalism (Budgett & Wild, 2014). Simulation-based inference activities foster students'

participation in the process of coordinating empirical evidence and theory (Coll-Serrano, Blasco-Blasco, & Alvarez-Jareno, 2011; Novak, 2014).

According to the randomization-based inference, the core elements of the practice and activity of statistical inference cannot be seized by the mathematical formalism or taught by an explicit direct instruction to novice individuals (Roy et al., 2014). Working with a simplified set of problems can be advantageous to the learning of probability because doing so assists with the calculations and permits observations of the underlying mathematical mechanisms of hypothesis testing. However, working in artificial settings does not foster students' cultivation of basic skills for inference and interpretation (Snee, 1993).

In teaching college statistics, the role of randomization-based inference methodology can substantially expand the opportunities for students to achieve significant learning outcomes (Tintle et al., 2014). Novak (2014) argued that the case for reform in randomization-based instruction is built on strong synergies between content, quality instruction, and simulation-based technologies. In teaching college statistics with randomizations, reflective instructional practice is desirable since certain categories of software, such as spreadsheets, could continue to obscure the conceptual foundation in an analogous way to the traditional theory-based methodology, whereas purpose design software for teaching and learning may make the process clearer (Pfannkuch & Budgett, 2014). Therefore, exposure to such dynamic visualization tools can assist the students' mental abilities and intuitive thinking in order to develop powerful mental conceptualizations (Sacristan et al., 2010).

Technology for Randomization-Based Inference Teaching Methodologies.

There are several available technologies for an introductory statistics course that cover randomization-based methods within textbooks and supporting materials. The main four technologies are S-PLUS with the companion textbook *The Practice of Business Statistics*, Applets with the companion textbook *Introduction to Statistical Investigation*, TinkerPlots with the companion textbook *Statistical Thinking: A Simulation Approach to Model Uncertainty*, and finally StatKey, which was developed for instructional purposes allowing students to visualize distributions. The current study implemented the approach of first presenting randomization-based methods followed up with the theory-based method, since the instructional strategy was to promote the students' ability to reconcile the randomization-based method with the traditional-theory method in order to conduct inference. The selected technology was StatKey, a free software available to any user that has a computer with access to the internet (See Table 1).

Table 1

Technology for Randomization-Based Methods

Technology	Textbook	Pedagogical Approach
S-Plus, 2007	The Practice of Business Statistics	Simulations based methods at the end of the course as an additional chapter
TinkerPlots	Statistical Thinking: A Simulation Approach to Model Uncertainty	Simulations based methods as the only inferential approach
Applets	Introduction to Statistical Investigations	Inference through simulations are covered from the beginning of the course alongside theory-based methods
StatKey	Statistics: Unlocking the Power of Data	Simulations base methods as introduction for distributional theory-based methods

The textbook used for the current study was *Statistics: Unlocking the Power of Data* by the Lock family (Lock et al., 2017), who developed the software. The textbook used real data and real applications, it was comprised of four academic sections as follows: Data, understanding inference, inference with normal and t-distributions, and inference for multiple parameters. The first section was analogous to the theory-based method because descriptive statistics does not require resampling techniques in order to understand it. Since students were working at the level of collecting and describing data, it should be to an extent a homogenous learning experience for the student regardless of the teaching methodology. The second section covered bootstraps confidence interval for means and proportion with one and two samples, and for matched pairs cases. The second section also covered randomization test as an alternative to the traditional approach to

hypothesis testing. The techniques employed to compute the bootstraps confidence intervals were the standard error (SE) of the bootstrap sampling distributions and then by the percentile approach. The textbook, in section three, shifted over to the traditional-based inference methods but continued to relate the students learning experience back to the randomization-based methods as a reinforcement strategy.

StatKey, as an educational technology tool, was the best choice for this project since the goal was to trigger cognitive conflict in order to stimulate deeper understanding of statistical concepts by the effort that students will have to make to reconcile the randomization-based methods with the traditional theory-based method. The pedagogical approach using StatKey has recently been validated by Maurer and Lock (2016) who found an overall increase in statistical literacy.

Barriers to Statistical Literacy

The body of research in statistics education reveals that students usually hold non-standard concepts in inferential statistics (Aquilonius & Brenner, 2015; Khazanov & Prado, 2010; Lane-Getaz, 2017; Reaburn, 2014). In the literature, the non-standard concepts apply to misconceptions, misperceptions, preconceptions, and naïve concepts (Khazanov & Prado, 2010). However, the most frequent term in the literature is misconception. The term misconception is defined as “a student conception that produces a systematic pattern of errors” (Smith, DiSessa, & Roschelle, 1994, p.125). The research demonstrates that a large proportion of college students have misconceptions about statistical inferences, particularly the acceptance of the null hypothesis on the basis of a larger p-value which emerge from a misapplication of deductive reasoning (Budgett et al., 2013). Aquilonius and Brenner (2015) identified that the most prevalent

misconception about p-values seems to concern the p-values as a conditional probability. Such misconceptions might have originated as the product of informal experiences students have with random events at the workplace, at home, and while playing games that involve chance. Another place in which misconceptions might have originated is at school while learning about mathematics and statistics (Aqilonius & Brenner, 2015).

There is robust data indicating that misconceptions about statistical literacy do not disappear as a result of traditional-theory based instruction about inference, which includes asymptotic sampling distributions test that required the assumptions of complex conditions (Khazanov & Prado, 2010; Tittle et al., 2011). According to Khazanov and Prado (2010), misconceptions can peacefully coexist with the right concepts. This symbiosis hinders students' application of statistical inference ideas consistently and with certainty.

In order to understand the barriers to statistical literacy, students must comprehend the behavior of sampling and the essence of probability. The intrinsic relationship between them has inspired researchers to identify barriers with students achieving statistical literacy. According to Garfield and Ben-Zvi (2007) "the main barrier is students' focus on the mathematical and mechanical aspects of knowledge, disempowering them to apply statistical content knowledge to solve problems arising from a specific context" (p. 375). They also claim that ideas of probability and statistics are very challenging for students to learn. Smith and Staetsky (2007) demonstrate the inability of students to apply statistics reasoning in everyday life, leaving students with little to no statistical literacy.

There are three basic stochastic misconceptions within the framework of an introductory undergraduate statistics course. The first one is the equiprobability misconception which is a students' predisposition to think that everything is equally likely to happen in a random experiment to different events. This misconception should be addressed by providing students with resources that recreate random phenomena (Khazanov & Prado, 2010). Purpose-built dynamic visualization software can recreate random phenomena, promoting students' mental images concepts which foster problem solving using a process of visual reasoning unlike verbal processes, which deal with information sequentially (Budgett & Wild, 2014).

Pfannkuch and Budgett (2014) conducted a case study to explore the construction of inferential concepts through the use of simulations. The underpinning framework of the study was based on three types of activities that a student may engage in when interacting with a dynamic visualization software in order to construct concepts. These steps are surface observation of the visual representation, property observation of the representation, and actions on the representation. Each of these activities is influenced by the reasoning that the representation of the statistical idea evokes for the student, thereby building schema through students' procedural and conceptual process views (Pfannkuch & Budgett, 2014). The results of the case study support the idea that visual imagery may contribute to a richer conceptual understanding of statistical ideas. Furthermore, they believe that dynamic visualizations enabled students to construct an appropriate scenario which visually and verbally define the behavior of the randomization test, thus building students' conceptual understanding of the core concepts of inferential statistics (Pfannkuch & Budgett, 2014).

The second misconception is the outcome orientation. Because students have an intuitive model of probability in which expected values represent single event outcomes instead of tendencies in a succession of cumulative single events (Garfield, 2002), students treat the probability of an occurrence or non-occurrence as an affirmation of certainty (Khazanov & Prado, 2010). Interpreting probability values in terms of single event outcomes can be produced by a misunderstanding of probability theories. Probability is only observable in a summation of consecutive trials; it is never visible in one event. But everyday life is made up of single events and students are accustomed to thinking in those terms. Therefore, a randomization-based inference approach presents an opportunity in correcting these misconceptions (Aquilonius & Brenner, 2015).

The last misconception is the representativeness bias. Students tend to think that all samples resemble the population from which they were obtained regardless of the size (Kahneman et al., 1982). Students' notion of the power of a well-drawn sample is not easily understood. Students often consider sample size as an arbitrary subset of the populations that do not include variability or sampling effects. This erroneous student perception of sampling leads to reason in flawed ways about statistical inference (Khazanov & Prado, 2010; Saldanha & Thompson, 2003). Consequently, these three misconceptions leave students with a weak overall understanding of inferential statistics. The following series of research explores the compounding effect of these misconceptions in an undergraduate introductory statistics course.

Turegun (2011) designed a pretest-posttest study in order to understand community college students' preconceptions of variance and their ability to think and reasoning statistically after taking an introductory statistics course. Turegun used the

ARTIST Measures of Spread Topic Scale. The study included analysis of student journals and interviews in order to gather more data about students' statistical literacy. The result indicated that students' scores were significantly higher than what would be expected by not taking the course; nevertheless, the overall scores were low. The 95% confidence interval for the average total scores on the pretest was between (CI 3.9, 5.2), and for the posttest between (CI 5.6, 7.4) which is low, even though the results are significant. One flaw of the study was that the sample size was 29. This is considered a small sample size, thereby impacting the researcher's ability to produce generalization. Furthermore, Turegun found that students underestimated their comprehension of variability. In fact, after Turegun inspected students' journals and interviews, he concluded that students lean towards using their own terminology and they misuse words such as variability and range. Turegun (2011) suggested that the statistics education community should consider these findings to promote students' development of an understanding of spread.

Aquilonious and Brenner (2015) explored how students reasoned about p-values. They grounded their qualitative study in Tall and Vinner's (1981) theoretical framework, which encoded concept definitions and concept images. Aquilonious and Brenner were interested in the concept definition and images of the significant level and the p-value. The sample size consisted of 16 community college students who were videotaped and interviewed. The students were given hypothesis test problems to solve within two hours. The analysis consisted of searching for the emergent themes based on grounded theory (Charmaz, 2002). The results showed that students lacked an understanding of p-values as an independent concept, even though they knew how to use the p-values for making

their statistical decisions as long as they compared it to an alpha value. This finding supports the argument that procedural knowledge does not translate into conceptual understanding. In fact, the students did not seem to consider the intimate connection between the concepts of p-value and sampling variability. None of the students could give a concept definition of the p-value without consulting their notes (Aquilonious & Brenner, 2015). However, this study could have been stronger if they had collected information on students' demographic characteristics and performed a cluster analysis, as well as incorporated a treatment group with a randomization-based curriculum in order to explore students' reasoning about p-values for both groups.

Lane-Getaz (2017) conducted a pretest-posttest design using the reasoning about p-values and statistically significant instruments in order to measure inference learning outcomes in an introductory level statistics course. The instrument contained 36 questions that assessed conceptual knowledge from across the domain of inference learning. Eighteen items measured whether the respondents recognize basic concepts. Twelve items measured if they differentiated connected concepts. Five items measured how the respondents interpret inferential results and two items whether they can evaluate procedural and inferential validity. The study was conducted at a small liberal arts college in the U.S. Upper Midwest with a sample size of 69 students majoring in psychology and sociology. The design did not include a control group and all 69 students were exposed to 55 minutes' lab sessions once a week under Cobb's (2007) three R's of inference: "Randomize the data, Repeat the process, and Reject models that put your data in the tails of the null distribution" (Lane-Getaz, 2017, p. 6). The results showed that most students achieve statistically significant gains for 20 items out of the 36.

One persistent challenge was estimating and shading the p-value region due to an underlying misconception of one-side hypothesis testing when given evidence in the opposite direction than hypothesized. Such challenges suggest that these cases require greater detail when teaching with dynamic simulation tools. One potential issue with the study is that the instrument only measured one inference learning outcome domain, thus leaving other domain areas of inferential statistics like confidence interval out of the picture. Finally, the reported Cronbach's coefficient is 0.69, which was slightly below the acceptable value in the research community of 0.70 and above (DeVellis, 2012, p. 110).

The misconceptions of variability are present even in students with training in statistics (Chance, DeMas, & Garfield, 2004). It is almost as though there is something unnatural about probabilistic thinking. This could be that students do not have an insight into the likelihood of an event since students' reason from everyday experience (Garfield & Ben-Zvi, 2007). These misconceptions can be directed by certain affordances available in randomization-based environments. Consequently, randomization-based inference can help undergraduate students understand probability concepts by providing them with simulations of random situations (Pfannkuch & Budgett, 2014). Even in the case that students do possess adequate statistical reasoning but fail to identify many situations in probability instruction as random, randomization-based inference can assist students' statistical reasoning.

Randomization-based inference can provide situations of the sampling process and make it easier for students to experience it. Probabilities represent stable properties of likelihood setups, but they do not represent fixed facts (Pfannkuch & Budgett, 2014). Students, when solving problems, learn to use reasoning that does not involve

uncertainty, and this particular scenario misleads their reasoning and thinking.

Randomization-based inference can help students understand probability values because they can reproduce random behavior instantaneously. In some sense, a large part of traditional theory-based inference instruction has led students to think most phenomena as not being random (Chance et al., 2004).

All these aforementioned barriers lead to students' inability to construct an adequate abstract representation of key ideas about statistical inference. These barriers, which we see across all students, can be compounded by some student characteristics, which might have an additive effect, exacerbating statistical illiteracy.

Student Characteristics as Obstacles to Statistical Literacy

Educational researchers have been long concerned in predicting the academic success and retention of college students in general. A special group in the prediction of academic success is at-risk students. At-risk students include ethnic minorities, non-native English speakers, non-traditional age students, and first-generation college students, who typically have low academic performance and higher dropout rates than any other groups (Stebbleton, Soria, & Huesman, 2014). The mathematical abilities tend to be poorer among these populations, as measured by lower critical thinking scores in standardized assessment prior to college (The College Board, 2013). This could compound the barriers to statistical literacy. This research seeks to examine how alternative pedagogies can help at risk students overcome these barriers. The following sections examine these student characteristics.

Non-Native Speaker

Non-native speaker refers to someone who has learned a particular language as a child or adult rather than as a baby (Cambridge Dictionary). In the Hispanic student population, non-native speakers tend to perform lower because of “language discontinuity,” which is an extension of the cultural discontinuity theory between their homes and the educational institution (Suárez-Orozco, 1998). This unique situation produces misunderstanding and inferior learning patterns across students (Suárez-Orozco, 1998). Key barriers encountered by Hispanics students that limit higher education completion include limited English proficiency, first generation status, caring for children at home, being single parents, struggling financially, delaying to attend college instead of attending immediately after high school, and working full-time or part-time while attending college (Burns, 2010). In fact, according to Garcia (2000) and Schmid (2001), English proficiency is among the most important predictors of Hispanics’ low educational achievement, since they are more likely to be non-English speakers and to live in low socio-economic status families than any other group.

Cárdenas and Kerby (2012) suggested that in general the Hispanic population will grow to 30% by 2050 as projected by the forecasting models. Consequently, it is essential to address the challenges that Hispanics students are facing in introductory statistics courses and their ability to perform. More research is needed to understand the relationship between these students’ characteristics and statistical literacy attainment. Identifying the mechanisms through which these demographic and academic variables and achievement are key issues of analyses. Then, development of a study model and the

empirical results generated might add to the understanding of best practices and processes for statistical literacy curriculum for Hispanic students.

Students at Minority Serving Institutions. Minority Serving Institutions (MSIs) play a fundamental role in the local and national economy by incorporating human capital and preparing minority students to fill critical jobs in our society. It is also a vehicle for social mobility among members of different social classes. One performance metric of school success for colleges and universities is the retention of students. MSIs depend on the success of all students attempting to achieve a degree. Low retention rates drive up the cost of education through inflated tuition, additional fees, and the increased use of taxpayer money (Swail, Redd, & Pena, 2003). In fact, the Florida State Legislature developed a new funding formula for state colleges and public universities partially based on students' retention and completion since 2014.

One form of MSIs is Hispanic Serving Institutions (HSI), which is an institution of higher education with an enrollment of undergraduate full-time equivalent students that is not less than 25% Hispanic, as per the definition of the United States Department of Education (2013). Laden (2001) adds to the definition by stating that 50% of the student population should be first generation college students.

During the last 40 years, Hispanic students have experienced important educational advances. HSIs serve Hispanic students' need to access higher education, which leads to greater economic attainment and can help break the cycle of poverty. These institutions, officially recognized by U.S. Congress in 1992, are relatively new to the arena of higher education and were developed through demographic enrollment changes mostly in the southern states (Bridges, Cambridge, Kuh, & Leegwater, 2005).

Presidents and chancellors of HSIs acknowledge several challenges their institutions will face if they do not plan ahead. The three main challenges are the lack of funding, poor academic students' readiness, and student retention and success (Santos & Cuamea, 2010).

In order to deal with these issues, it is important to understand the student population and the barriers to retention. Significant educational gaps in metrics such as educational achievement and attainment persist among Hispanics students (National Center for Education Statistics, 2012). First generation Hispanics are less likely to complete college and have lower school attainment levels than do second generation Hispanics. It is important to mention that Hispanics have differences, depending on their country of origin, which come with different characteristics, immigration experiences, language proficiency, as well as educational and poverty levels (U.S. Census Bureau, 2015). Seidman (2005) reported that commonalities among minority students that contribute to attrition are economic challenges, academic readiness, and being first generation.

HSIs, in particular, are essential points of access as they now enroll 64% of all Latino college students (Calderón Galdeano, & Santiago, 2015). Arbelo-Marrero and Milacci (2016) conducted a phenomenological study about the academic persistence of undergraduate nontraditional students at HSIs. The themes identified through the data analysis suggested that key factors that affect students' academic persistence include the interaction that takes place between students and family members, peers, the relationships students have developed with faculty and administrators in the institutional setting,

English language competencies, and students' resilience in overcoming challenges within these factors.

Consequently, researching instructional models that contribute to the retention and completion in introductory statistics courses of minority students will assist colleges and universities in establishing a curriculum that supports retention, especially when statistics is part of the category of gate keeper courses. The effect of teaching and learning inferential statistics with a randomization-based methodology at MSIs has been neglected in the literature. Much of the population described in the literature of simulation-based methods has focused on PWIs (Aquilonius & Brenner, 2015; Chance et al., 2016; Lane-Getaz, 2017; Maurer & Lock, 2016; Pfannkuch & Budgett, 2014). One should not assume that the results obtained from previous studies can extrapolate to a different population. With the increase in the size of the Hispanic population and the growing number of HSIs, the examination of statistics learning represents a significant gap in the literature.

Gender

Schwery, Hulac, and Schweinle (2016) documented that the gender gap emerges early in students' academic studies and widens as students' progress through middle school, high school, and college. Based on data collected in norm-referenced achievement tests, males have long received higher scores than females. From 1972 to 2013, males outperformed females by an average of 40 points on the mathematics section of the SAT (College Board, 2013). However, Schwery et al. (2016) pointed out that the gender gap is closing, and when mathematics achievement is measured using mathematics classroom grades, females now outperform males in the classroom. During

the senior high school years, instructors rate females' mathematics performance higher than males' mathematics performance despite SAT data, where males still outperform females (Schwery et al., 2016). However, on the SAT only 3% of the variance in test scores are accounted for by gender; the remaining 97% seems to be due to natural proclivity, training, and other factors (Cummins, 2014).

Researchers have accounted for gender differences in mathematics courses in different ways. One is the idea that “common cultural discourses associate mathematics with masculinity, making it more challenging for women to identify themselves as effective mathematicians” (Solomon, 2012, p. 173). Pre-existing gender stereotypes differentially influence men and women via implicit social cognitions (Nosek & Smyth, 2011). The stereotype threat that males are better at mathematics than females may result in a reduction in females' academic self-efficacy because females believe that they have inadequate mathematics abilities (Schwery et al., 2016). Students who have lower mathematics self-efficacy are more likely to have unpleasant experiences with mathematics, and given the fact that self-efficacy is a strong predictor of achievement, addressing this stereotype threat becomes important in when teaching statistics. Finally, researchers point out that the lack of female role models in mathematics might impact female mathematics achievement (e.g. Beilock, Gunderson, Ramirez, & Levine, 2010). In college, this could translate into fewer women participating in careers in the field of science, technology, engineering, and mathematics (Nosek & Smyth, 2011). Furthermore, in the social sciences the only contact point between a systematic process of logical reasoning and judgment is statistics, which surprisingly has been given no attention in HSIs, contrary to mathematics, where all the effort has been concentrated.

Nontraditional Students

The construct of the nontraditional student is defined differently across the literature. However, this project defines nontraditional as any undergraduate student over the age of 25. The National Center for Education Statistics (NCES) said: “Age acts as a surrogate variable that captures a large, heterogeneous population of adult students who often have family and work responsibilities as well as other life circumstances that can interfere with successful completion of educational objectives” (NCES, 2011a, para. 1). The NCES reports that degree completion rates at the undergraduate level for students between the ages of 25 and 29 years are 16% for Hispanic students, 20% for African American students, and 40% for Caucasian students (NCES, 2014). Thus, there is a vital need to address the learning needs of an increasing nontraditional college student population. Furthermore, nontraditional students at two-year colleges are far less likely than those at four-year institutions to complete a degree (Brock, 2010).

There are several factors that could account for differences in success for the nontraditional age student. In the 2012 National Survey of Student Engagement, 46% of adult learners reported working at least 30 hours a week, and close to 75% of adult learners at four-year institutions reported caring for a dependent (National Survey of Student Engagement, 2012). Gilardi and Guglielmetti (2011) found that non-traditional students in a non-residential context showed that employment can represent an external restriction for the continuation of undergraduate studies. This is especially true for temporary jobs in which the nontraditional student does not possess any bargaining power and are forced to work even when their schedules shift in the middle of the semester. This discontinuity impacts students’ academic attendance, thus affecting

students' self-efficacy, self-esteem, and implementation of self-regulation, all of which are essential phases of academic success, particularly in mathematics related content (Good, Rattan, & Dweck, 2012).

Brodish and Devine (2009) claim that nontraditional undergraduate college students' lean towards mathematics anxiety testing performance as a result of the stereotype threat associated with age. According to Steele and Aronson (2004), the stereotype-threat construct is a response resulting from a personal belief that taking an exam, or being evaluated in some way, puts one at risk of confirming a negative stereotype about one's group. This situation could be strong enough to sway a student's implicit beliefs of intelligence and could also impact their personal sense that they belong in mathematics courses (Good et al., 2012). In general, the literature points out that nontraditional students tended to exhibit less efficacious beliefs about performance capabilities about testing and math abilities (Good et al., 2012; Hollis-Sawyer, 2011).

First Generation

First-generation college students are usually defined as those whose parents have not earned bachelor's degrees. Continuing-generation students have at least one parent with a bachelor's degree (Stebbleton & Soria, 2012). The number of first-generation college students in the U.S. has gradually raised to about 21% of the student population (NCES, 2011) with only 11% of first-generation students earning a bachelor's degree after an average of six years in college. Ethnic minority and low-income students are often the first members of their families to go to college (Jehangir, 2010). First-generation college students face several academic and social challenges. The parents of these students lack the knowledge of the adjustment process associated with higher

education, leaving the students with few role models to help them navigate the process (Woosley & Shepler, 2011). A rate of 11% college completion for first generation is in contrast to 55% of continuing-generation students who earned a bachelor's degree (Balemian & Feng, 2013).

According to Woosley and Shepler (2011), first-generation students arrive at the college culture with poor academic preparation and poor educational resilience. They become aware that "their high school curriculum was less than rigorous and that the academic expectations inherent in baccalaureate programs can be somewhat overwhelming, resulting in self-doubt." (Woosley & Shepler, 2011, p. 704). Some of the academic obstacles that first-generation students experience include that they are less likely to take college level courses in high school and displaying lower average scores on standardized admission exams and critical-thinking assessments (Balemian & Feng, 2013). These obstacles are compounded by the fact that most first-generation students' work and they have family obligations that can compromise their academic and social integration on campus. Overcoming these academic obstacles is crucial since first semester grades and self-reported confidence in mathematics predict college persistence for first-generation students (Dika & D'Amico, 2016).

When the challenges of first-generation college students are added to issues of gender, non-traditional age, and being a non-native speaker, the odds of Latinos succeeding through college graduation is lowered (Longwell-Grice et al., 2016). Therefore, understanding the pathway in which all these demographic characteristics that define the population for the study interplay becomes important when analyzing the

findings of the effects of a randomize-base inference methodology in an undergraduate introductory statistics course.

Best Practices for Statistical Literacy

The role of randomization-based inference in teaching and learning inferential statistics has been explored to bring potential solutions to the most current identified problems in statistical literacy (Chance et al., 2016; Lane, 2017; Maurer & Lock, 2016; Tittle et al., 2011). These roles include the use of simulation as a dynamic visualization tool that reduces the computational load; an alternative for developing and communicating statistical reasoning; and a vehicle for assessment of student learning (Budgett, 2013). In addition, randomization-based technology can bring into the statistics classrooms' rich real-world problems with statistical applications and give students and professors more opportunities for feedback and reflection (Novak, 2014).

One of the best practices, suggested by Budgett and Wild (2014), is shifting the focus of statistics curricula from mathematical calculations to activities of applied nature by providing students with problems in different contexts thus simulating real life events. The same line of suggestions was proposed by Garfield and Ben-Zvi (2008). They advocated for active learning and introducing activities where students can construct knowledge by engaging them in data collection, reflection, and exploration of statistical concepts. Finally, Mills (2002) advocated for simulation-based programs which allow students to explore statistical concepts in discovery-world environments by providing examples that have appeared in the media, government reports, and news. These strategies combined to support the postulate that statistical literacy is highly correlated with the implementation of a randomization-based curriculum and should become a focus

of statistical education. However, according to Gould, Davis, Patel, and Esfandiari (2010), the ease of computations by using randomization should not disguise students from the theories and assumptions of the statistical analysis that they are using because it could seem like students think the re-randomization distribution was the observed data distribution. This is the reason why an important component of this research study will measure the students' ability to reconcile randomization-based inference with traditional theory-based inference.

Several key research articles support the use of randomization-based inference methodology as a best practice for achieving statistical literacy. Tintle et al. (2011) compared traditional algebra-based introductory statistics courses that used asymptotic testing as the underlying base for inferential statistics to a randomization-based approach to statistical inference. In this study, the randomization-based curriculum emphasized the core logic of statistical inference by using randomization tests, by presenting sampling distributions in an intuitive manner based on randomization test, by introducing confidence intervals as an outcome of test of significance rather than the other way around, and by de-emphasizing descriptive statistics techniques. The sample was 240 students from Hope College, a liberal arts college in Holland, Michigan.

Tintle et al. (2011) measured statistical literacy with the CAOS test. The design of the experiment was a pretest posttest ANOVA analysis in order to compare aggregate changes in the CAOS scores between the groups and with respect to the national sample, while matched pairs t-test were used for significant learning gains. Additionally, differences in item-level posttest scores were evaluated by pretest using a logistic regression model predicting whether a student correctly answered the questions on the

posttest. Finally, an item level analysis used a Bonferroni correction to set the significance level at 0.05 over 40 (number of questions from the instrument). The results suggested that there was a significant improvement in students' knowledge of test of significance, simulation, and the purpose of randomization. However, there were several flaws in the study. One was the absence of students' characteristics as a predictor of achievement in the simulation-based inference curriculum. The second was at the instructor level, since the researchers did not control the confounding instructor effect on each curriculum. Finally, they did not address the assumption of independence of errors between students, which was likely violated because learning outcomes for students attending the same lectures were related. This is a problem for all comparative studies in education where measurements are collected at the individual student level but the pedagogical intervention is given at the class level, creating sufficient doubts with respect to the statistical significance of the results and their generalization to the population at large (Maurer & Lock, 2016).

The second article is a follow up to the first one. Tintle et al. (2012) explored retention of statistical concepts from a cohort of students who took the randomization-based curriculum to a cohort of students who used the traditional theory-based curriculum. Tintle et al. (2012) addressed a remaining question: whether the learning gains observed were only temporary, or the randomization-based inference fostered conceptual knowledge to continue with the students after the course ended. This investigation concentrated on student retention as the difference between posttest and four months post course scores on the CAOS test. The sample size for the traditional theory-based approach and the randomization-based were 78 and 76 respectively.

Similar to Tintle et al. (2011), Tintle et al. (2012) performed aggregate comparisons of the two cohorts by using pair t-tests for changes in score over time, and a multiple regression model predicting each student posttest CAOS score by pretest score, participation status, four-month retention period, and cohort. The researchers also ran interactions between cohort and four-month retention status. The model yielded a p-value of 0.012 on the interaction term. A logistic regression model predicting gender by participation status and cohort was run, including an interaction between participants' status and cohort. The model yielded a p-value of 0.91 indicating no evidence of different gender participation rates between cohorts. However, this study would have been stronger if they had considered a multivariate analysis with a larger sample that include student level factors that influence retention in undergraduate introductory statistics courses.

Chance et al. (2016) measured student performance in a simulation-based inference curriculum in a college-level introductory statistics course from a multi-institution assessment effort by instructors, some for the first time. The researchers examined several pretest and posttest measurements of students' attitudes through the Survey of Attitudes Towards Statistic (SATS) instrument and a 30-question concept-based inventory. The latter instrument was developed by the researchers by adapting, extending, and using questions from the Goals and Outcomes Associated with Learning Statistics (GOALS) and CAOS test. This was one of the most comprehensive research projects ever conducted in the field of statistics education. Their design applied hierarchical modeling to the data from student level variables like GPA, major, grade level, number of previous math courses, age, type of degree seeking and instructor level

variables like gender, tenure status, years of experience, and length of the class. The developers of the curriculum and intervention were advisors of the Randomization Base Curriculum Developers (RBCD), a group which was the result of a National Science Foundation (NSF) grant from.

However, the fidelity of treatment was not possible since this study collected samples from different institutions including four high schools, two community colleges, 25 four-year colleges, and six research universities. Instructors varied in the implementation of the attitude and concept instruments, particularly with the level of incentives provided to students. Some instructors offered quiz credit, some homework credit. Another group offered just extra credit, and finally, there was a group that offered nothing. Additionally, the concept inventory test lacked the proper validity and reliability testing. In fact, after the study concluded, the researchers changed the concept inventory test, which was suggested as the next step. Finally, there were only two groups of traditional theory-based methods classes, which created an unequal sample size among the groups. The results showed that the level of prior knowledge a student brings into the course and how confident they feel about their ability of learning statistics were strong predictors of achievement, regardless of instructors' characteristics.

According to Garfield (2007), ready-to use simulations have become increasingly popular. They are easy to access and use, and they require little prior knowledge on the part of learners. These emergent simulations are extremely helpful since the norm is to dedicate six to eight weeks of the semester to the topic of statistical inference, which might not be sufficient time to develop the concepts. However, they face challenges. The first one is that the origin of the representation, the process and algorithms, and the

connections between users' actions and numerical changes are not evident in the situation. In other words, the interaction between parameters and graphics is unclear for the students, if the mapping is not performing adequately (Velleman & Moore, 1996).

In recent years, a group of statistics research educators have developed and used randomization-based approach for inferential statistics. Nonetheless, the statistics reform movement on students' conceptual understanding has fallen short on permeating college classrooms (Chance et al., 2016; Lane, 2017). delMas (2014) explored the trends in students' conceptual understanding of statistics from 23,645 respondents over a nine-year period. He found that the CAOS factor scores have been stable with a possible increasing trend in mean scores for the questions assessing students' understanding of test of significance (delMas, 2014). One critique of the study is that because the sample was not collected specifically to measure the effects of the simulation-base inference approach the results do not indicate that students' statistical understanding has or has not increased over the study period.

In general, dynamic visualization tools for inference seems to help students understand the components and what each graphical representation means. Such visualizations work better when combined with verbalizations and language that assist students to reason and argue from the simulated data (Budgett et al., 2013). However, educational tools for simulations inference-based methods are fairly new in undergraduate introductory statistics courses and therefore more data needs to be collected to measure the effects on the different populations.

The Gaps in the Research

This review of the literature aimed to examine several components of statistical literacy in order to improve the pedagogical approach and the students' ability to master the learning outcomes associated with an introductory statistics course in light of emergent changes, explicitly the trend to include randomization-based methods. As more instructors start to incorporate a randomization-based curriculum, they want to know how statistically literate their students are through the new methodology.

But as this literature review demonstrates, there are several gaps in the literature about inferential statistics education. The first gap relates to the population in which researchers generalize their findings. In the two most recent articles about simulation-based inference by Tintle et al. (2011) and Tintle et al. (2012), the samples come from Hope College, a single Midwestern college, and do not represent a diverse pools of college students. Tintle et al. (2012) acknowledged that “the sample was biased towards better students” (p.30). Chance et al.’s (2016) sample was only diverse at the institutional level (high school, community colleges, colleges, four-year colleges, and research universities) but it was not diverse for the students’ demographics characteristics. Finally, in Maurer and Lock’s (2016) work, the sample came from an undergraduate introductory statistics course from Iowa State University.

There is a dearth of research on randomization-based methods in MSIs and their unique set of characteristics that define this population of students, including at HSIs. At HSIs, a critical component in an introductory statistics course is English language proficiency, which is necessary in order to understand inferential questions that are highly embedded in word problems. Perhaps a randomization-based approach can

mitigate the burdens faced by non-native speaker students. There are no studies currently looking at HSIs. This is something that needs to be addressed since the Hispanic population represents 18% of the U.S. population and it is projected to grow by 57% from 2015 to 2050 (Krogstad, 2014)

Finally, the research design used in these studies could be improved. Even though the majority of them used a pretest and posttest for students who completed an introductory statistics course (Chance et al., 2016; Maurer & Lock, 2016; Tintle et al., 2011; Tintle et al., 2012), they either lack a high-quality assessment with validity and reliability measures, or they lack demographic information from students in order to explain variability within students. In general, these studies answered some thought-provoking questions, but they were just exploring the surface. Only Maurer and Lock (2016) investigated the statistical literacy achievement level by forcing students to mentally reconcile the differences between simulation-based methods with traditional-theory based methods on how each approach obtain confidence intervals and p-values in order to make a judgment about the nature of the hypothesized proposal. Therefore, there is an unexplored area in this type of interventions since the vast bulk of the literature on simulation-based approaches emphasized on the theoretical assistant in shortening the concepts of hypothesis testing (Maurer & Lock, 2016).

This Project

This project operates under the learning model of conceptual change described in Chapter 1. The learning model embraces that students favor their multiple cognitions being consistent with one another because when their cognitions are inconsistent or dissonant, students feel troubled and are encourage to make them consistent

(Bernstein et al., 1994; Liu & Matthews, 2005). The cognitive conflict could be seen as an extension of cognitive constructivism. Lecoutre (1992) argued that poor understanding occurred when students were unable to construct an abstract representation of the situation. However, students succeed in constructing a satisfactory abstract representation for themselves when they were exposed to circumstances that trigger critical reflection, specifically when working with chance models (Lecoutre & Rezrazi, 1998). Tall and Vinner (1981) developed the idea that cognitive dissonance recreated different evoked images simultaneously in a student's mental schema until proper construction builds the full concept image.

This project aimed at challenging students to think and reason critically to bring about inferential statistical reasoning that was deep and conceptual. The instructor's aid in order to scaffold student understanding of: 1. Randomness, 2. Estimation of parameters with the use of the standard error from the bootstrapping distribution, 3. The percentile approach from the bootstrapping distribution, and 4. Students performed randomization test in order to learn hypothesis testing. These main simulations helped to ignite critical reflection once students encountered the distributional-theory based approach with their respective meaning implicit in the z-table and t-table.

Because of the lack of knowledge about teaching statistics to undergraduate Hispanic students, this project added to the body of the literature about how this population performs under both a traditional and a randomization-based approach. It was expected that the random based approach would increase student outcomes because a high proportion of this population are non-native speakers, which hinders understanding

of the behavior of probability density functions for continuous random variables due to the dense narrative involved in solving each problem.

In order to test the differences between traditional and randomized based inference, this project followed the instructional sequence of the textbook *Statistics: Unlocking the Power of Data* by Lock et al. (2017). In the first seven weeks of instruction, the pedagogy and content remained the same as the control group. During weeks 9, 10, and 11, a randomization-based inference methodology was introduced to the experimental group followed up by the traditional-theory based inference methodology in weeks 12 through 15. This intervention was modeled from the design of Maurer and Lock's (2016) research article. The bootstrap simulations for confidence intervals and randomization tests for hypothesis testing operated as stimuli for conflict to the traditional-theory based, which was based on asymptotic tests (See Table 2 for teaching sequence).

Table 2

Curricula Design

	Traditional Curriculum Weeks 1-7	Randomization-Base Inference Weeks 9-12	Theory-Base Inference Weeks 13-15	
	1. Introduction	9. Sampling Distribution	12. Normal Distribution and CLT	
CAOS	2. One Variable Descriptive	10. Bootstrap Confidence Intervals	13. Inference: One Proportion	CAOS
PRE TEST	3. Two Variable Descriptive	11. Randomization-Based Test	14. Inference: One Mean	POST TEST
	4. Linear Regression		15. Inference: Two Mean	
	5. Experimental Design and Sampling Methods			
	6. General Probability 7. Binomial Distribution			

Students from the control group and experimental group were exposed to identical content and pedagogy for all non-inference related topics in the course in an effort to control for teacher level factors so the study could focus on the student level for the analysis. By the end of the 16 weeks, all groups were to have covered how to conduct inference using asymptotic tests. However, the experimental group learned the fundamental concepts of inference using randomization-based approaches prior to the learning the traditional-theory based inference methodology.

Summary

The literature review aimed to explore the challenges towards statistical literacy among undergraduate students in an introductory statistics course. The literature review explored the best practices, technologies, and instruments available in order to measure learning outcomes associated with a randomization-based inference approach as more instructors start to incorporate simulations in their curriculum. The next chapter describes in depth the design of the experiments in order to conduct the study of randomization-based inference methods while triggering critical reflection.

CHAPTER 3:
METHODOLOGY

Methods and Research Design

The methods and the statistical techniques adopted in the study are described herein. The population, sample, sampling design, instrumentation, and the instructional strategy are defined in order to advance the study. In addition, the procedure for data collection and analysis are outlined with the required tools for the study. The research examined the effectiveness of a randomization-based inference teaching model in an undergraduate introductory statistics course on student achievement of statistical literacy. The subjects were students at a large HSI college in the South.

The study utilized a quantitative methodology. Quantitative methods make use of numerically measured variables and statistical techniques to help you test hypothesis. (Little, 2013). Quantitative approaches are aligned with deductive reasoning that aims to generalize the findings from a sample to a larger population (Little, 2013). When conducting research in the social sciences, in education, or with people in general, random assignment is almost unfeasible because researchers are studying groups that are naturally occurring (Little, 2013, p11).

The study explored an instructional strategy used with undergraduate students in a class setting. Because it was not possible to randomly assign students to classes, this project used a quasi-experimental design. The design looked like an experimental design but lacks random assignment, as a result groups were not probabilistic equivalent (Trochim, Donnelly, & Arora, 2016). The most distinguishable quasi-experimental

designs are the non-equivalent group design (NEGD) and the regression-discontinuity design (Trochim et al., 2016). For the present study, the NEGD was the best fit because the study involved two groups (control and experimental) for comparison. Controlling for the conditions that made the two groups equivalent was unachievable (Trochim et al., 2016). The pretest-posttest NEGD involved selecting a control group and experimental groups, upon which a response variable, statistical literacy, was tested, without any random pre-selection process.

The students enrolled in the statistics courses could not be randomly chosen, as mentioned before, so the primary characteristic of a true experiment was infeasible for this study. Therefore, the addition of the pretest measure helped to address the problem of assignment bias that exists with all nonequivalent group research (Trochim et al., 2016). Although the addition of a pretest to the NEGD reduces some threats to internal validity, it does not eliminate them completely. There are six categories of time related threats with any NEGD: history, maturation, testing effects, instrumentation, regression, and mortality that can potentially threaten the internal validity. In order to minimize the influence of these threats, this study had a comparison group, the instructional strategy started after the deadline for students' academic withdrawal, the way that the data was collected in the pretest was the same as the posttest, and the results were interpreted in the context of the study.

The research measured the changes that occur in the statistical literacy of undergraduate college students in an introductory statistics course with the use of randomization-based simulations. The predictor variable was the use of a randomization-based teaching methodology in the experimental group. The control group was instructed

under traditional theory-based teaching, allowing the two groups to be compared in statistical literacy scores.

The effect of covariates was also considered in the students' achievement of statistical literacy. The covariates included gender, first-generation status, traditional or nontraditional age student, native or non-native speaker student, prior experience in statistics, student status (part-time or full-time), STEM or non-STEM students, highest math class taken, Hispanic, financial aid, and GPA. Statistical literacy scores of the students for each group were measured, using the Comprehensive Assessment of Outcomes in Statistics (CAOS) test. The variables were subjected to statistical analysis, specifically multiple linear regression analysis, and quantitative methods were appropriate for the study.

Population and Sampling

The population for the present study was undergraduate college students attending a large HSI in the Southeast of the U.S. Participants were recruited from Miami Dade College (MDC), the largest HSI in the United States (Miami Dade College, 2014). Miami Dade College had the largest undergraduate Hispanics student enrollment in the U.S. (71% of the credit-enrollment). Year-to-year credit enrollment for the college has remained for the past three years at around 92,000 students. The student population had 54% who are native English speakers, 38% who were native Spanish speakers, and 33% of credit students were of a nontraditional age if 25 years and above. The student body was 58% female versus 42% male. (Miami Dade College, 2014). Miami Dade College has had an open-door policy where no SAT or ACT scores are required and only 32% of incoming students tested as college ready. Finally, 72 % of students were U.S. citizens

(Miami Dade College, 2014). Therefore, MDC was good source of the population for this project.

Sample

The sample size was determined through an *a priori* power analysis using the G*Power software version 3.1 (Faul, Erdfelder, Buchner, & Lang, 2009). The statistical test selected was a linear multiple regression: fixed model, R square increase (F test). Five factors were considered in the power analysis: effect size, significance level, power of test, number of tested predictors, and total number of predictors. Significance level refers to the tolerance level of rejecting a true null hypothesis. The power of test refers to the probability of rejecting a false null hypothesis. In most quantitative studies, the significance level is set at 5%, indicating that there is up to 5% potential risk of concluding that a difference between groups exists when there is no actual difference (Field, 2013). A 95% power of test was used because it is considered to be statistically powerful (Faul et al., 2009) and, for this case, does not require an overly large the sample size. The effect size indicates the estimated degree of relationship between the predictors and the criterion variable. Effect size is normally categorized into small, medium, and large. The effect size for multiple linear regression is usually measured by Cohen's $f^2 = r^2 / (1 - r^2)$ and it is interpreted as 0.02=small, 0.15=medium, and 0.35=large (Cohen, 1988). Therefore, using 5% significance level, 95% power of test, medium effect size, and linear multiple regression with one main predictor: group membership, the minimum *a priori* required sample size was 89. That is at least 45 students for each group (traditional theory-based and simulation-based). In order to account for missing data and attrition of participants, the number of students was increased to 90 per group at the beginning of the

semester. The final sample was 70 for the simulation-based group and 55 for the theory-based group. Therefore, the study's total sample size of 125 was above the minimum a priori total sample size of 89. While the sample was not randomly selected and the strategy was not randomly assigned, the researcher inferred that it was representative of students who enrolled in an introductory statistics course.

Sampling Design

The sample size was achievable by selecting two sections, with each one having up to 90 students potentially enrolled in the section. Two instructors agreed to use their courses for the purposes of this study, with each instructor teaching one of the methods. The design aimed to reduce teaching variability, since only two instructors were involved. Furthermore, both instructors were tenured, and they had a minimum of seven years of teaching experience with the course. Both instructors had the same average positive response feedback rate on the students' evaluations and they met the accreditation criteria by the Southern Association of Colleges and Schools (SACS); which meant having earned a master's degree with no less than eighteen graduate credits in the field of statistics. All classes covered the same topics in accordance with the Florida Department of Education guidelines. In Florida, Elementary Statistics (STA2023) was a 3-credit course that consists of 130 minutes of instruction per week. The prerequisite for STA2023 at MDC was either Intermediate Algebra (MAT1033) or Mathematics for Liberal Arts I (MGF 1106). A passing grade in any of these courses was considered the baseline knowledge in order to enter in a STA2023 course.

The course description for this class, STA2023 Statistical Methods 3 credits, as found in the college catalog is the following: "This course will introduce students to

statistical methods. Students will learn topics to include collecting, grouping and presenting; measures of central tendency and dispersion; probability; testing hypotheses; confidence intervals; and correlation. Prerequisite: MAT 1033” (Miami Dade College, 2016, p. 236).

Students enrolled in the dual-enrollment program were not included in the analysis. Even though it was a small number of students, they were under 18 years of age, and the inclusion of data from dual-enrollment students that generally have taken several Advanced Placement (AP) courses might produce results not representative of the general undergraduate Hispanic college population. They were identified by a screening question in the survey and if they answer positively to under 18 years of age, they were redirected to “thank you for participating” end of the survey. Students who were repeating the course were also screened out at the start of the survey. These groups of students were removed from the data in order to optimize the internal validity of the study.

Ethical Considerations

All the participants were treated in accordance to the ethical guidelines of Florida International University and MDC’s Institutional Review Board (IRB). All potential participants were asked to read the electronic consent form via SurveyMonkey. The consent form adhered to the Florida International University consent form template. Only those participants who agreed by clicking on the agree button were allowed to participate in the study. The data collection process ensured that the participants understood the purpose of the study, responsibilities as participants, and risks involved in participating. Furthermore, participants were informed that they could withdraw from the study at any point without any negative consequences on their part.

In any research, confidentiality of participants and the data are of utmost importance. Confidentiality of participants was ensured by not sharing any identifying information such as their name. Student IDs were used to link participants across the data collection process, allowing the pretest, posttest, demographic and academic questionnaire to be linked to a single individual. Without these provisions, the level of improvement could not be tracked. Participants were assured, as indicated in the informed consent form, that the researcher securely kept all data gathered from them. Electronic files of the data and other related information were kept in a password-protected drive. All electronic files were kept secured until the completion of the study. After this timeframe, all files were permanently deleted. Data from the participants who chose to withdraw in the middle of the study followed the same method of disposal.

The CAOS test had been approved by the University of Minnesota Institutional Review Board for use by human subjects.

Instruments

Comprehensive Assessment of Outcomes in Statistics (CAOS)

The instrument used to measure statistical literacy was the Comprehensive Assessment of Outcomes in Statistics (CAOS) test. The CAOS test is the first and only comprehensive standard assessment that measures all the learning outcomes of the traditional theory-based curriculum that correspond with an introductory undergraduate statistics course (delMas, Garfield, Ooms, & Chance, 2007). The CAOS consisted of 40 online multiple-choice questions, equally weighted, that evaluated students' statistical reasoning "with a focus on statistical literacy, conceptual understanding, and reasoning about variability" (delMas et al., 2007, p. 31). The CAOS test is different than instructor-

designed exams where students are asked procedural knowledge and computation (delMas et al., 2007). These 40 items can be divided into nine topic-based subscales which were “data collection and design, descriptive statistics, graphical representations, boxplots, bivariate data, probability, sampling variability, confidence intervals, and test of significance” (Tittle et al., 2012, p. 13), as seen in Table 3. The entire CAOS test can be completed in approximately 60 minutes (delMas, 2014) (See Appendix D).

Table 3

CAOS Item Numbers by Topic

Topic	Number of Questions	Item Number
Data Collection and Design	4	7, 22, 24, 38
Descriptive Statistics	3	14, 15, 18
Graphical Representations	9	1, 3-6, 11-13, 33
Boxplots	4	2, 8, 9, 10
Bivariate Data	3	20, 21, 39
Probability	2	36, 37
Sampling Variability	5	16, 17, 32, 34, 35
Confidence Intervals	4	28-31
Tests of Significance	6	19, 23, 25, 26, 27, 40

Validity and Reliability. The CAOS test was developed through a “three-year process of acquiring and writing items, testing and revising items, and gathering evidence of reliability and validity” (delMas et al., 2007. p. 31.). In 2007, with the support of the

National Science Foundation (NSF), the CAOS test was launched. During the development of the instrument, a representative sample of undergraduate students was selected from two-year and four-year public and private institutions colleges and universities students. The “average posttest CAOS score was 54%, compared to 44.9% on the pretest” (delMas et al., 2007, p. 34). While this difference represented a statistically significant growth in content knowledge, the outcome showed that students only added 9% to their score at the end of the course.

delMas et al. (2007) reported the analysis of content validity that revealed that there was

100% agreement among 18 expert raters that the instrument measures important basic learning outcomes and 94% agreement that it measures important learning outcomes. In addition, all raters agreed that the CAOS measures outcomes for which they would be disappointed if they were not achieved by students who succeed in their statistics courses. Based on this evidence, the assumption was made that the CAOS is a valid measure of important learning outcomes in a first course in statistics (delMas et al., 2007, p. 373).

The CAOS test was also considered reliable in consistently measuring the construct of statistical literacy. An analysis of the internal consistency of the 40 questions on the instrument, based on a sample of 23,645 students, produced a Cronbach’s alpha coefficient of 0.78 (University of Minnesota, 2016) where in the original analysis it was 0.82 (delMas et al., 2007), which is pretty consistent. Different standards for an acceptable level of internal consistency have been recommended; however, researchers agree that a Cronbach’s alpha coefficient of 0.70 and higher is deemed acceptable (DeVellis, 2012). Therefore, the CAOS test was deemed to have acceptable internal consistency for students enrolled in a college-level, non-mathematical introductory

statistical course. Because the instrument has a pre-existing established reliability and validity, a pilot study did not need to be conducted for the current study.

Measures being used. For this project, the measure being used was the total summated scores from the CAOS test in order to test the effectiveness of the intervention. This data was gathered from the item-level response for each student that collects the instrument and it was provided to researchers with an IRB approval.

The CAOS test is one of the most comprehensive content assessments currently available. However, it is critical to mention that its purpose was to assess the concepts associated with the traditional theory-based curriculum. The new concepts that were incorporated in the randomization-based curriculum, like bootstrapping confidence intervals and a randomization test, were not directly assessed by the CAOS test. For this study, the purpose of the simulations was to aid individuals in their capability to develop multiple representations of the concepts associated with inferential analysis rather than asymptotic test results which were disconnected from real data analysis and inference (Tintle et al., 2012). The ability to manipulate dynamic visualization tools allowed the students to develop a concept image that it was absent in the traditional theory-based approach (Budgett et al., 2013), which was being measured by the CAOS.

Demographic and Academic Survey

A survey was used to collect demographic and academic information from the participants. The collected information was used as predictors for the study, which included age, gender, first-generation status, native speaker, prior-experience with statistics, student status (part-time or full-time), STEM or not STEM students, Race, highest math class taken, financial aid, and GPA. The student consent forms were

completed before the pretest and the survey with the posttest. The CAOS instrument was embedded into the demographic and academic survey via a web link.

The demographic and academic survey was collected via SurveyMonkey. The questions included the following: 1. Gender: Male or Female; 2. Age (enter age in years); 3. First Generation College Students: Yes, or No; 4. Native English Speaker: Yes, or No; 5. Full Time Student: Yes, or No; 6. Major (selected from drop down menu); 7. Prior-Experience with Statistics: Yes, or No; 8. High School GPA; 9. College GPA (enter current GPA); 10. Highest Math Class Taken (selected from drop down menu); 11. Race; 12. What is your major: (selected from drop down menu); 13. Financial Aid: Yes or No; and 14. Current class standing: (selected from drop down menu) (See Appendix D).

Instructional Strategy

The participants for this study were divided into two groups: the traditional theory-based group and the simulation-based group. The theory-based group had, at the end of the semester, 55 students and the simulation-based group 70 students. Furthermore, the two instructors involved in the study had at least than seven years teaching experience with the course in order to reduce the variance in student achievement associated with the instructor. Both curricula comply with the course guidelines set by the Department of Mathematics at Miami Dade College, covering those listed in Table 4. The required textbook for both courses is *Elementary Statistics: Picturing the World*, 6th edition by Ron Larson and Betsy Farber 2015, and students could use scientific calculators. Each group had lecture-based teaching and each one had three types of learning evaluations: weekly homework assignments (20%), a midterm exam (40%), and a cumulative final exam (40%). Even though these evaluations were not

being used in this study, they helped promote a similar amount of studying for both groups.

Students from the two groups were exposed to the same curriculum for all non-inference related topics in the course. Both groups included the same textbook materials, lectures, homework assignments, and midterm exam during the first half of the semester. After the midterm, the content diverged. The design of the sessions was adapted from the research of Maurer and Lock (2016). The following section examines the differences.

Traditional Theory-Based Group – Instructor 1

After the midterm exam, this section progressed through the traditional theory-based approach by learning the normal distribution, sampling distribution, the central limit theorem and the use of the normal tables. Then students were introduced to confidence intervals and hypothesis testing and they were not being exposed to simulations. Table 4 illustrates the curricula design.

Table 4

Traditional Curricula – Theory-Based Inference Expected Time: Spring 2018

WEEK 1	WEEK 2-7	WEEK 8	WEEK 9-11	WEEK 12-15	WEEK 16
STUDENTS ENTRY POINT	2. Designs & Sampling Methods	Midterm Exam	9. Normal Distribution	12. Inference: One Mean.	STUDENTS EXIT POINT
Introduction	3. Univariate Descriptive		10. Sampling Distribution	13. Inference: One Mean & One Prop.	Final Exam
CAOS Pretest and Survey	4. Bivariate Descriptive		11. Central Limit Theory.	14. Inference: One Prop.	CAOS Posttest and Survey
	5. General Probability			15. Inference: Two Means	
	6. Discrete Random Variables				
	7. Binomial Distribution				

Simulation-Based Group - Instructor 2

After the midterm exam, the simulation group began the randomization-based inference approach in week nine by first teaching the concepts of sampling distributions. Then the class used computer simulations and sampling variability as a foundation for exploring parameter estimates using bootstrap confidence intervals and randomization-based tests. Confidence intervals were generated via the standard error by using the bootstrap distribution standard deviation. Also, students were presented with the

bootstrap percentile-based methods. Instructor lectures and homework used the dynamic visualization tool StatKey software package (Lock et al., 2017) to conduct all the simulation-based inference activities. The activities were taken from the companion book for StatKey, *Statistics: Unlocking the Power of Data* (2017), which was developed by the same authors of the software (Lock et al., 2017). In week twelve, the instructor covered normal distributions and how they were used to conduct inferences on means and proportions. In addition, confidence intervals via computing the margin of error were covered. Table 5 illustrates the curricula design.

During weeks 12 through 15, the curriculum came back together for the two classes. By the end of the semester both groups were exposed to how to conduct inference using normal theories and tables. Still, the experimental group learned the core concepts of statistical inference using simulations-based inference methods preceding to the learning of traditional theory-based inference methods. The aim of this design was to trigger cognitive conflict since students had to reconcile both approaches thus deepening the conceptual understanding of inferential analysis.

Table 5

Simulation-Based Group Expected Time: Spring 2018

WEEK 1	WEEK 2-7	WEEK 8	WEEK 9-11	WEEK 12-15	WEEK 16
STUDENTS ENTRY POINT	2. Designs & Sampling Methods	Midterm Exam	9.Sampling Distributions	12. Inference: One Mean.	STUDENTS EXIT POINT
Introduction	3. Univariate Descriptive		10. Bootstrap Confidence Intervals.	13. Inference: One Mean & One Prop.	Final Exam
CAOS Pretest and Survey	4. Bivariate Descriptive		11. Random. Hypothesis Test	14. Inference: One Prop.	CAOS Posttest and Survey
	5. General Probability			15. Inference: Two Means	
	6. Discrete Random Variables				
	7. Binomial Distribution				

The main goal of this pretest-posttest NEGD was to eliminate differences in non-inference related topics to the greatest extent possible. Students were exposed to the same jargon and ideas leading up to inferential analysis. The last official day for students' academic withdrawal was in week nine, which means that the simulations would not play a role with respect to students dropping the class. As a result, it was expected that all students who began the randomization-based inference methodology would complete the course.

The traditional theory-based group met Monday and Wednesday from 2:05 PM to 3:20 PM, and the simulation-based group met Tuesday and Thursday from 5:40 PM to 6:55 PM. Each of the STA 2023 courses were taught twice a week with a class lecture time of 65 minutes, making up a total of 130 minutes per week. Finally, the final exam was equal for both groups since, in accordance with the Mathematics Department guidelines, students needed to compute inferential analysis with an asymptotic test.

Fidelity of Treatment

For both groups, the pretest and posttest CAOS assessment was taken in the mathematics computer lab under supervision of their respective instructor. The pretest was administered at the end of the first week of class and the posttest was given during the last week of class. To be included in the study, the student had to complete the 40 online multiple choice CAOS items in no less than 10 minutes and no more than 60 minutes. The CAOS scored 1 for a correct answer and 0 for an incorrect answer, and all the questions were equally weighted. For the pretest, students received a 100% homework grade for taking the CAOS test, while on the posttest students received a performance-based grade, specifically a 100% homework grade if they scored 70% or higher on the CAOS test. Each student's statistical literacy score was measured using a discrete scale but treated as continuous.

In order to ensure a smooth process, the class instructors introduced the assessments during the first day of class to the students. The study's purpose and procedures were explained and after the first day all students received an email reminder with the date and time in order to take the CAOS test in the mathematics lab. The demographic and academic survey was run through SurveyMonkey and it remained open

until both groups had taken it. A similar procedure occurred with the posttest for each group. Students were informed during the last week of class and they received another email reminder.

Data Analysis

All data collected from the pretest, posttest, and the demographic-academic survey were compiled into a single data file and prepared for data analysis. The data collected from the CAOS test were imported to the Statistical Procedures for Social Sciences software (IBM SPSS Statistics Version 25 for Windows) for data analysis. The online submission of the survey through SurveyMonkey was recorded in spreadsheets that was converted to SPSS documents.

The exploratory analysis included the descriptive statistics for both groups. Since the design was a pretest–posttest NEGD, it was important to explore the bivariate distributions of data points between pretest versus posttest for each group within the same plot in order to explore the selection bias as both groups were not probabilistic equivalent that resulted from the lack of random assignments. Preceding any analysis, a comprehensive data screening was completed to ensure that no outliers distort the data and subsequent analyses. A test for univariate outliers was conducted through a visual inspection of box-plots. A test for bivariate outliers was conducted by checking the scatter plot and inspecting the correlation coefficient.

It was hypothesized that students who received the instructional strategy would experience the benefits of increase conceptual understanding towards mastery of content. An alpha level of .05 was used for all tests. The following section explains the data analysis for each research question.

Research Questions 1 and 2

Since the linear multiple regression was a parametric test, there was a need to first examine whether the data gathered adhered to the statistical assumptions of this type of statistical test. There were 7 assumptions that were checked before running the regression model (Field, 2013).

1. Variable Type: the response variable had to be continuous, either interval or ratio level of measurement. For this project, the response variable was statistical literacy, which was measured by the CAOS test thus the response variable was treated as continuous. Therefore, this assumption was satisfied.

2. Linearity: the relationship model must be linear between the response variable and the predictors (Field, 2013). This assumption was explored first with the matrix scatterplot for linearity diagnostics, allowing the researcher to visually determine if a linear relation occurs between the response variable and each predictor variable. A plot of standardized residuals against standardized predicted value checked the linearity of the model.

3. Independence: all outcome values come from different individuals of the sample (Field, 2013). Observations between groups should be independent, which means the groups should be made up of different people. This assumption was satisfied as the classes did not contain the same students.

4. Independent Errors: for any pair of observations, the residual term should be uncorrelated (Field, 2013). The standardized residual plot against the standardized predicted values helped to determine this assumption. The dots should be scattered.

Therefore, if the dots were scattered, then the errors would meet the assumption of being normally distributed.

5. Homoscedasticity: for each value of the predictors the variance of the error term should be constant (Field, 2013). In the standardized residuals against standardized predicted values, if the dots were scattered, it would mean that the variance of residuals was constant.

6. No Multicollinearity: the predictors must not be highly correlated among them (Field, 2013). This assumption refers to the potential situation if two or more predictor variables are linearly related. If this happens, it increases the standard error of the regression coefficients. Avoiding multicollinearity can be achieved by looking at the correlation matrix, specifically the Pearson correlation values, the cutoff values will be between -0.70 and 0.70. In addition to the correlation matrix, the Variance Inflation Factor (VIF) helped to determine collinearity among predictor variables whereas a $VIF > 10$ was deemed unsatisfactory (Field, 2013).

7. Normally-distributed Errors: the residuals in the model must be random (Field, 2013). The dots from the standardized residuals against standardized predicted value should be scattered. Then it can be concluded that the residuals were normally distributed. In addition, the P-P plot helped to determine if this assumption was satisfied if the dots generally follow the diagonal line.

Question 1. Do students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group when controlling for pre-CAOS test scores?

This research question was answered by running a multiple regression with the main predictor, group membership, while controlling for pre-test. The response variable was the post-test total CAOS assessment score. The researcher looked at the F test and the R-square. The researcher looked at the beta to determine in the model the predictive value for posttest CAOS score. The regression techniques used here were with the enter method.

Question 2. Are students' demographic and academic characteristics—age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, GPA, financial aid, and highest math class taken—associated with statistical literacy scores?

For research question 2, the Pearson correlation matrix determined which predictors were associated with Post-CAOS test scores.

Research Questions 3

Question 3: Do students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group when controlling for demographic and academic characteristics?

Research question 3 was specifically geared toward the group difference while controlling for all the demographics and academic characteristics. The regression technique used here was stepwise, which looked for the best model with significance predictors that accounted for the variance in post CAOS test scores. The response variable was statistical literacy measured in the posttest.

Summary

It is noteworthy to mention that the pretest-posttest quasi-experimental design of this study only allowed for identification of association, not causation. The sample was comprised of students from one college—specifically, Miami Dade College–Kendall campus—who enrolled in an introductory statistics course. The design limited the generalizability of the findings. The two participants’ instructors were debriefed on stereotype threat in order to pursue a healthy learning environment and maximize the impact on students’ achievement.

CHAPTER 4: RESULTS

The goal of the research was to empirically determine whether there was a difference in student achievement in an undergraduate introductory statistics course between students who participated in classes using a simulation-based teaching strategy and those who participated in classes that used traditional instruction in a Hispanic Serving Institution. This goal is paramount since Hispanics represented 24% of Florida's population (Pew Research Center, 2014) and the projections for 2050 is that Hispanics will represent approximately 30% of the U.S. population (Krogstad, 2014). Surprisingly, there is an absence of research on Hispanic students' statistical literacy. Such a lack of research represents a significant gap in the literature given the fact that Hispanics are the largest minority group in the U.S. and their educational attainment has larger implications (Arbelo-Marrero & Milacci, 2016) in the progress of their communities and the economic development of the country. In 2015, the 55 million Latinos living and working in the U.S. were responsible for \$2.13 trillion, an equivalent 11.8% of America's \$18.04 trillion gross domestic product (GDP). By 2020, the researchers estimate that Latinos will fuel nearly a quarter of all U.S. GDP growth which represent 12.7% of the country's total GDP. Therefore, aiding to power that growth will be the number of young Latinos who will be joining the workforce as the Baby Boomers retire (Blanco, 2017). It was hypothesized that students learning statistics under the simulation-based approach would have higher levels of learning beginning statistics than those using the traditional approach as measured by the Comprehensive Assessment of Outcomes in Statistics

(CAOS) test. Chapter 4 describes the results of the testing of this hypothesis. The results from the demographic and academic survey are presented in this chapter in order to answer if there are predictors with a significant contribution to the variance associated with Post-CAOS scores.

Data Cleaning

The data from the Pre-CAOS test and Post-CAOS test were combined into a single dataset. The participants were matched by group (teaching modality), name and student identification number to ensure accuracy. However, some participants were removed from the data as follows: four students were removed for being under the age of 18, one did not complete the pre-demographic and academic survey, and 73 students who did not complete all four instruments (pre and post CAOS test, pre and post demographic-academic survey) were not included in the dataset. This left a sample size of 125 as follows: 70 participants in the simulation-based teaching modality and 55 in the traditional teaching modality for the final analysis. The sample size did not compromise the power of the F-test since the *a priori* calculations for a 95% power only required a total sample size of 89 for one tested predictor (group membership), at a significant level of 0.05, and with a medium effect size of 0.15. However, experimental mortality can become a threat to the internal validity since the attrition across the comparison groups is different. Where such attrition is higher in one group compared to another, it can become more difficult to conclude that the outcome measure, post-CAOS scores, is the result of the teaching modality and not the attrition.

The sample was taken from the population that attends specifically one MDC campus out of eight. The campus is located in the Kendall area from the Miami Dade

County. Therefore, the findings of this study can only be generalized to this particular campus and not to the overall population of students who attend MDC across the county. Hispanic students represent 71% of the Miami Dade College-wide undergraduate student body. The student population had 54% who are native English speakers and 38% who were native Spanish speakers. The student body was 58% female versus 42% male (Miami Dade College, 2014). Race and gender proportions are similar to the sample of this study.

Scale Creation

The CAOS test is the first and only comprehensive standard assessment that measures all the learning outcomes of the traditional theory-based curriculum that corresponds with an introductory undergraduate statistics course (delMas et al., 2007). The CAOS consisted of 40 online multiple-choice questions, equally weighted, that evaluated students' statistical reasoning "with a focus on statistical literacy, conceptual understanding, and reasoning about variability" (delMas et al., 2007, p. 31). These 40 items can be divided into nine topics which were "data collection and design, descriptive statistics, graphical representations, boxplots, bivariate data, probability, sampling variability, confidence intervals, and test of significance" (Tintle et al., 2012, p. 13). The entire CAOS test can be completed in approximately 60 minutes (delMas, 2014). The CAOS test was also considered reliable in consistently measuring the construct of statistical literacy. An analysis of the internal consistency of the 40 questions on the instrument yielded a Cronbach's alpha coefficient of 0.82 (delMas et al., 2007). Researchers agree that a Cronbach's alpha coefficient of 0.70 and higher is deemed acceptable (DeVellis, 2012).

The total CAOS percent scores for each participant who took the test was reported by the researchers at the University of Minnesota who process and manage the CAOS test. A subscale was created by the researcher of this study adding together the 15 questions that were in the three subscales directly impacted by the simulations—sampling distributions, confidence intervals, and hypothesis testing. However, the reliability coefficient Cronbach’s alpha with all 15 questions yielded a value of 0.609. According to Darren George and Paul Mallery (2003), a Cronbach’s alpha below 0.7 is below acceptable.

Descriptive Statistics

Table 6 shows continuous variables collected—age, high school grade point average, and current grade point average for students in the simulation-based group versus the theory-based group. It can be observed that there is no extreme difference between both groups. Table 7 shows categorical variables collected: sample size and respective conversion to percentage of gender, race, first language, class standing, first-generation college students, full or part-time student, financial aid, highest math class taking in high school, and STEM major by class.

Table 6
Descriptive Statistics for Continuous Variables

Class	Age			HS GPA			Current GPA		
	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>	<i>N</i>	Mean	<i>SD</i>
Simulations	70	23	4.51	70	3.26	.51	70	3.17	.53
Traditional	55	22	7.87	53	3.23	.44	52	3.16	.39

Table 7
Descriptive Statistics for Categorical Variables by Class

	Simulations		Traditional	
	<u>N</u>	<u>%</u>	<u>N</u>	<u>%</u>
Gender				
Female	44	62.9	38	69.1
Male & Other	26	37.1	17	30.9
Total	70	100	55	100
Race				
Asian	0	0	2	3.6
Black	3	4.3	2	3.6
Hispanic	61	87.1	43	78.2
White	5	7.1	6	10.9
Mixed/Other	1	1.4	2	3.6
Total	70	100	55	100
First Language				
English	26	37.1	26	47.3
Other	44	62.9	29	52.7
Total	70	100	55	100
Class Standing				
Freshman	7	10	13	23.6
Sophomore	33	47.1	30	54.5
Junior	18	25.7	8	14.5
Senior	12	17.1	4	7.3
Total	70	100	55	100
First Generation College Student				
Yes	22	31.4	25	45.5
No	48	68.6	30	54.5
Total	70	100	55	100
Full- or Part-time Student				
Part-time	24	34.3	11	20
Full-time	46	65.7	44	80
Total	70	100	55	100
Financial Aid				
Yes	46	65.7	33	60
No	24	34.3	22	40

Total	70	100	55	100
Highest Math Taken in HS				
Pre-Algebra	4	5.7	2	3.6
Algebra 1	15	21.4	7	12.7
Algebra 2	20	28.6	25	45.5
Algebra 3	1	1.4	0	0
College Algebra	1	1.4	2	3.6
Analysis of Function	2	2.9	2	3.6
Pre-Calculus	13	18.6	6	10.9
Calculus	9	12.9	3	5.5
College Readiness	1	1.4	2	3.6
Statistics	2	2.9	4	7.3
Trigonometry	2	2.9	0	0
Advanced Math Topics	0	0	1	1.8
Other	0	0	1	1.8
Total	70	100	55	100
STEM Major				
Yes	40	57.1	21	38.2
No	30	42.9	34	61.8
Total	70	100	55	100

In Table 7 it can be observed that almost two-thirds of the participants are females for both groups taking introductory statistics courses. In regard to race, the simulation-based group had 87% and the traditional theory-based 78% of Hispanics. More than 50% reported their native language to be Spanish, and around 50% are sophomores for both groups. Another finding from the sample is that less than 45% are first-generation college students. The findings from the descriptive statistics is close to the overall report from the Miami Dade College that stated that 52% are first-generation college students (Dominicis, 2016). In addition more than 65% are full-time students. However, full-time student status does not imply that the students do not work. According to the Miami Dade College report (2016), 20% of the student body work full-time and 69% work part-time. Finally, at least 60% of the participants received financial aid.

Race, Language and Highest Math Taken in HS Recoding

Language was dichotomized into English versus other, race was dichotomized into Hispanic versus Non-Hispanic, and highest math taken in high school was dichotomized into a group of courses up to and including Algebra II and a group of courses above Algebra II. From Table 8 it can be seen that the majority of the sample was Hispanic with 83.2%. The majority of the students have taken courses up to Algebra II (62.4%).

Table 8

Frequency of Hispanic and Highest Math Taken in HS

	Frequency	Percent
Non-Hispanic	21	16.8
Hispanic	104	83.2
Algebra 2 and below	78	62.4
Above Algebra 2	47	37.6

Table 9 presents a breakdown of the dichotomized variable race and highest math course taken in high school by group membership. In both groups, the majority is Hispanic. A Chi-square test was conducted, and it was found that there is no difference between the two groups $\chi^2(1) = 1.77, p = .183$. It can also be seen from the table that in both groups the majority of the students reported that the highest math course taken was up to Algebra II. In addition, a Chi-square test was conducted, and there was no significant difference between groups in the distribution of the highest mathematics course variable $\chi^2(1) = 1.87, p = .171$.

Table 9

Descriptive Statistics for Hispanic and Highest Math Taken by Class

	Simulations		Traditional	
	<i>N</i>	%	<i>N</i>	%
Hispanic				
Non-Hispanic	9	12.9	12	21.7
Hispanic	61	87.1	43	79.2
Total	70	100	55	100
Highest Math				
Algebra 2 or lower	40	57.1	38	69.1
Higher than Algebra 2	30	42.9	17	30.9
Total	70	100	55	100

Prior-Statistical Experience

Most participants in the sample had not taken a statistics class before as shown in Table 10. According to Chance et al. (2016), prior-statistical experience correlates significantly with higher CAOS test scores rather than higher mathematics course since statistics cannot be placed in the hierarchy of mathematics level courses. A Chi-square test was conducted and there was no significant difference between groups $\chi^2(1) = .480$, $p = .489$.

Table 10

Frequency of Prior Statistical Experience by Class

	No	Yes
Traditional	42 76.4%	13 23.6%
Simulations	57 81.4%	13 18.6%

Pre and Post CAOS Score Descriptive Statistics

The response variable, post-CAOS scores, for each group is normally distributed according to the skewness statistics, as observed in Table 11. Kurtosis for the pre-CAOS score shows there is a pull in the score below 40% as shown in Appendix E. Even though the post CAOS is not normally distributed according to the Shapiro-Wilks test, the need for transformation is not necessary and it is found in other research using CAOS scores. The main assumption of regression is that the theoretical errors are normally distributed; hence the need to use the observed errors. Additionally, if the response variable is not too far skewed, like in the current sample, then regression is still appropriate. Furthermore, by checking the normal p-plot, it can be seen that, for the most part, the residuals are following the line and are not grossly deviating.

Table 11

CAOS Score Descriptive Statistics.

	N	Mean	SD	Skewedness	Kurtosis	Shapiro- Wilks			
				SE	SE				
Pre Total Percent	125	34.96	8.63	1.35	0.22	3.40	0.43	0.91	0.000
Post Total Percent	124	42.08	14.73	1.47	0.22	2.68	0.43	0.88	0.000

Table 12 shows the mean scores of the pre and post CAOS test scores for both teaching modalities. The pretest mean for both groups is almost the same, at around 35%, and it is similar to those found in the literature. However, the posttest mean is different for both groups with an average of 46.89% for the simulation-based group and an average of 36.09% for the traditional group. Therefore, the difference is $46.89 - 36.09 = 10.8$ percentage points.

Table 12
CAOS Pre- and Post-Test Descriptive Statistics by Class

	Mean	SD
Simulations		
Pre-total Percent	35.36	7.7
Post Total Percent	46.89	16.47
Traditional		
Pre-total Percent	34.67	9.83
Post Total Percent	36.09	8.83

T-Test Analysis

An independent samples t-test was conducted to assess if there were differences between the groups prior to providing the simulation-based teaching approach. The T-test analysis was done by testing for differences in the pre-CAOS test scores with the class being the grouping variable. An assumption of an independent samples t-test is that there is homogeneity of variance. That is, it is expected that the variance for each group in each independent samples t-test should be roughly equivalent. A Levene's Test for Equality of Variances was conducted. When Levene's test is significant, it indicates that there are significant differences in the variance, thereby indicating that the homogeneity of variance assumption has been violated. Levene's test was not significant ($F = .127$, $p = .722$), so no adjustment to the degrees of freedom for the t-test was necessary.

The resulting t-test was not significant $t(123) = .579$, $p = .564$. This means that the pre-intervention statistical understanding of students in the simulation-based teaching group ($M = 35.36$, $SD = 7.7$) was not significantly different from that of students in the

traditional teaching approach ($M = 34.67$, $SD = 9.83$). Because there were no differences between the two groups' pre-test, any differences in CAOS scores between the two groups following the teaching modality can be asserted as resulting from the instructional program in the context of a quasi-experimental design (Trochim et al., 2016).

Research Question 1

Do students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group in an undergraduate introductory statistics course?

Ho: The mean CAOS score in the simulation-based group is less or equal than the mean CAOS score in the theory-based group.

Ha: The mean CAOS score in the simulation-based group is higher than the mean CAOS score in the theory-based group.

In order to test this hypothesis, it is necessary that the assumptions associated with a linear regression model are satisfied. There are several assumptions that must be checked before running the regression model (Field, 2013).

Test of Assumptions

1. Variable Type: The response variable is statistical literacy which is measured by the CAOS test thus the response variable is numerical. This test is designed to assign 0 to each wrong response and 1 to each correct response. The CAOS test score was obtained by summing the number of items scored correctly.

2. Linearity: the relationship model must be linear between the response variable and the predictors (Field, 2013). The scatterplot in Figure 1 of standardized residuals against standardized predicted values checked the linearity of the model since the dots do

not show any divergent pattern, fan out, or convergent pattern, funnel in, the linearity is being satisfied.

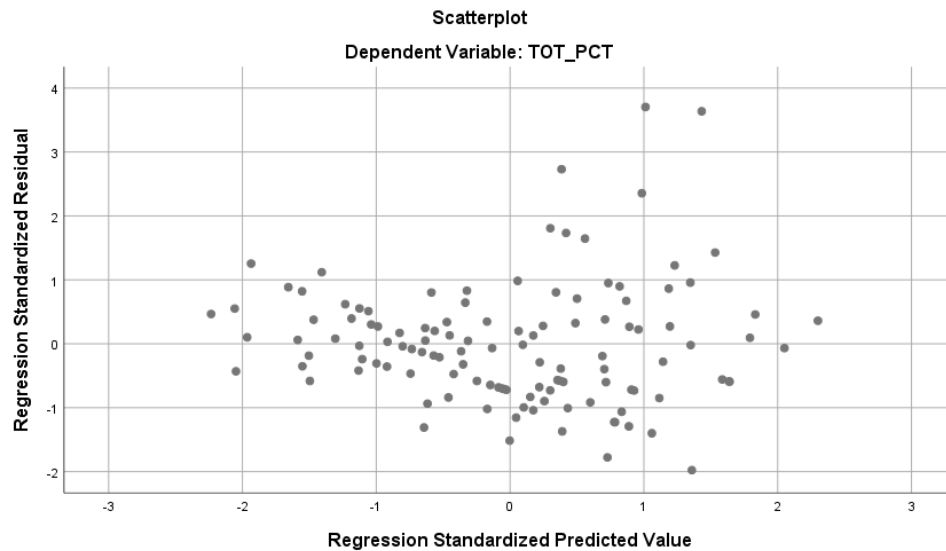


Figure 1: Residuals for the Regression of Predictors on Total CAOS scores

3. Independence: this assumption is satisfied as the two groups do not contain the same students.

4. Independent Errors: for any pair of observations, the residual term should be uncorrelated (Field, 2013). The standardized residual plot against the standardized predicted values helped to determine this assumption. Furthermore, the Durbin Watson = 1.863, showing no auto-correlations.

5. Homoscedasticity: for each value of the predictors the variance of the error term should be constant (Field, 2013). According to Tabachnick and Fidell (2007), the residuals and the variance of the residuals should be about the same to meet the assumption of homoscedasticity. The scatter plot should have an approximate shape of a

rectangular and concentrated around the fit line. The scatterplot of the residuals for this project generally fits this with only a couple of dots outside of the clustering. There is no clear pattern within the residuals, and the plot is generally randomly distributed, meeting the description given in Allison (1999).

6. No Multicollinearity: the predictors must not be highly correlated among them (Field, 2013). This assumption refers to the potential situation if two or more predictor variables are linearly related. If this happens, it will increase the standard error of the regression coefficients. Avoiding multicollinearity can be achieved by looking at the Variance Inflation Factor (VIF) whereas a $VIF > 10$ is deemed unsatisfactory (Field, 2013). All of the VIFs were below 2. (See Appendix E, Table E.1)

7. Normally-distributed Errors: the residuals in the model must be random (Field, 2013). The dots from Figure 1, the standardized residuals against standardized predicted value, should be scattered. Then it can be concluded that the residuals are normally distributed. In addition, the P-P plot helps to determine if this assumption is satisfied if the dots generally follow the diagonal line as shown in Figure 2.

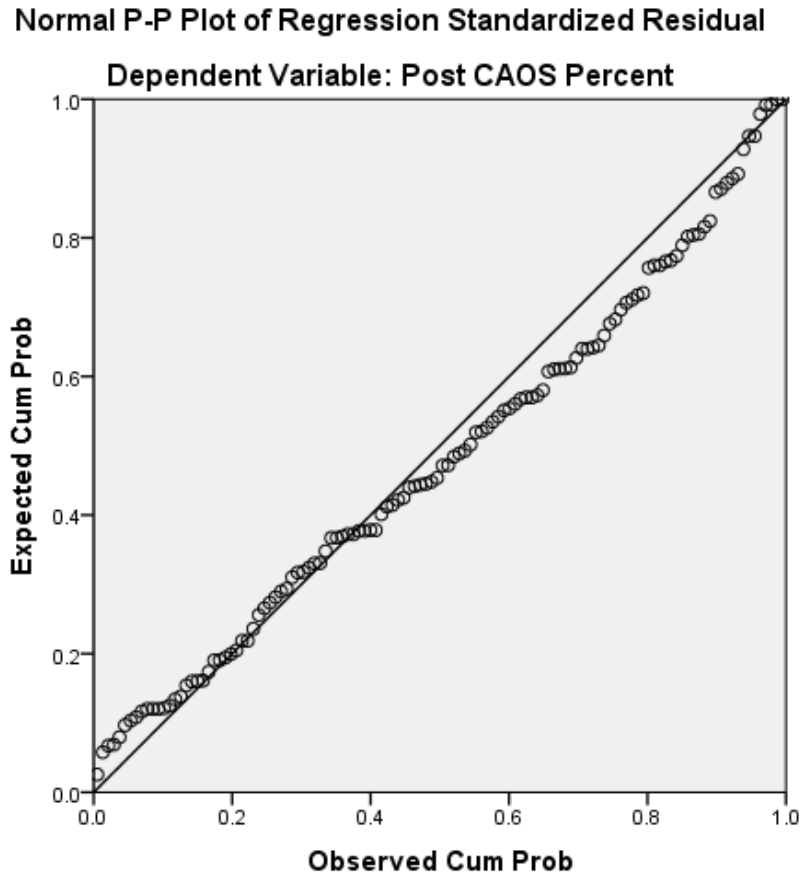


Figure 2: P-Plot of Post CAOS Residuals

Regression

In order to test RQ 1, an enter method multiple linear regression was used to look at the effects of group while controlling for Pre CAOS (Table 13). The only predictors were group membership (simulation-based group vs. theory-based group) and Pre-CAOS scores. This model was found to be a significant predictor of Post CAOS score, $F(2, 121) = 19.20, p < .001$. The coefficient of determination, R^2 , showed that group membership and Pre CAOS accounted for 24.1% of the variance in the Post CAOS scores. Both Group and Pre CAOS scores were significant predictors of Post CAOS

Score. For Group, the simulation-based class had higher scores than theory-based class, even when controlling for Pre-CAOS scores, $B=10.66$, $t(123)= 4.55$, $p < .001$. For Pre-CAOS scores, as you increase the pre-test score by 1 unit, you increase the post-test score by 0.54, $B=0.54$, $t(123)= 4.02$, $p < .001$.

Table 13

Regression of Group, Pre-CAOS test, on Post CAOS Percent

	B	Std. Error	Beta	
(Constant)	16.99	5.01		**
Group	10.66	2.35	0.36	***
Pre CAOS	0.54	0.14	0.32	***
F			19.20	***
df			2, 121	
R2			.241	

Note= N = 124,
*p<.05, **p<.01, *** p<.001

Research Question 2

Are students' demographic and academic background characteristics—age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, highest math class taken, financial aid and GPA—associated with statistical literacy scores?

Ho: Demographic and academic characteristics do not associate with statistical literary scores.

Ha: At least one demographic or academic characteristic associate with statistical literacy scores.

At the bivariate level the Pearson correlation matrix yields significant results for Group Membership, Pre-CAOS scores, and Highest Math taken in HS as show in Table 14. These three factors significantly correlated with the posttest CAOS total percentage scores. Group increased posttest scores $r(123) = .374, p < .001$. Therefore, being in the simulation-based teaching approach group increased the post-CAOS total score. On the other hand, CAOS pretest was positively correlated with CAOS posttest $r(123) = .334, p < .001$. Higher pre-test scores had higher post-test scores. Finally, the highest math course taken increased posttest scores $r(123) = .261, p < .01$. The rest of the predictors were not significantly correlated with Post CAOS Score. At the multivariate level, the question is answered through research question 3.

Table 14

Pearson Correlations

	<u>Post CAOS Score</u>
Age	.00
Gender	-.02
First language spoken	.15
First generation college student	-.15
Full or part time student	.11
STEM major	-.02
Current GPA	.10
Group	.374 ^{***}
CAOS pre-test percent	.334 ^{***}
Taken a statistics class before	.08
Financial Aid	-.05
Hispanic	-.110
Highest Math	.261 ^{**}

*p<.05, **p<.01, ***p<.001

Research Question 3

Do students in a randomization-based inference teaching modality group have higher statistical literacy scores than students in the traditional theory-based inference teaching methodology group when controlling for demographic and academic characteristics?

Ho: The mean CAOS score in the simulation-based group is less or equal than the mean CAOS score in the theory-based group, even when controlling for demographic and academic characteristics.

Ha: The mean CAOS score in the simulation-based group is higher than the mean CAOS score in the theory-based group, even when controlling for demographic and academic characteristics.

The analysis used the stepwise multiple linear regression method. The assumptions were the same as the prior analysis, as the stepwise is still a regression, with the same response variable. All of the VIFs were below 2 (see Appendix E, Table E.1) indicating that there is no multicollinearity, and the Durbin-Watson=1.946, indicating that there is no autocorrelation among the residuals. Figure 3 showed the scatterplot where the response variable is Post CAOS scores. This figure showed a random allocation of Post CAOS scores residuals that take on a rectangular shape with no clustering or systematic pattern. Therefore, Figure 3 showed the assumption of homoscedasticity was satisfied (Tabachnick & Fidell, 2007). Figure 4 showed the normal p-p plot of standardized residuals. Therefore, if the data follow a normal distribution with mean μ and variance σ^2 , then a plot of the theoretical percentiles of the normal distribution versus the observed sample percentiles should be approximately linear (Tabachnick & Fidell, 2007). The resulting plot of Figure 4 was approximately linear; thus, the error terms are normally distributed. Finally, Figure 5 showed the histogram of the residuals. It was used to check if the variance was normally distributed. A symmetric bell-shaped histogram which is approximately evenly distributed around zero implies that the normality assumption was satisfied which is the case of Figure 5 (Tabachnick & Fidell, 2007).

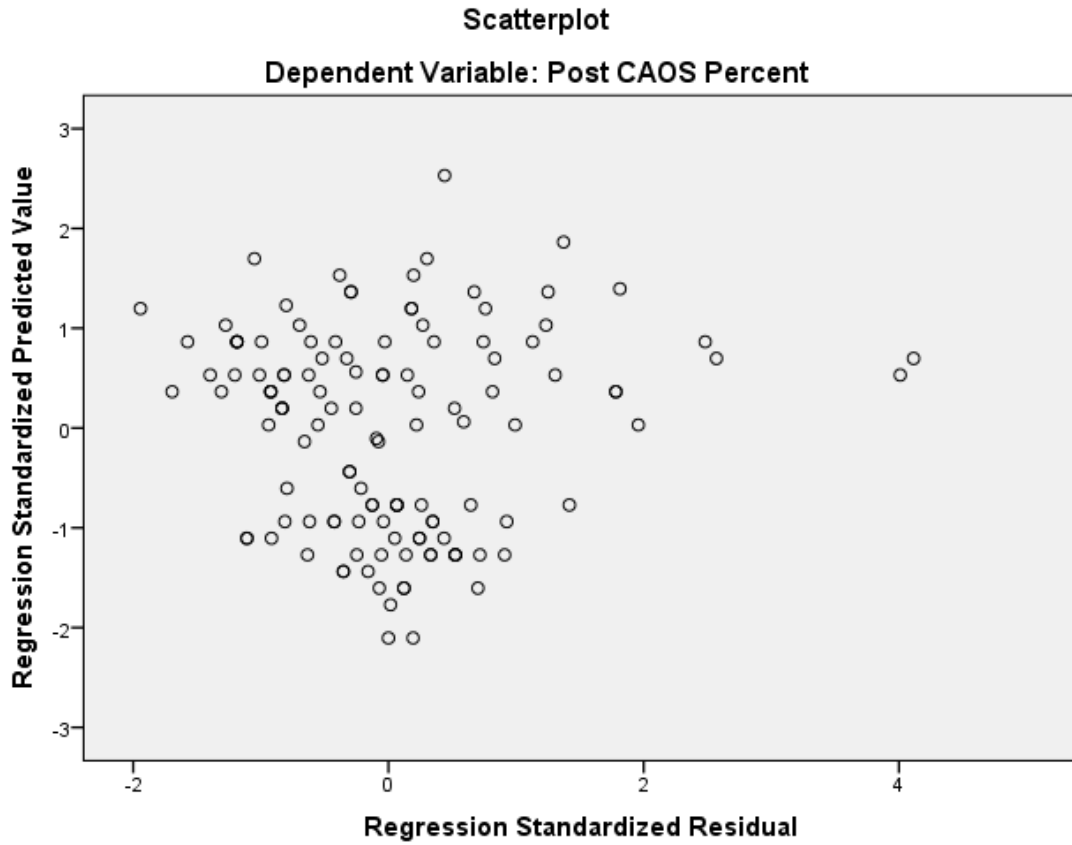


Figure 3: Residuals for the Regression of Predictors on total Post CAOS scores

Normal P-P Plot of Regression Standardized Residual
Dependent Variable: Post CAOS Percent

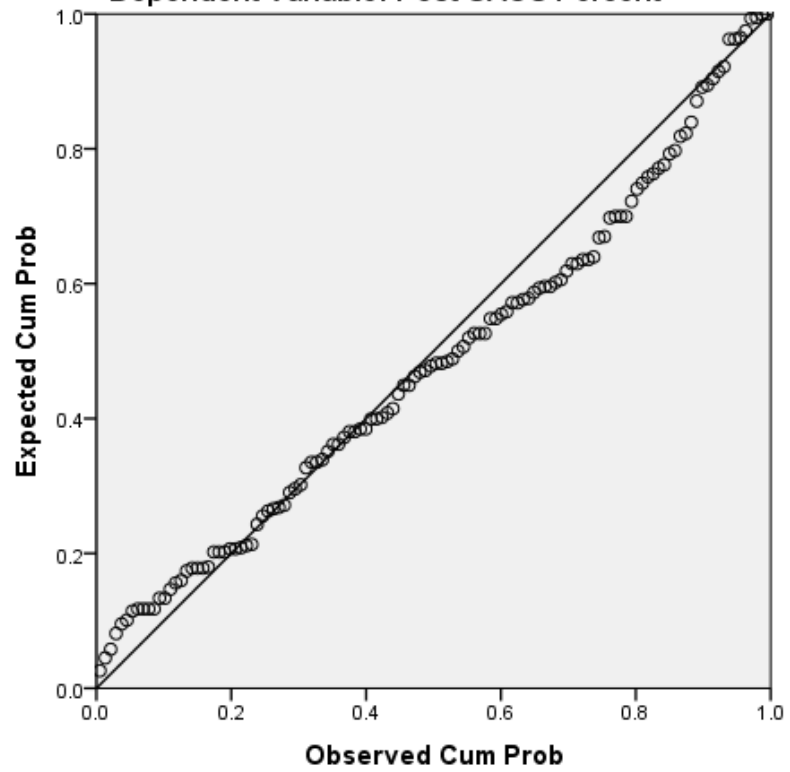


Figure 4: P-Plot of the Residuals

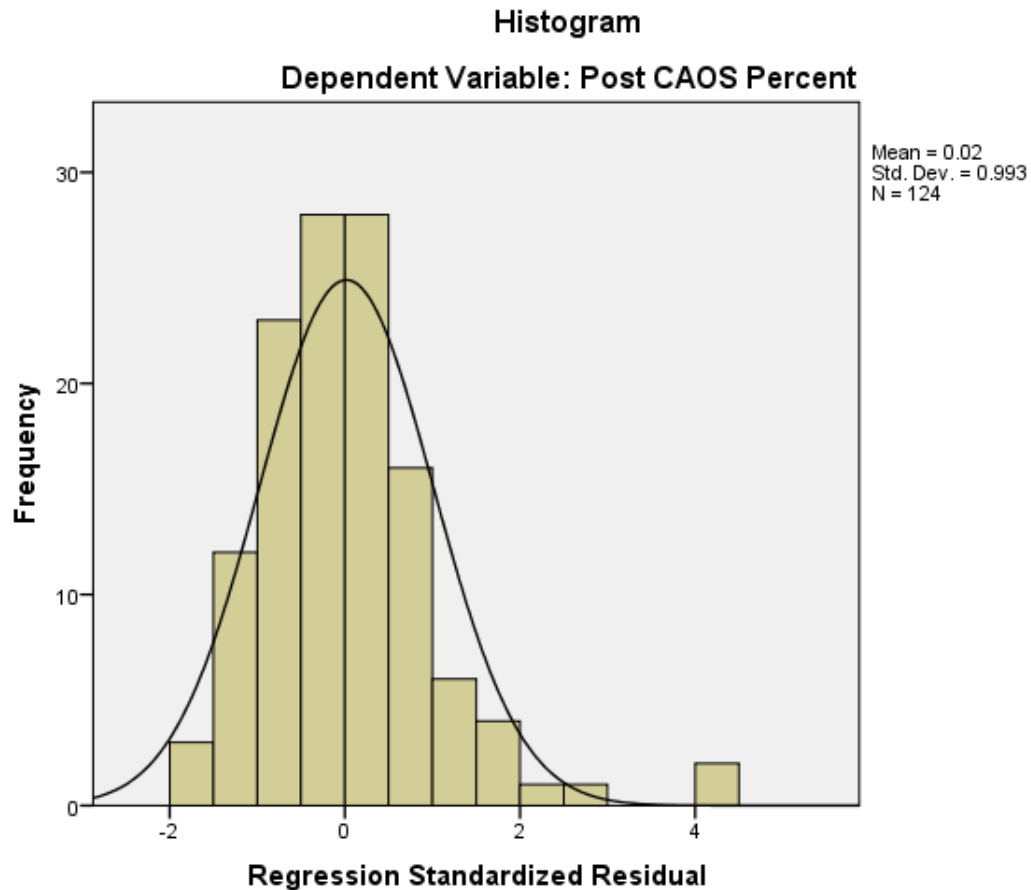


Figure 5: Histogram of the Residuals

Regression 2

The model used the stepwise regression method. The response variable was Post CAOS scores. The predictor variables were group membership (simulation-based group vs. theory-based group), Pre-CAOS scores, age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, Hispanic, highest math course taken in high school, and GPA. The full model showed that only Group and Pre-CAOS scores were the only significant predictors of Post CAOS

scores. None of the other variables were significant. The model was a significant predictor of Post CAOS score, $F(2, 121) = 16.96, p < .001$. The coefficient of determination, R^2 , showed the model accounted for 22.2% of the variance in the Post CAOS scores. For group membership, simulation-based teaching methodology had higher scores than theory-based teaching methodology, even when controlling for Pre CAOS test score, $B=11.22, t(123)= 4.70, p < .001$. For every one unit increase in pre CAOS test score, the post CAOS score increased by .46 units, $B=0.46, t(123)= 3.07, p = .003$. In the appendix section it can be seen the full model with all the predictors.

Table 15

Stepwise Regression of Group, Pre CAOS test, Gender, Language and Control Variables on Post CAOS.

	B	Std. Error	Beta	
(Constant)	19.49	5.41		***
Group	11.22	2.39	0.38	***
Pre CAOS percent	0.46	0.15	0.25	***
F			16.96	***
df			2, 121	
R2			.222	

Note= N = 124,
* $p < .05$, ** $p < .01$, *** $p < .001$

CHAPTER 5

DISCUSSION

Chapter five is divided into five sections which provide a summary of the study, followed by a synthesis and discussion of the results. The first section is a summary of the research problem and the results of the study. The second section is a discussion of the implications of the findings for the practice and how practitioners should apply it. The third section is about the limitations of the study. The fourth section is about the recommendations for future research. Finally, the fifth section is the conclusion based on the research.

Summary of the Findings

The purpose of the present study was to investigate the effectiveness of a simulation-based teaching methodology versus the traditional theory-based teaching methodology for undergraduate students in an introductory statistics course at a Hispanic Serving Institution. The study also examined what factors, if any, contributed to students' success in mastering statistical literacy as measured by the CAOS test. According to Cobb (2015), a simulation-based teaching methodology allows students to comprehend better inferential statistics since sampling variability, confidence intervals, and hypothesis testing require comprehension of the central limit theorem. The idea behind the central limit theorem is abstract for students who tend to limit themselves to reading probabilities from tables that emerge from theoretical distributions rather than distributions emerging from the sample, which is the only information available to solve a problem.

The present study consisted of a quasi-experimental research design, specifically a nonequivalent group design, which involved the collection and analysis of data from a

pretest, a posttest, and an online demographic and academic survey for two undergraduate introductory statistics classes. One group used the simulation-based teaching modality (program group), and the other used the theory-based teaching modality (comparison group). In order to accomplish the design for this research, the researcher surveyed students at Miami Dade College, Kendall campus, a Hispanic Serving Institution (HSI) in the United States during the spring semester of 2018. An online survey was used to collect information about demographic and academic factors that might impact students' success of mastering statistical literacy in order to examine the simulation-based teaching modality effect.

Three research questions were posed in this study. The first research question to be answered was *Did students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group when controlling for pretest?* The second research question was *Which students' demographic and academic characteristics—age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, highest math class taking, and GPA—are associated with statistical literacy scores?* And the final research question was *Did students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group when controlling for demographic and academic characteristics?*

Research Question 1. To answer the first research question, the researcher ran a multiple linear regression with group membership as the predictor of the response variable—post CAOS test scores while controlling for pretest scores. This analysis tested

if students in the simulation-based group performed significantly better than students in the traditional theory-based group as measured by the post-semester CAOS scores. The results demonstrated that students in the simulation-based group perform significantly better than the students in the traditional theory-based group. According to Chance et al. (2016), the cognitive development of statistical ideas like sampling variability, confidence intervals, and hypothesis testing is enhanced when students are exposed to simulations of these statistical concepts. Such concepts can be perceived as mathematical objects that can be manipulated by changing parameters that define it, rather than relying upon theoretical distributions in which students in an introductory statistics course lack the mathematical background to understand where these distributions and probabilities arise (Pfannkuch & Budgett, 2014).

Pfannkuch and Budgett (2014) stated “The sampling distribution can be perceived as an object, a probability distribution with its own properties and as being formed from the process of drawing multiple random samples and recording from each sample a statistic such as the mean” (pp. 1-2). Cobb (2015) observed that students find the mathematical approach to inference obscure, and it is well documented that hypothesis testing is one of the most challenging topics for students to comprehend despite the decision-theoretic framework used, either the Neyman-Pearson or the Fisherian framework (Budgett et al., 2013).

Lane-Getaz (2017), in her a pretest-posttest quasiexperimental study, suggested that with appropriate simulation-based teaching, students in an introductory statistics course can overturn the misconceptions and misinterpretations of the Null Hypothesis Significance Test procedures, p-values, and statistical significance as reflected in the

scores of the Reasoning about P-values and Statistical Significance (RPASS-10) test used in Lane-Getaz's study.

Finally, it is important to mention that in the present research students were induced to reflect on their own ideas that were externalized during the simulations. To attain the reflection of ideas, the researcher exhibited a series of questions during the statistical simulations related to the inferential concepts in order to guide students to manipulate the simulations and to observe the results gradually. The same learning model was used in Liu's (2010) study, where students critically reflect on their ideas; if they did not understand the concepts after following very specific activities, they were asked to repeat the entire phase.

The results from research question one support that the use of bootstrap confidence intervals and a randomization test as a teaching modality resulted in better scores for the group of students who received this type of instruction when compared to the traditional-theory based.

Research Question 2. The second research question posed in the study was which students' demographic and academic characteristics—age, gender, first generation, prior experience with statistics, student status (part/full time), native speaker, STEM or not-STEM major, highest math class taken, financial aid, and GPA—were associated with statistical literacy scores? The aim of the question was to explore the association between students' demographic and academic characteristics and the CAOS posttest to determine if there were significant correlations. The results from the Pearson correlation matrix showed that only pretest, group membership, and highest math class taken were

significantly correlated with posttest scores. However, in the stepwise regression model, none of the demographic nor academic characteristics were significant.

The result aligns with prior work, including that of Tintle, Topliff, Vanderstoep, Holmes, and Swanson (2012), of Chance et al. (2016), and by Maurer and Lock (2016). Neither gender nor native speaker were significant at the bivariate level even though the findings of the study of Corral et al. (2014) reported that students who speak English as a second language have significantly less chance of academic success, and the study of Schwery et al. (2016) reported that males tend to perform better than females on standardized tests.

A possible explanation for the gender finding in this study was the ratio of males to females, with females being the majority of the sample in both classes. For every male in each group, there were approximately two females. The unbalanced gender sample size might have meant there were not enough males to compare against the females. Another chance is that females might have been primed for the course more than the males when they took the pretest, so they were ready for the course in a way that they would not have been without the pretest. However, this is part of the nature of the pretest-posttest design, and this is defined as testing threat (Trochim et al., 2016).

A potential explanation for the finding of no correlation between non-native English speakers and posttest scores might be due to the framework of the demographic question. The survey question asked, “What is your native language? (first language spoken)” rather than asking “Is English your stronger language?” Most of the students attending Miami Dade College Kendall Campus are second generation Hispanics, and even though since birth the main language spoken at home is Spanish, once they start

formal educational, they learn to speak and to write English more academically than they do in Spanish. By the time they have finished high school, they are proficient in the English language despite learning Spanish first.

Research Question 3. The third research question posed in this study was *Did students in a randomization-based inference teaching methodology group have higher statistical literacy scores than students in the theory-based inference teaching methodology group when controlling for demographic and academic characteristics?*

The third research question dealt at the multivariate level. In order to answer this question, a stepwise multiple linear regression was run. The stepwise regression ran several models until the final model only included significant predictors that accounted for the variance in posttest scores as measured by the CAOS test. The only two significant predictors for posttest scores were pretest and group membership.

The pretest provided a baseline for each student and allowed the researcher to determine if there was any difference between the two groups before the instruction began. In this study, there was no significant difference in statistical literacy scores at the pretest level. On the other hand, group membership had to do with the teaching modality. Therefore, as seen with Pfannkuch and Budgett (2014), the adaptable use of simulations aided students to advance abundant schemas and conceptual understanding. Such level of comprehension is because students developed the ability to move at will in any given representational system, thereby allowing them to participate actively in procedural and conceptual exchanges with the representations. None of the students' demographic and academic characteristics were significant, even though the highest math class taken was significant at the bivariate level at the .01 significant level.

The results of the present study demonstrate that statistical literacy was better comprehended under the simulation-based teaching modality where the emphasis was in inferential techniques and in understanding the terminology during the second half of the semester. This was achieved by promoting the connection between asymptotic and randomization tests, by introducing confidence intervals as an outcome of test of significance rather than the other way around, and by using real-world contexts problems in the instruction, as endorsed by Watson (2011) and Cobb (2007), since such in-class direction allowed corrections of common misconceptions associated with inferential statistics. The following misconceptions were addressed in class: confusing test statistics and p-values; confusing samples and populations; confusing significant level and Type I error; believing p-values are independent of sample size; and misinterpreting the p-value as the probability the alternative hypothesis is true or that the p-value is the probability that the null hypothesis is false. Furthermore, Budgett and Wild (2014) discussed that real-life problems combined with simulations-based inference stimulate sensory cognition particularly for novice students, a phenomenon reflected in the significance of the results of research question three.

Implications

The results from the present study supported the hypothesis that statistical literacy can be significantly enhanced with the use of simulation-based teaching modality as compared to the theory-based instruction modality of computations and procedures (GAISE-ASA, 2005; Garfield & Ben-Zvi, 2008) even after controlling for demographic and academic differences. As a result, such results suggest that instructors need to create a favorable context for students to learn how to reason with statistical concepts. In

addition, instructors may need to have clear statistical literacy learning goals in their courses. Major statistical literacy learning goals of this study were understanding the purpose of randomization in an experiment, ability to determine which of two boxplots represents a larger standard deviation, understand that correlation does not imply causation, the ability to recognize a correct interpretation of a p-value, and understand how sampling error is used to make an informal inference about a sample mean, among others (delMas et al., 2007). These statistical literacy learning goals are evaluated in the CAOS test, and since its reliability is 0.82, results indicate that it is an effective way to measure students' mastering of statistical literacy rather than relying on instructors' self-made tests. Students need to be taught how to relate the different statistical concepts they learned by making critical conceptual connections. The simulation-based teaching modality facilitated the connections among the concepts of sampling variability, confidence intervals, and hypothesis testing by using the resampling techniques of bootstrap confidence intervals and permutation tests.

Another suggestion for the practice is about evaluating students with a standardized assessment that possesses validity and reliability considerations. As a result, the CAOS instrument could be used at the end of an introductory undergraduate statistics course to provide feedback about essential statistical literacy topics to evaluate students' learning outcomes. Priority should be given to the following outcomes: teaching the p-value as an integral part of a larger statistical process, highlight the concept of the distribution under the null hypothesis, stress the difference of variation within and between, and accentuate the distinctness of sample statistics from population parameters. As a result, further scrutiny can be given to comprehend how these topics have been

taught in the curriculum and why students are erroneously answering some questions, leading to curricula upgrades. For example, the traditional theory-based teaching modality overplays the normal distribution, leading students to select, most of the time, graphs that bring to mind a normal distribution, without cautiously thinking about all data offered in the problem. The CAOS instrument is a tool for diagnosing students' misunderstandings and guiding changes and advancements in statistics education.

There has been work done to change undergraduate introductory statistics courses based on the recommendations by Cobb (2007) and by the Guidelines for Assessment and Instruction in Statistics Education (GAISE; ASA, 2005). New statistics curricula have been created, using a simulation-based approach to teaching inferential statistics by focusing on randomization tests, specifically by using the techniques of bootstrap confidence intervals when the distribution is unknown and permutation test for any test statistic (Garfield et al., 2012; Tintle et al., 2011). As a result, researchers should examine how strong students are doing on this new curricula and gauge if there is a particular curriculum that is superior at increasing student performance.

Finally, particular attention should be given to Liu's (2010) learning model. She established four phases in order to attain conceptual change: externalization, reflection, construction, and application. According to Liu (2010), students should be led by the instructors in order to make them aware of their preconceptions before the instructional program. Hence, the purpose of this stage is to enable students to be aware of their own ideas about the statistical concepts they had learned.

Limitations of the Study

A number of limitations regarding this study are notable.

One limitation of the study was incomplete measures of variables in the demographic and academic survey. Several students were unable to convert their high school GPA to the U.S. scale since several of them come from other countries.

Additionally, even though students received a verbal explanation from the researcher about the nature of the study, students might not have perceived it as valuable to them and did not finish the survey. As a consequence, the sample size was 70 in the simulation-based group and 55 in the traditional theory-based group.

In addition, student withdrawal in the courses could have been an issue. If students who withdrew from the traditional theory-based teaching group were the more academically outstanding compared to those who withdrew from the simulation-based teaching modality (e.g., because they were less motivated), the mean post-CAOS test score of the traditional theory-based group could be lower than would have been expected. While there is no way to test this, it is a possibility. Subsequently, the difference in the scores could not be justified exclusively by the teaching modality but also by the attrition. This becomes a threat to the internal validity of the results, regardless that attrition of participants is an aspect typically related to data collection and analysis (Creswell, 2005).

Another limitation of this research was the lack of random assignment to the groups. As a result, causation cannot be established in the results of this research. It is just a correlational study. The academic readiness of the students is different across groups; this could have been taken care of by searching at the scores on placement exams

but since MDC is an open institution such requirement is not warranted. At the instructor level, one limitation was that for both groups the instructor was not the same. Even though the instructors' feedback evaluation may yield similar results in the areas evaluated, there is always an intrinsic teaching style that cannot be measured through the instructors' feedback evaluation. In addition, there was only one class for each type of teaching modality where data was collected. Future studies should expand the number of classes in order to collect data.

Response variability in the CAOS might also involve assessment exhaustion and lack of exam taker motivation (Weathington et al., 2010). Even though participants in both groups received credit for completing the assessment, receiving credit only for completion might influence students' motivation to excel on the CAOS test. In regard to environmental variability, both groups were tested in the classroom. However, one source of measurement error is that both groups were not tested at the same time since the courses met at different times during the week. The comparison group met Mondays and Wednesday from 2:05 PM to 3:20 PM, and the simulation-based group met Tuesday and Thursday from 5:40 PM to 6:55 PM. The descriptive statistics showed that the traditional class that meets at 2:05 PM had 80% of full time students while the simulation-based class that meets at 5:40 PM had only 65.7% of full time students. This might have had an effect on the overall results.

Prior research has found the same limitations since they are difficult to circumvent. Chance et al. (2016) reported the effect of attrition in their model and advised randomizing the sample. Maurer and Lock (2016) reported that the volatility rate for Type I errors increased under minor violations to model assumptions. They disclosed

that all comparative educational studies have this central challenge where pedagogical modalities are conducted on the class level and measurements are taken at the individual student level. Even with these limitations, this research found sufficient evidence to support the claim that teaching an undergraduate introductory statistics course under the simulation-based teaching modality yields significantly better results in statistical literacy scores.

Recommendations

There are several recommendations for future studies. First, at the demographic and academic survey level, one strong suggestion is to ask the Institutional Research Board of the college where the study is conducted for students' grade point average rather than collecting the information via a survey. A proportion of students at MDC get their high school diploma from foreign countries, which used a different scale and they are unaware of the scale conversion. Therefore, it might happen that a proportion of Hispanic students who decide to attend a HSI have done their high school in other countries.

Another important recommendation about the survey is in regard to the question about being a native speaker. The question asked students if they are native speakers but according to Merriam-Webster (n.d.) the definition of a native speaker is "a person who learned to speak the language of the place where he or she was born as a child rather than learning it as a foreign language." The problem with the question was that several students have a higher proficiency in their foreign language rather than their native language, since they arrived to the USA at a young age. On the other hand, second generation Hispanics living in South Florida who were born in the U.S. learned Spanish first and then learned English once they went to school. This question did provide the

researcher with information about how proficient students were in English. The recommendation is to ask students what language they are the most comfortable as a follow up question.

In regard to the research design, several suggestions should be considered. For a more in-depth study, a hierarchical linear model could be executed, one level is at the student, as done in this study, and the other at the instructor level. Chance et al. (2016) collected data on the instructor experience level by the number of years teaching the subject and on the instructor's experience on the simulation-based teaching modality. Another variable collected was gender of the instructor since there has been an interest in the role of instructor gender on student achievement. The rationale that Chance et al. (2016) used in their study was that the hierarchical model is an attempt to explain section to section variability in student conceptual gain. The current result of this study showed that the most significant predictors of student gains are the students' pre-test scores and student GPA. The same study found that after adjusting for pre-attitudes measures towards statistics, pre-concept scores, and the instructor level variables, the students in the simulation-based curriculum achieve higher gains, particularly with more experienced instructors, with the resampling methods used in simulations. Instructors have different motivations and perceptions of how to disseminate the content in an introductory statistics course; therefore, this should be a variable to consider. Another recommendation is to evaluate students' attitudes and anxiety level rather than just achievement level to better comprehend their outcomes. Chance et al. (2016) measured attitudes towards statistics and found that students who expected the course to be less difficult tended to have higher gains.

Finally, the study should be replicated in Miami Dade College and Florida International University with more classes in order to incorporate students from the morning, afternoon, night, and weekend courses to determine if it yield similar results. Furthermore, HSIs should conduct similar research to determine the effectiveness of simulation-based teaching in order to generalized the findings. Finally, a pivotal recommendation supported by Lane-Getaz (2017) is to emphasize the concept of the distribution under the null hypothesis. The distribution of the null make available a point of evaluation, since usually the results are from a single sample. Furthermore, it is at this point that the p-value must be taught as an integral part of a larger statistical process in order to interpret the results.

Conclusion

This research study provided support for the simulation-based teaching modality over the traditional theory-based mode of instruction in an undergraduate introductory statistics course at a HSI. The results provided significant information for the statistics education community since no previous research has been conducted at HSI. As noted by Krogstad (2015), Hispanics represent a growing force in the higher education arena. Therefore, it is paramount to revise and advance instruction and curricula in order to help Hispanics achieve completion given the fact that statistics act as a gatekeeper course for STEM and non-STEM students. As found with Cobb (2007) and Rossman (2007), simulation-based teaching mimics real-world experience for the students and no educational activity can overshadow this hands-on experience in manipulating statistics concepts.

Despite the limitations of the present research study, it is vital that statistics educators implement simulations like bootstrap confidence intervals and randomization tests in order to improve the understanding of hypothesis testing and p-values, critical concepts of inferential statistics. Finally, statistics educators who are curious in understanding what statistical literacy knowledge their students have can profit from the CAOS assessment because it yields priceless feedback to the instructors who conduct research with introductory statistics courses.

REFERENCES

- Arbelo-Marrero, F., & Milacci, F. (2016). A phenomenological investigation of the academic persistence of undergraduate Hispanic nontraditional students at Hispanic Serving Institutions. *Journal of Hispanic Higher Education, 15*(1), 22-40.
- Aliaga, M., Cuff, C., Garfield, J., Gould, R., Lock, R., Moore, T., Rossman, A., Stephenson, R., Utts, J., Velleman, P. & Witmer, J. (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College report*. Alexandria, VA: American Statistical Association. Retrieved from <http://www.amstat.org/education/gaise/>
- Allen, K., Stone, A., Rhoads, T. R., & Murphy, T. J. (2004). The statistics concepts inventory: Developing a valid and reliable instrument. *Proceedings of the 2004 American Society for Engineering Education Annual Conference and Exposition* (pp. 1, 15). Salt Lake City, UT. Online: <https://engineering.purdue.edu/SCI/pubs/ASEE%202004%20SCI.pdf>
- Allen, K. (2006). *The Statistics Concept Inventory: Development and analysis of a cognitive assessment instrument in statistics*. (Unpublished doctoral dissertation). University of Oklahoma, Norman, OK.
- Allen, K., Rhoads, T. R., & Terry, R. (2006, October). Work in progress: Assessing student confidence of introductory statistics concepts. In *Frontiers in Education Conference, 36th Annual* (pp. 13-14). IEEE.
- Allison, P. D. (1999). *Multiple regression: A primer*. Pine Forge Press
- Aquilonius, B. C., & Brenner, M. E. (2015). Students' Reasoning About p-values. *Statistics Education Research Journal 14*(2), 7-27.
- Arbelo-Marrero, F., & Milacci, F. (2016). A phenomenological investigation of the academic persistence of undergraduate Hispanic nontraditional students at Hispanic Serving Institutions. *Journal of Hispanic Higher Education, 15*(1), 22-40.
- Balemian, K., & Feng, J. (2013). First generation students: College aspirations, preparedness and challenges. *Research College Board*. Retrieved from <http://research.collegeboard.org/sites/default/files/publications/2013/8/presentation-apac-2013-firstgeneration-college-aspirations-preparednesschallenges>.
- Beilock, S. L., Gunderson, E. A., Ramirez, G., & Levine, S. C. (2010). Female teachers' math anxiety affects girls' math achievement. *Proceedings of the National Academy of Sciences, 107*, 1060–1063.

- Ben-Zvi, D. (2004). Reasoning about data analysis. In D. Ben-Zvi and J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 121-146). Dordrecht, The Netherlands: Kluwer.
- Ben-Zvi, D., & Arcavi, A. (2001). Junior high school students' construction of global views of data and data representations. *Educational Studies in Mathematics*, 45, 35-65.
- Bernstein, D.A., Clarke-Stewart, A. Roy, E.J., Srull, T.K. & Wickens, C.D. (1994). *Psychology, third edition*. Houghton Mifflin Company.
- Blanco, O. (2017, July 1). *Latinos key to U.S. economic growth*. Retrieved from CNN Money: <https://money.cnn.com/2017/07/01/news/economy/latinos-u-s-economy/index.html>.
- Blejec, A. (2002). Teaching statistical concepts with simulated data. In B. Phillips (Ed.), *Proceeding of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistics Institute.
- Bridges, B., Cambridge, B. Kuh, G., & Leegwater, L. (2005). Student engagement at minority serving institutions: Emerging lessons from the BEAMS project. *New Directions for Institutional Research*, 2005(125), 25-43.
- Brock. T. (2010). Young adults and higher education: Barriers and breakthroughs to success (National Poverty Center Working Paper Series No. 10-08). *Future of Children*, 20(1), 109-132. Retrieved from http://npc.umich.edu/publications/u/working_paper10-08.pdf.
- Brodish, A. B., & Devine, P. G. (2009). The role of performance-avoidance goals and worry in mediating the relationship between stereotype threat and performance. *Journal of Experimental Social Psychology*, 45(1), 180-185.
- Budgett, S., Pfannkuch, M., Regan, M., & Wild, C. J. (2013). Dynamic visualizations and the randomization test. *Technology Innovations in Statistics Education*, 7(2), 1-21.
- Budgett, S., & Wild, C. J. (2014). Students' visual reasoning and the randomization test. *International Conference on Teaching Statistics*. Flagstaff, AZ: IAES.
- Burns, K. (2010). An issue: Community college student success variables: A review of the literature. *The Community College Enterprise*, 16(2), 33-61.
- Bureau of Labor Statistics. (2016). *Occupational Outlook Handbook*. U.S. Department of Labor.

- Calderón Galdeano, E., & Santiago, D. A. (2015). *Hispanic-Serving Institutions (HSIs) Fact Sheet: 2013–14*. Retrieved from <http://www.edexcelencia.org/hsi-cp2/research/hsis-fact-sheet-2013-14>.
- Cambridge Dictionary. (2017, August 11). *Non-native speaker*. Retrieved from Cambridge Dictionary: <http://dictionary.cambridge.org/us/dictionary/english/non-native-speaker>.
- Cardenas, V., & Kerby, S. (2012, August 8). *The state of Latinos in the United States: Although this growing population has experienced marked success, barriers remain*. Washington, DC: Center for American Progress. Retrieved from <https://www.americanprogress.org/issues/race/reports/2012/08/08/11984/thestate-of-latinos-in-the-united-states/>.
- Carter, J., Brown, M., & Simpson, K. (2017). From the classroom to the workplace: How social science students are learning to do data analysis for real. *Statistics Education Research Journal*, 80-101.
- Carver, R. H., & Stephens, M. (2014). It is time to include data management in introductory statistics. In *Ninth International Conference on Teaching Statistics*. Online: http://iaseweb.org/icots/9/proceedings/pdfs/ICOTS9_C134_CARVER.pdf.
- Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.). *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer.
- Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, 1(1). Retrieved from: <http://escholarship.org/uc/item/8sd2t4rr>.
- Chance, B., & Rossman, A. (2014). Using simulation-based inference for learning introductory statistics. *WIREs Computational Statistics*, 6, 211–221.
- Chance, B., Wong, J., & Tintle, N. (2016). Student performance in curricula centered on simulation-based inference: A preliminary report. *Journal of Statistics Education*, 24(3), 114-126.
- Charmaz, K. (2006). *Constructing grounded theory: A practical guide through qualitative analysis* (1st ed.). Thousand Oaks, CA: Sage.
- Cobb, G. W. (2007). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education*, 1(1).

- Cobb, G. (2015). Mere renovation is too little too late: We need to rethink our undergraduate curriculum from the ground up. *The American Statistician*, 69(4), 266–282.
- College Board. (2013). *The 2013 SAT report on college and career readiness*. New York, NY: Retrieved from <http://research.collegeboard.org/programs/sat/data/archived/cb-seniors-2013>.
- The College Board. (2017, July). *College Board AP*. Retrieved from Score Distribution : <https://apscore.collegeboard.org/scores/about-ap-scores/score-distributions/>.
- Coll-Serrano, V., Blasco-Blasco, O., & Alvarez-Jareno, J. A. (2011). Dynamic graphics in Excel for teaching statistics: understanding the probability density function. *Teaching Mathematics and its Applications*, 30(4), 222-226.
- Corral, D., Gasman, M., Nguyen, T. H., & Samayoa, A. (2015). An examination of existing and emerging Hispanic Serving Institutions' Latino initiatives and culture. Center for Minority Serving Institutions. Retrieved from http://repository.upenn.edu/gse_pubs/344.
- Creswell, J. (2005). *Educational research planning, conducting, and evaluating quantitative and qualitative research*. Upper Saddle River, NJ: Pearson Education, Inc.
- Cumming, G. (2010). Understanding, teaching and using p values. *International Conference on Teaching Statistics*. Voorburg: IASE.
- Cummins, D. D. (2014). *Why the gender difference on SAT math doesn't matter*. Retrieved from [https:// www.psychologytoday.com/blog/good-thinking/201403/why-the-gender-difference-sat-math-doesntmatter](https://www.psychologytoday.com/blog/good-thinking/201403/why-the-gender-difference-sat-math-doesntmatter)
- delMas, R. (2014). Trends in students' conceptual understanding of statistics. *International Conference on Teaching Statistics*, Flagstaff: IASE.
- delMas, R., Garfield, J., & Ooms, A. (2005, July). Using assessment items to study students' difficulty reading and interpreting graphical representations of distributions. In *Fourth Forum on Statistical Reasoning, Thinking, and Literacy (SRTL-4)*.
- delMas, R., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal*, 6, 28–58.
- De Veaux, R. (2014). Introductory statistics in the 21st Century. *International Conference on Teaching Statistics*. Flagstaff: IASE.

- DeVellis, R.F. (2012). *Scale development: Theory and applications*. Los Angeles: Sage.
- Diez, D. M., Barr, C. D., & Çetinkaya-Rundel, M. (2014). *Introductory statistics with randomization and simulation*. OpenIntro.
- Dika, S. L., & D'Amico, M. M. (2016). Early experiences and integration in the persistence of first-generation college students in stem and non-stem majors. *Journal of Research in Science Teaching*, 53(3), 368–383.
- Dominicis, E. (2016). *Miami Dade College - American Library Association*. Miami
- Efron, B. (1979). Bootstrap methods: Another look at the jackknife. *The Annals of Statistics*, 7, 1-26.
- Festinger, L. (1957). *A theory of cognitive dissonance*. Evanston, IL: Row, Peterson, 1.
- Faul, F., Erdfelder, E., Buchner, A., & Lang, A. G. (2009). Statistical power analyses using G* Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, 41(4), 1149-1160.
- Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. SAGE .
- Fisher, R. A. (1936). "The coefficient of racial likeness" and the future of craniometry. *The Journal of the Royal Anthropological Institute of Great Britain and Ireland*, 66, 57-63.
- Forbes, S., Chapman, J., Harraway, J., Stirling, D., & Wild, C. (2014). Use of data visualization in the teaching of statistics: A New Zealand perspective," *Statistics Education Research Journal*, 13, 187–201. Available at http://iaseweb.org/documents/SERJ/SERJ13%282%29_Forbes.pdf.
- Gal, I. (2002). Adults' statistical literacy: Meanings, components, responsibilities. *International Statistical Review*, 70(1), 1–25.
- Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25-34.
- Garfield, J. (2002). The challenge of developing statistical reasoning. *Journal of Statistics Education*. 10(3). Retrieved from: <http://ww2.amstat.org/publications/jse/v10n3/garfield.html>.
- Garfield, J. (2003). Assessing statistical reasoning. *Statistics Education Research Journal*, 2(1), 22-38.

- Garfield, J., & Ben-Zvi, D. (2005). A framework for teaching and assessing reasoning about variability. *Statistics Education Research Journal*, 4(1), 92-99.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396.
- Garfield, J., & Ben-Zvi, D. (2008). *Developing students' statistical reasoning: Connecting research and teaching practice*. New York, NY: Springer.
- Garfield, J., delMas, R., & Chance, B. (2002). *The Assessment Resource Tools for Improving Statistical Thinking (ARTIST) Project*. NSF CCLI grant ASA-0206571. Retrieved from <https://apps3.cehd.umn.edu/artist/index.html>.
- Garfield, J., delMas R., & Zieffler, A. (2012). Developing statistical modelers and thinking in an introductory, tertiary-level statistics course. *The International Journal on Mathematics Education*, 44(7), 883-898.
- Garfield, J., delMas, R., & Zieffler, A. (2010). Assessing important learning outcomes in introductory tertiary statistics courses. In Bidgood, P., Hunt, N., & Jolliffe, F. (Eds.), *Assessment methods in statistical education* (pp. 75-86). John Wiley & Sons, Ltd.
- Garcia, E. (2000). *Student cultural diversity: Understanding and meeting the challenge*. New York: Houghton Mifflin Company.
- GAISE. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report*. American Statistical Education. Retrieved from http://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf
- Gilardi, S., & Guglielmetti, C. (2011). University life of non-traditional students: Engagement styles and impact on attrition. *The Journal of Higher Education*, 82(1), 33-53.
- Good, C., Rattan, A., & Dweck, C. S. (2012). Why do women opt out? Sense of belonging and women's representation in mathematics. *Journal of Personality and Social Psychology*, 102(4), 700-717.
- Gould, R., Davis, G., Patel, R., & Esfandiari, M. (2010). Enhancing conceptual understanding with data driven labs. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_C208_GOULD.pdf

- Grant, H., & Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of Personality and Social Psychology*, 85(3), 541-553.
- Haines, B. (2015). Conceptualizing a framework for advanced placement statistics teaching knowledge. *Journal of Statistics Education*, 4, 27-437.
- Hassad, R. A. (2011). Constructivist and behaviorist approaches: Development and initial evaluation of a teaching practice scale for introductory statistics at the college level. *Numeracy*, 4(2), Article 7.
- Hesterberg, T., Monaghan, S., Moore, D. S., Clipson, A., & Epstein, R. (2003). Bootstrap methods and permutation tests [Companion Chapter]. In Moore, D. S., McCabe, G. P., Duckworth, W. M., & Sclove, S. L. (eds.) *The Practice of Business Statistics*. New York: W. H. Freeman and Company.
- Herk, M. (2016, June 11). *The Skills Gap*. Retrieved from Committee for Economic Development: <https://www.ced.org/blog/entry/the-skills-gap-and-the-seven-skill-sets-that-employers-want-building-the-id>.
- Hispanic Associations of Colleges & Universities . (2017). *2017 Fact Sheet*. Retrieved from Hispanic Higher Education and HSIs: https://www.hacu.net/hacu/HSI_Fact_Sheet.asp.
- Hollis-Sawyer, L. (2011). A math-related decrement stereotype threat reaction among older nontraditional college learners. *Educational Gerontology*, 37(4), 292–306.
- Horton, N. J. (2015). Challenges and opportunities for statistics and statistical education: Looking back, looking forward. *The American Statistician*, 69(2), 138-145.
- Hubbard, S. & Stage, F. (2009). Attitudes, perception, and preferences of faculty at Hispanic serving and predominantly Black institutions. *The Journal of Higher Education*, 80(3), 270-289.
- Jehangir, R. R. (2010). *Higher education and first-generation students: Cultivating community, voice, and place for the new majority*. New York, NY: Palgrave Macmillan.
- Kahneman, D. (2013). *Think, fast and slow*. Farrar, Straus and Giroux, New York.
- Kahneman, D., Slovic, P., & Tversky, A. (1982). *Judgment under uncertainty: Heuristic and biases*. Cambridge, UK: Cambridge University Press.

- Khazanov, L., & Prado, L. (2010). Correcting students' misconceptions about probability in a introductory college statistics course. *Adults Learning Mathematics International Journal*, 5(1), 23-35.
- Knold, C. (1995). Issues in assessing conceptual understanding in probability and statistics. *Journal of Statistics Education* 3(1), 1-9.
- Konold, C., Pollatsek, A., Well, A., & Gagnon, A. (1997). Students analyzing data: Research of critical barriers. In Garfield, J. B. & Burrill, G. (Eds.), *Research on the role of technology in teaching and learning statistics: Proceedings of the 1996 IASE Round Table Conference*, 151-167. Voorburg, The Netherlands: *International Statistical Institute*.
- Krathwohl, D. R. (2002). A revision of Bloom's taxonomy: An overview. *Theory into Practice*, 41(4), 212-218.
- Krogstad, J. M. (2014). *With fewer new arrivals, Census lowers Hispanic population projections*. Retrieved from Pew Research Center: <http://www.pewresearch.org/fact-tank/2014/12/16/with-fewer-new-arrivals-census-lowers-hispanic-population-projections-2/>.
- Krogstad, J. M. (2016, July 28). *5 facts about Latino and education*. Retrieved from Pew Research Center: <http://www.pewresearch.org/fact-tank/2016/07/28/5-facts-about-latinos-and-education/>.
- Laden, B. (2001). Hispanic-Serving Institutions: Myths and realities. *Peabody Journal of Education*, 76(1), 73-92.
- Lane-Getaz S (2013). Development of a reliable measure of students' inferential reasoning ability. *Statistics Education Research Journal*, 12(1), 20-35.
- Lane-Getaz, S. J. (2017). Is the p-value really dead? Assessing inference learning outcomes for social science students in a introductory statistics course. *Statistics Education Research Journal*, 16(1), 357-399.
- Lecoutre, M. P. (1992). Cognitive models and problem spaces in “purely random” situations. *Educational Studies in Mathematics*, 23(6), 557-568.
- Lecoutre, M. P., & Rezrazi, M. (1998). Learning and transfer in isomorphic uncertainty situations: the role of the subject’s cognitive activity. In *Statistical Education—Expanding the network, Proceedings of the fifth International Conference on Teaching of Statistics* (Vol. 3, pp. 1105-1111). Vooburg, Netherlands: ISI Permanent Office.

- Lehrer, R., & Schauble, L. (2002). Distribution: A resource for understanding error and natural variation. In B. Phillips (Ed.), *Proceeding of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg. The Netherlands: International Statistics Institute.
- Limon, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: A critical appraisal. *Learning and Instruction*, 11(4-5), 357-380.
- Little, T. D. (2013). *Longitudinal structural equation modeling*. Guilford Press.
- Liu, T.C. (2010). Developing simulation-based computer assisted learning to correct students' statistical misconceptions based on cognitive conflict theory, using "correlation" as an example. *Educational Technology & Society*, 180-192.
- Liu, C. H., & Matthews, R. (2005). Vygotsky's philosophy: Constructivism and its criticisms examined. *International Education Journal*, 6(3), 386-399.
- Liu, Y., & Thompson, P. W. (2009). Mathematics teachers' understandings of proto-hypothesis testing. *Pedagogies*, 4(2), 126-138.
- Lock, R., Lock, P., Morgan, K., Lock, E., & Lock. D. (2013). *Statistics: Unlocking the power of data*. Hoboken, NJ: Wiley.
- Lock, R., Lock, P., Morgan, K., Lock, E., & Lock. D. (2017). *Statistics: Unlocking the power of data*. Wiley, Hoboken, NJ.
- Longwell-Grice, R., Adsitt, N. Z., Mullins, K., & Serrata, W. (2016). The first ones: Three studies on first-generation college students. *NACADA Journal*, 36(2), 34-46.
- Maurer, K., & Lock, E. (2015). Bootstrapping in the introductory statistics curriculum. *Technology Innovations in Statistics Education*.
- Maurer, K., & Lock, D. (2016). Comparison of learning outcomes for simulation-based and traditional inference curricula in a designed educational experiment. *Technology Innovations in Statistics Education*, 9(1).
- McDonald, S. D., & Vrana, S. R. (2007). Interracial social comfort and its relationship to adjustment to college. *The Journal of Negro Education*, 130-140.
- Miami Dade College. (2016). *Catolog 2016-2018*. Miami: Miami Dade College.
- Miami Dade College. (2014). *MDC*. Retrieved from MDC: <http://www.mdc.edu/about/facts.aspx>.

- Mills, J. D. (2002). Using computer simulation methods to teach statistics: A review of the literature. *Journal of Statistics Education*, 10(1), 1-20.
- Merriam-Webster (n.d.) Native Speaker." Retrieved from <https://www.merriam-webster.com/dictionary/native%20speaker>.
- National Center for Education Statistics (2011a). *National assessment of educational progress* (NAEP). Washington, D.C.: U.S. Department of Education, Retrieved from <http://nces.ed.gov/nationsreportcard/mathematics/>.
- National Center for Education Statistics. (2011b). *Digest of Education Statistics, 2010* (NCES 2011-015). Retrieved from <http://nces.ed.gov/fastfacts/display.asp?id=98>.
- National Center for Education Statistics. (2014). *The condition of education 2014* (NCES2014-083, Educational Attainment). Retrieved from <http://nces.ed.gov/fastfacts/display.asp?id=27>.
- National Center for Education Statistics. (2016, August). *Status and trends in the education of racial and ethnic groups 2016*. Retrieved from <https://nces.ed.gov/pubs2016/2016007.pdf>.
- National Survey of Student Engagement. (2012). *Promoting student learning and institutional improvement: Lessons from NSSE at 13*. Bloomington, IN: Indiana University Center for Postsecondary Research.
- Nisbett, R. (1993). The use of statistical heuristics in everyday inductive reasoning. In R. Nisbett (Ed.), *Rules for reasoning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Nosek, B. A., & Smyth, F. L. (2011). Implicit social cognitions predict sex differences in math engagement and achievement. *American Educational Research Journal*, 48, 1125–1156.
- Novak, E. (2014). Effects of simulation-based learning on students' statistical factual, conceptual and application knowledge. *Journal of Computer Assisted Learning*, 30, 148–158.
- Núñez, A. M., Crisp, G., & Elizondo, D. (2016). Mapping Hispanic Serving Institutions: A typology of institutional diversity. *The Journal of Higher Education*, 87(1), 55-83.
- Pew Research Center. (2014). *Pew Research Center Hispanics Trends*. Retrieved from Latinos by Geography: <http://www.pewhispanic.org/states/state/fl/>.
- Pfannkuch, M., Regan, M., Wild, C., Budgett, S., Forbes, S., Harraway, J., Parsonage, R.

- (2011). Inference and the introductory statistics course. *International Journal of Mathematical Education in Science and Technology*, 42(7), 903-913.
- Pfannkuch, M., & Budgett, S. (2014). Constructing inferential concepts through bootstrap and randomizations-test simulations: A case study. *International Conference on Teaching Statistics*. Flagstaff, AZ: IAES.
- Principi, A., & Lamura, G. (2009). Education for older people in Italy. *Educational Gerontology*, 35(3), 246–259.
- Reaburn, R. (2014). Introductory statistics course tertiary students' understanding of p-values. *Statistics Education Research Journal*, 13(1), 53-65.
- Reed-Rhoads, T., Murphy, T. J., & Terry, R. (2006). *The statistics concept inventory*. Retrieved from Purdue University, The Statistics Concepts Inventory Project Web site: <https://engineering.purdue.edu/SCI>.
- Nicholson, J., & Ridgway, J. (2017). A response to White and Gorard: Against inferential statistics: How and why current statistics teaching gets it wrong. *Statistics Education Research Journal*, 16(1), 66-73.
- Roy, S., Rossman, A., Chance, B., Cobb, G., VanderStoep, J., Tintle, T., & Swanson, T. (2014). Using simulation/randomization to introduce p-value in week 1,” in *Proceedings of the 9th International Conference on Teaching Statistics*, Flagstaff, AZ. Available at http://icots.info/9/proceedings/pdfs/ICOTS9_4A2_ROY.pdf
- Rumsey, D. J (2002). Statistical literacy as a goal for introductory statistics courses. *Journal of Statistics Education*. 10(3), 6-13.
- Sacristan, A., Calder, N., Rojano, T., Santos-Trigo, M., Friedlander, A., & Meissner, H. (2010). The influence and shaping of digital technologies on the learning and learning trajectories of mathematical concepts. In C. Hoyles, & J. Lagrange (Eds.), *Mathematics education and technology - Rethinking the terrain: The 17th ICMI Study* (pp. 179-226). New York, NY: Springer.
- Saldanha, L., & Thompsom, P. (2003). Conceptions of sample and their relationship to statistical inference. *Educational Studies in Mathematics*, 5, 257-270.
- Santos, A. & Cuamea, K. (2010). Challenges facing Hispanic Serving Institutions in the first decade of the 21st century. *Journal of Latinos and Education*, 9(2), 90-107.
- Schmid, C. (2001). Educational achievement, language minority and the new second generation. *Sociology of Education* 73(Extra Issue), 71-87.

- Schwery, D., Hulac, D., & Schweinle, A. (2016). Understanding the gender gap in mathematics achievement: The role of self-efficacy and stereotype threat. *National Association of School Psychologists, 10*(4), 386-396.
- Scott-Jones, J. S., & Goldring, J. E. (2017). Telling stories, landing planes and getting them moving—a holistic approach to developing students' statistical literacy. *Statistics Education Research Journal, 17*(1).
- Seidman, A. (2005). Minority student retention: resources for practitioners. *New Directions for Institutional Research, 125*, 7-24.
- Smith III, J. P., DiSessa, A. A., & Roschelle, J. (1994). Misconceptions reconceived: A constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*(2), 115-163.
- Smith, T. M. F., & Staetsky, L. (2007). The teaching of statistics in UK universities. *Journal of the Royal Statistical Society, Series A, 170*, 581-622.
- Snee, R. (1993). What's missing in statistical education? *The American Statistician, 47*, 149-154.
- Solomon, Y. (2012). Finding a voice? Narrating the female self in mathematics. *Educational Studies in Mathematics, 80*, 171–183.
- Stebbleton, M. J., & Soria, K. M. (2012). Breaking down barriers: Academic obstacles of first-generation students at research universities. *The Learning Assistance Review, 17*(2), 7-19.
- Stebbleton, M. J., Soria, K. M., & Huesman, R. L. (2014). First-generation students' sense of belonging, mental health, and use of counseling services at public research universities. *Journal of College Counseling, 17*(1), 6-20.
- Steele, C. M., & Aronson, J. A. (2004). Stereotype threat does not live by Steele and Aronson (1995) alone. *American Psychologist, 59*(1), 47–48.
- Suárez-Orozco, M. (1998). *Mexican immigration in interdisciplinary perspectives*. Boston, Harvard University, David Rockefeller Center for Latin American Studies.
- Suárez-Orozco, M. & Pérez, M. (2008). *Latinos: Remaking America*. University of California Press.
- Swail, W., Redd, K., & Perna, L. (2003). *Retaining minority students in higher education: A framework for success*. ASHE-ERIC Higher Education Report No. 2.

Washington, DC: The George Washington University, School of Education and Human Development.

- Tabachnick, B. G. & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Boston, MA: Allyn and Bacon.
- Tall, D., & Vinner, S. (1981). Concept image and concept definition in mathematics with particular reference to limits and continuity. *Educational Studies in Mathematics*, 12 (2), 151-169.
- Tappin, L.A. (2000). Statistics in a nutshell. *Journal of Statistics Education*, 8(1). Retrieved from: <http://ww2.amstat.org/publications/jse/secure/v8n1/tappin.cfm>.
- Tintle, N. L., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2015). *Introduction to statistical investigations*. New York: Wiley.
- Tintle, N., Swanson, T., VanderStoep, J., Roy, S., Rossman, A., Cobb, G., Rogers, A., & Chance, B. (2014). *Quantitative evidence for the user of simulation and randomization in the introductory statistics course*. In Proceedings of the 9th International Conference on Teaching Statistics, Flagstaff, AZ.
- Tintle, N., Topliff, K., VanderStoep, J., Homes, V.-L., & Swanson, T. (2012). Retention of statistical concepts in a preliminary randomization- based introductory statistics curriculum. *Statistics Education Research Journal*, 11, 21–40.
- Tintle, N., VanderStoep, J., Holmes, V-L., Quisenberry, B., & Swanson, T. (2011). Development and assessment of a preliminary randomization- based introductory statistics curriculum. *Journal of Statistics Education*, 19(1). n1.
- Trafimow, D., & Marks, M. (2015). Publishing models and article dates explained . *Basic and Applied Social Psychology*, 37, 1-2.
- Trochim, W., Donnelly, J., & Arora, K. (2016). *Research Methods The Essential Knowledge Base*. Cengage.
- Tuan, N. T., & Jay, I. (2016). The Mind Could Accommodate Inconsistencies. *Systemic Practice and Action Research*, 29(2), 173-182.
- Turegun, M. (2011). *A model for developing and accessing community college students' conceptions of the range, interquartile range, and standard deviation*. (Unpublished doctoral dissertation). University of Oklahoma, Norman, OK. Retrieved from <http://www.stat.auckland.ac.nz/~iase/publications/dissertations/dissertations.php>.

- Tversky, A., & Kahneman, D. (1974). Heuristics and biases: Judgement under uncertainty. *Science*, 185, 1124-1130.
- U.S. Census Bureau. (2015). *American Community Survey*. Retrieved September 5, 2017, from <http://www.census.gov/acs>.
- United States Census Bureau (2016). *Population Distribution and Change: 2000 to 2010*. Retrieved from <https://www.census.gov/prod/cen2010/briefs/c2010br-01.pdf>.
- United States Department of Education . (2013). *Developing Hispanic-serving institutions program - Title V*: Retrieved from: <https://www2.ed.gov/programs/idueshsi/index.html>.
- Utts, J. (2003). What educated citizens should know about statistics and probability. *The American Statistician*, 57(2), 74-79.
- Velleman, P. F., & Moore, D. S. (1996). Multimedia for teaching statistics: Promises and pitfalls. *The American Statistician*, 50, 217-225.
- Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society, *Journal of the American Statistical Association*, 88(421), 1-8.
- Watson, J.M. (1997). Assessing statistical thinking using the media. In I. Gal & J.B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 107–121). Amsterdam: IOS Press.
- Watson, J. M. (2002). Inferential reasoning and the influence of cognitive conflict. *Educational Studies in Mathematics*, 51(3), 225-256.
- Watson, J.M. (2011). Foundations for improving statistical literacy. *Statistical Journal of the IAOS*, 27, 197-204.
- Weathington, B. L., Cunningham, C. J., & Pittenger, D. J. (2010). *Research methods for the behavioral and social sciences*. Hoboken, NJ: Wiley.
- Wosley, S.A., & Shepler, D.K. (2011). Understanding the early integration experiences of first-generation college students. *College Student Journal*, 45(4), 700- 714.
- Zieffler, A., and Catalysts for Change (2015). *Statistical thinking: A simulation approach to uncertainty* (3rd ed.), Minneapolis, MN: Catalyst Press.
- Zieffler, A., & Garfield, J. (2008). What does research suggest about the teaching and learning of introductory statistics at the college level? A review of the literature. *Journal of Statistics Education*, 16(2). Retrieved from: <https://www2.amstat.org/publications/jse/v16n2/zieffler.html>.

Ziegler L (2014). *Reconceptualizing statistical literacy: Developing an assessment for the modern statistics course*. Ph.D. thesis, University of Minnesota.


APPENDICES

APPENDIX A: IRB FROM FIU



Office of Research Integrity
Research Compliance, MARC 414

MEMORANDUM

To: Dr. Maria L. Fernandez
CC: File
From: Maria Melendez-Vargas, MIBA, IRB Coordinator 
Date: November 3, 2017
Protocol Title: "AN INSTRUCTIONAL STRATEGY WITH SIMULATIONS USED TO INCREASE STATISTICAL LITERACY AMONG HISPANIC STUDENTS"

The Social and Behavioral Institutional Review Board of Florida International University has approved your study for the use of human subjects via the Expedited Review process. Your study was found to be in compliance with this institution's Federal Wide Assurance (00000060).

IRB Protocol Approval #: IRB-17-0355 IRB Approval Date: 11/02/17
TOPAZ Reference #: 106144 IRB Expiration Date: 11/02/18

As a requirement of IRB Approval you are required to:

- 1) Submit an IRB Amendment Form for all proposed additions or changes in the procedures involving human subjects. All additions and changes must be reviewed and approved by the IRB prior to implementation.
- 2) Promptly submit an IRB Event Report Form for every serious or unusual or unanticipated adverse event, problems with the rights or welfare of the human subjects, and/or deviations from the approved protocol.
- 3) Utilize copies of the date stamped consent document(s) for obtaining consent from subjects (unless waived by the IRB). Signed consent documents must be retained for at least three years after the completion of the study.
- 4) Receive annual review and re-approval of your study prior to your IRB expiration date. Submit the IRB Renewal Form at least 30 days in advance of the study's expiration date.
- 5) Submit an IRB Project Completion Report Form when the study is finished or discontinued.

Special Conditions: N/A.

For further information, you may visit the IRB website at <http://research.fiu.edu/irb>.

MMV/em

APPENDIX B: CITI PROGRAM COMPLETION

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM) COMPLETION REPORT - PART 1 OF 2 COURSEWORK REQUIREMENTS*

* NOTE: Scores on this Requirements Report reflect quiz completions at the time all requirements for the course were met. See list below for details. See separate Transcript Report for more recent quiz scores, including those on optional (supplemental) course elements.

- **Name:** Eric Hernandez (ID: 2634572)
- **Email:** ehern012@fiu.edu
- **Institution Affiliation:** Florida International University (ID: 970)
- **Institution Unit:** Teaching and Learning
- **Phone:** 305-237-0336

- **Curriculum Group:** Responsible Conduct of Research
- **Course Learner Group:** Social and Behavioral Responsible Conduct of Research Course
- **Stage:** Stage 1 - RCR
- **Description:** This course is for investigators, staff and students with an interest or focus in Social and Behavioral research. This course contains text, embedded case studies AND quizzes.

- **Report ID:** 7287864
- **Completion Date:** 30-Dec-2015
- **Expiration Date:** N/A
- **Minimum Passing:** 80
- **Reported Score*:** 90

REQUIRED AND ELECTIVE MODULES ONLY	DATE COMPLETED	SCORE
Responsible Conduct of Research (RCR) Course Introduction (ID: 1522)	30-Dec-2015	No Quiz
Research Misconduct (RCR-Basic) (ID: 16604)	30-Dec-2015	5/5 (100%)
Data Management (RCR-Basic) (ID: 16600)	30-Dec-2015	4/5 (80%)
Authorship (RCR-Basic) (ID: 16597)	30-Dec-2015	5/5 (100%)
Peer Review (RCR-Basic) (ID: 16603)	30-Dec-2015	4/5 (80%)
Using Animal Subjects In Research (RCR-Basic) (ID: 13301)	18-Jan-2014	5/5 (100%)
Conflicts of Interest (RCR-Basic) (ID: 16599)	30-Dec-2015	5/5 (100%)
Collaborative Research (RCR-Basic) (ID: 16598)	30-Dec-2015	4/5 (80%)
Research Involving Human Subjects (RCR-Basic) (ID: 13566)	30-Dec-2015	4/5 (80%)
Responsible Conduct of Research (RCR) Course Conclusion (ID: 1043)	30-Dec-2015	No Quiz

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing Institution identified above or have been a paid Independent Learner.

Verify at: www.citiprogram.org/verify/?kbe6f856c-fb10-4425-8051-856eae46a0cb-7287864

CITI Program
Email: support@citiprogram.org
Phone: 888-529-5929
Web: <https://www.citiprogram.org>

APPENDIX C: CONSENT FORM

FIU IRB Approval:	11/02/2017
FIU IRB Expiration:	11/02/2018
FIU IRB Number:	IRB-17-0355



ADULT ONLINE CONSENT TO PARTICIPATE IN A RESEARCH STUDY

AN INSTRUCTIONAL STRATEGY WITH SIMULATIONS USED TO INCREASE STATISTICAL LITERACY AMONG HISPANIC STUDENTS

The Informed Consent Form (ICF) is a form of voluntary agreement. With this ICF human participants can read, understand, and agree to voluntarily participate in the research study. Please note that you must be at least 18 years old to participate in this voluntary study.

Hello, my name is Eric Hernandez. The purpose of this study is to examine the effects of statistical simulations on students' mastery of content in a college introductory statistics course, specifically STA 2023. If you decide to be in this study, you will be one of 220 people in this research.

Participation in this study will take no more than 130 minutes of your time as follows: no more than 5 minutes to answer the demographic survey questions and no more than 60 min to answer 40 online multiple choice questions in the Comprehensive Assessments of Outcome in Statistics (CAOS) test. If you agree to be in the study, you will have to do this at the end of the first week of class and repeat the same process the last week of the semester. You will receive homework credit for completing these questionnaires. If you choose not to participate, an alternate MySTATlab assignment requiring the same amount of time and involving the same content will be provided for you to earn the same homework credit.

All information from this survey will be kept strictly confidential. Only the researcher will have access to the data which will be kept in a password protect file. Your identity will remain confidential since you will provide an identification marker to link your demographic survey with your assessment score (CAOS test).

There are no foreseeable risks to you for participating in this study or any unique discomfort from answering these questions. It is expected that this study will benefit society by adding to the literature the missing knowledge about statistical literacy among Hispanic students and in light of the changes in teaching statistics, there has yet to be any work to see if the new methods are increasing Hispanic undergraduate statistical literacy. There is a lack of research that explores whether the new teaching model helps Hispanic students overcome the barriers to statistical reasoning and individual barriers like language.

Additionally, this study will serve as a scaffold for institutions of higher education when developing instructional strategies that enhance students' conceptual understanding in a first undergraduate statistics course.

FIU IRB Approval:	11/02/2017
FIU IRB Expiration:	11/02/2018
FIU IRB Number:	IRB-17-0355

There is no cost or payment to you. If you have questions before and/or while taking the instrument, please stop and contact Eric Hernandez.

I would like to emphasize that your identity will remain safe and protected. In any sort of report we might publish findings from this study, we will not include any information that will make it possible to identify you as a subject. Research records will be stored securely and only the researcher will have access to the records.

All participants will need to complete this ICF in order get access to the online surveys generated through the SurveyMonkey website. Once the online surveys are completed, the researcher will utilize the data for statistical summaries only for the purposes of this study

If you have questions about this research study, you may contact Maria L. Fernandez at 305-348-0193 or Eric Hernandez at 305-237-2453.

If you would like to talk with someone about your rights of being a subject in this research study or about ethical issues with this research study, you may contact the FIU Office of Research Integrity by phone at 305-348-2494 or by email at ori@fiu.edu.

Your participation in this research is voluntary, and you will not be penalized or lose benefits if you refuse to participate or decide to stop.

I am at least 18 years old and give my consent to voluntarily participate in this study. I have read and understood the above information.

***1. I AGREE TO VOLUNTARILY PARTICIPATE AND I AM AT LEAST 18 YEARS OF AGE.**

YES NO

***2. If you agree, please validate your Informed Consent Form**

Name:

Name Initials:

Email Address:

***3. INFORMED CONSENT FORM
COMPLETED
MM/DD/YYYY**

APPENDIX D: SURVEY

**Comprehensive Assessment of Outcomes
for a first course in Statistics (CAOS)**

CAOS 4

**Developed by the Web ARTIST Project
<https://app.gen.umn.edu/artist/>**

**Funded by a grant from the National Science Foundation
NSF CCLI ASA- 0206571**

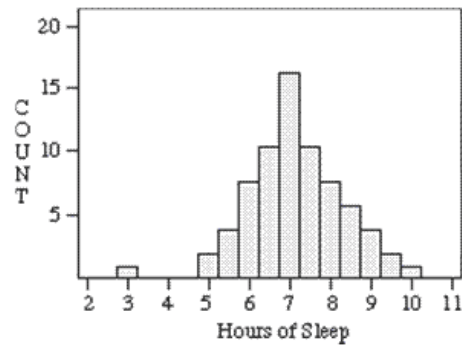
**Principal Investigators:
Joan Garfield and Bob delMas, University of Minnesota
Beth Chance, Cal Poly – San Luis Obispo
Post-doctoral Research Assistant:
Ann Ooms, University of Minnesota**

Version 31

September 8, 2005

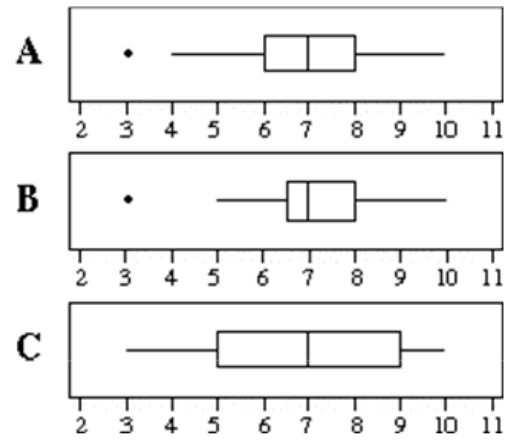
ARTIST CAOS 4 POSTTEST

The following graph shows a distribution of hours slept last night by a group of college students.



1. Select the statement below that gives the most complete description of the graph in a way that demonstrates an understanding of how to statistically describe and interpret the distribution of a variable.
 - a. The bars go from 3 to 10, increasing in height to 7, then decreasing to 10. The tallest bar is at 7. There is a gap between three and five.
 - b. The distribution is normal, with a mean of about 7 and a standard deviation of about 1.
 - c. Most students seem to be getting enough sleep at night, but some students slept more and some slept less. However, one student must have stayed up very late and got very few hours of sleep.
 - d. The distribution of hours of sleep is somewhat symmetric and bell-shaped, with an outlier at 3. The typical amount of sleep is about 7 hours and overall range is 7 hours.

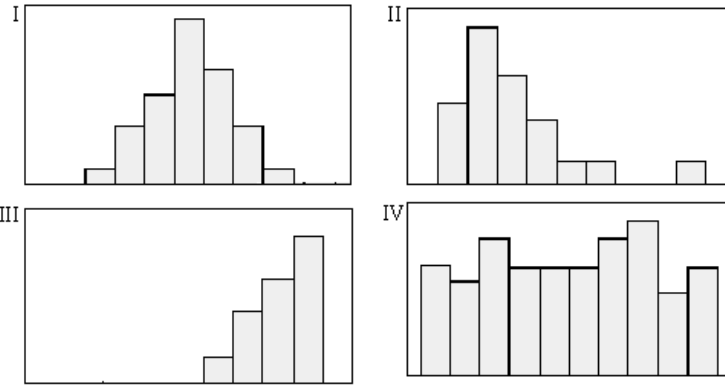
2. Which box plot seems to be graphing the same data as the histogram in question 1?



- a. Boxplot A.
- b. Boxplot B.
- c. Boxplot C.

Items 3 to 5 refer to the following situation:

Four histograms are displayed below. For each item, match the description to the appropriate histogram.



3. A distribution for a set of quiz scores where the quiz was very easy is represented by:
 - a. Histogram I.
 - b. Histogram II.
 - c. Histogram III.
 - d. Histogram IV.

4. A distribution for a set of wrist circumferences (measured in centimeters) taken from the right wrist of a random sample of newborn female infants is represented by:
 - a. Histogram I.
 - b. Histogram II.
 - c. Histogram III.
 - d. Histogram IV.

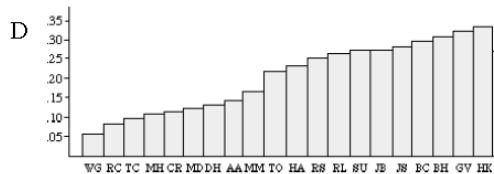
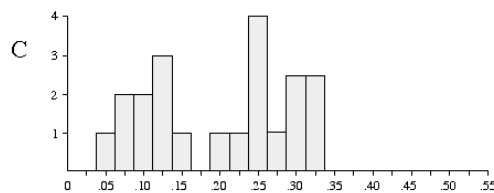
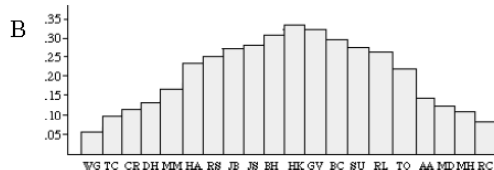
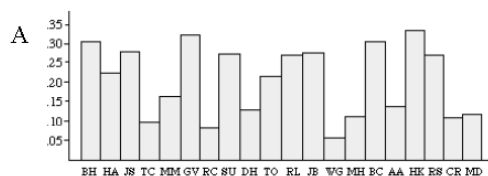
5. A distribution for the last digit of phone numbers sampled from a phone book (i.e., for the phone number 968-9667, the last digit, 7, would be selected) is represented by:
 - a. Histogram I.
 - b. Histogram II.
 - c. Histogram III.
 - d. Histogram IV.

6. A baseball fan likes to keep track of statistics for the local high school baseball team. One of the statistics she recorded is the proportion of hits obtained by each player based on the number of times at bat as shown in the table below. Which of the following graphs gives the best display of the distribution of proportion of hits in that it allows the baseball fan to describe the shape, center and spread of the variable, proportion of hits?

Player	Proportion of hits
BH	0.305
HA	0.229
JS	0.281
TC	0.097
MM	0.167
GV	0.333
RC	0.085

Player	Proportion of hits
SU	0.270
DH	0.136
TO	0.218
RL	0.267
JB	0.270
WG	0.054
MH	0.108

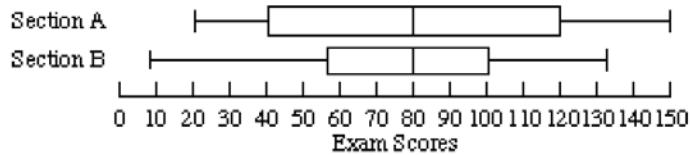
Player	Proportion of hits
BC	0.301
AA	0.143
HK	0.341
RS	0.261
CR	0.115
MD	0.125



7. A recent research study randomly divided participants into groups who were given different levels of Vitamin E to take daily. One group received only a placebo pill. The research study followed the participants for eight years to see how many developed a particular type of cancer during that time period. Which of the following responses gives the best explanation as to the purpose of randomization in this study?
- To increase the accuracy of the research results.
 - To ensure that all potential cancer patients had an equal chance of being selected for the study.
 - To reduce the amount of sampling error.
 - To produce treatment groups with similar characteristics.
 - To prevent skewness in the results.

Items 8 to 10 refer to the following situation:

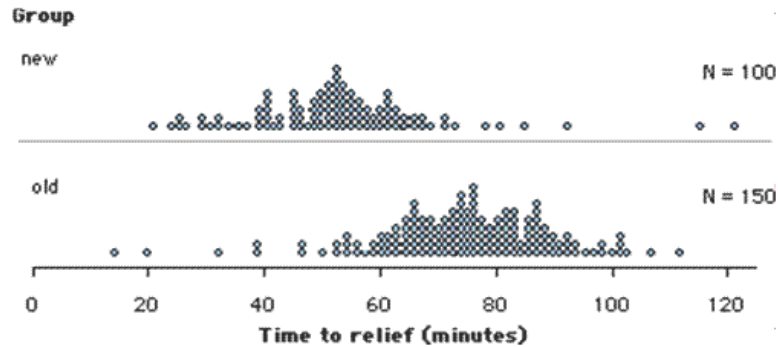
The two boxplots below display final exam scores for all students in two different sections of the same course.



8. Which section would you expect to have a greater standard deviation in exam scores?
- Section A.
 - Section B.
 - Both sections are about equal.
 - It is impossible to tell.
9. Which data set has a greater percentage of students with scores at or below 30?
- Section A.
 - Section B.
 - Both sections are about equal.
 - It is impossible to tell.
10. Which section has a greater percentage of students with scores at or above 80?
- Section A.
 - Section B.
 - Both sections are about equal.

Items 11 to 13 refer to the following situation:

A drug company developed a new formula for their headache medication. To test the effectiveness of this new formula, 250 people were randomly selected from a larger population of patients with headaches. 100 of these people were randomly assigned to receive the new formula medication when they had a headache, and the other 150 people received the old formula medication. The time it took, in minutes, for each patient to no longer have a headache was recorded. The results from both of these clinical trials are shown below. Items 11, 12, and 13 present statements made by three different statistics students. For each statement, indicate whether you think the student's conclusion is valid.



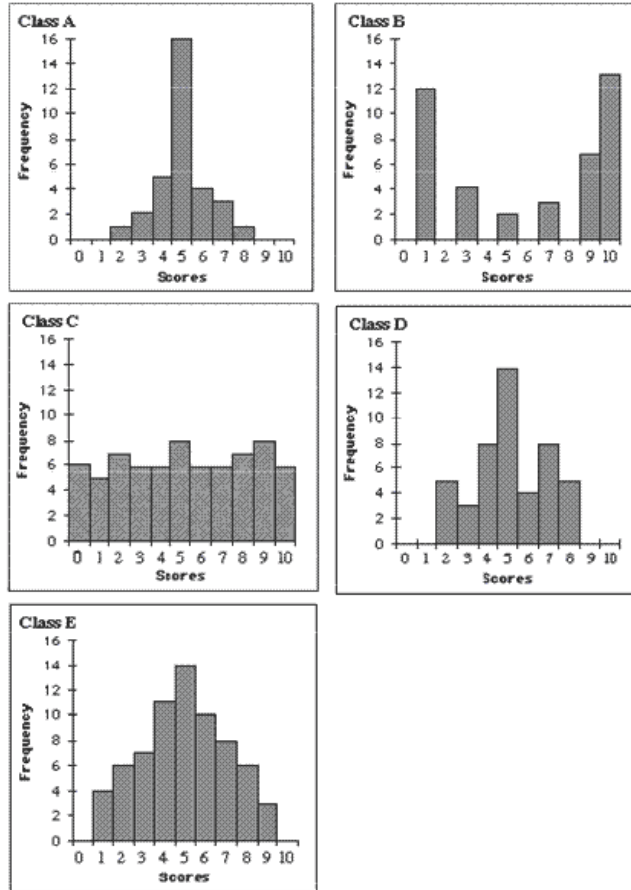
11. The old formula works better. Two people who took the old formula felt relief in less than 20 minutes, compared to none who took the new formula. Also, the worst result - near 120 minutes - was with the new formula.
 - a. Valid.
 - b. Not valid.

12. The average time for the new formula to relieve a headache is lower than the average time for the old formula. I would conclude that people taking the new formula will tend to feel relief about 20 minutes sooner than those taking the old formula.
 - a. Valid.
 - b. Not valid.

13. I would not conclude anything from these data. The number of patients in the two groups is not the same so there is no fair way to compare the two formulas.
 - a. Valid.
 - b. Not valid.

Items 14 and 15 refer to the following situation:

Five histograms are presented below. Each histogram displays test scores on a scale of 0 to 10 for one of five different statistics classes.



14. Which of the classes would you expect to have the lowest standard deviation, and why?
- Class A, because it has the most values close to the mean.
 - Class B, because it has the smallest number of distinct scores.
 - Class C, because there is no change in scores.
 - Class A and Class D, because they both have the smallest range.
 - Class E, because it looks the most normal.

15. Which of the classes would you expect to have the highest standard deviation, and why?
- Class A, because it has the largest difference between the heights of the bars.
 - Class B, because more of its scores are far from the mean.
 - Class C, because it has the largest number of different scores.
 - Class D, because the distribution is very bumpy and irregular.
 - Class E, because it has a large range and looks normal.
16. A certain manufacturer claims that they produce 50% brown candies. Sam plans to buy a large family size bag of these candies and Kerry plans to buy a small fun size bag. Which bag is more likely to have more than 70% brown candies?
- Sam, because there are more candies, so his bag can have more brown candies.
 - Sam, because there is more variability in the proportion of browns among larger samples.
 - Kerry, because there is more variability in the proportion of browns among smaller samples.
 - Kerry, because most small bags will have more than 50% brown candies.
 - Both have the same chance because they are both random samples.
17. Imagine you have a barrel that contains thousands of candies with several different colors. We know that the manufacturer produces 35% yellow candies. Five students each take a random sample of 20 candies, one at a time, and record the percentage of yellow candies in their sample. Which sequence below is the most plausible for the percent of yellow candies obtained in these five samples?
- 30%, 35%, 15%, 40%, 50%.
 - 35%, 35%, 35%, 35%, 35%.
 - 5%, 60%, 10%, 50%, 95%.
 - Any of the above.

18. Jean lives about 10 miles from the college where she plans to attend a 10-week summer class. There are two main routes she can take to the school, one through the city and one through the countryside. The city route is shorter in miles, but has more stoplights. The country route is longer in miles, but has only a few stop signs and stoplights. Jean sets up a randomized experiment where each day she tosses a coin to decide which route to take that day. She records the following data for 5 days of travel on each route.

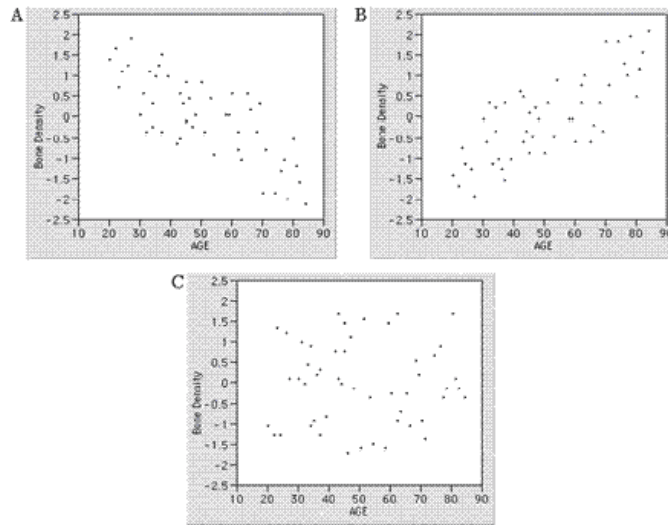
Country Route - 17, 15, 17, 16, 18

City Route - 18, 13, 20, 10, 16

It is important to Jean to arrive on time for her classes, but she does not want to arrive too early because that would increase her parking fees. Based on the data gathered, which route would you advise her to choose?

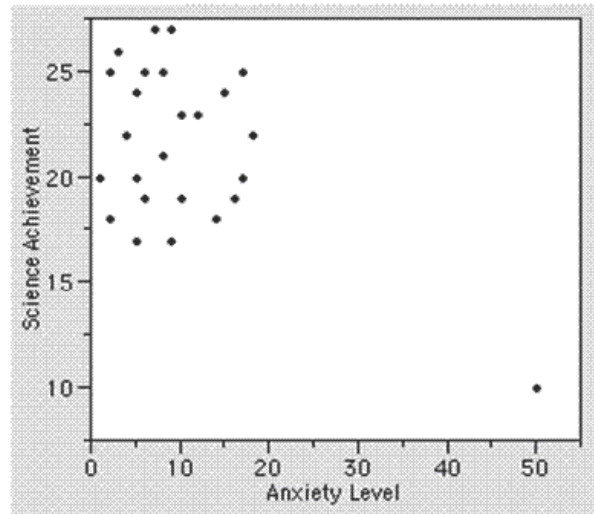
- a. The Country Route, because the times are consistently between 15 and 18 minutes.
 - b. The City Route, because she can get there in 10 minutes on a good day and the average time is less than for the Country Route.
 - c. Because the times on the two routes have so much overlap, neither route is better than the other. She might as well flip a coin.
19. A graduate student is designing a research study. She is hoping to show that the results of an experiment are statistically significant. What type of p -value would she want to obtain?
- a. A large p -value.
 - b. A small p -value.
 - c. The magnitude of a p -value has no impact on statistical significance.

20. Bone density is typically measured as a standardized score with a mean of 0 and a standard deviation of 1. Lower scores correspond to lower bone density. Which of the following graphs shows that as women grow older they tend to have lower bone density?



- a. Graph A.
- b. Graph B.
- c. Graph C.

21. The following scatterplot shows the relationship between scores on an anxiety scale and an achievement test for science. Choose the best interpretation of the relationship between anxiety level and science achievement based on the scatterplot.



- a. This graph shows a strong negative linear relationship between anxiety and achievement in science.
- b. This graph shows a moderate linear relationship between anxiety and achievement in science.
- c. This graph shows very little, if any, linear relationship between anxiety and achievement in science.
22. Researchers surveyed 1,000 randomly selected adults in the U.S. A statistically significant, strong positive correlation was found between income level and the number of containers of recycling they typically collect in a week. Please select the best interpretation of this result.
- a. We can not conclude whether earning more money causes more recycling among U.S. adults because this type of design does not allow us to infer causation.
- b. This sample is too small to draw any conclusions about the relationship between income level and amount of recycling for adults in the U.S.
- c. This result indicates that earning more money influences people to recycle more than people who earn less money.

Items 23 and 24 refer to the following situation:

A researcher in environmental science is conducting a study to investigate the impact of a particular herbicide on fish. He has 60 healthy fish and randomly assigns each fish to either a treatment or a control group. The fish in the treatment group showed higher levels of the indicator enzyme.

23. Suppose a test of significance was correctly conducted and showed no statistically significant difference in average enzyme level between the fish that were exposed to the herbicide and those that were not. What conclusion can the graduate student draw from these results?
- The researcher must not be interpreting the results correctly; there should be a significant difference.
 - The sample size may be too small to detect a statistically significant difference.
 - It must be true that the herbicide does not cause higher levels of the enzyme.
24. Suppose a test of significance was correctly conducted and showed a statistically significant difference in average enzyme level between the fish that were exposed to the herbicide and those that were not. What conclusion can the graduate student draw from these results?
- There is evidence of association, but no causal effect of herbicide on enzyme levels.
 - The sample size is too small to draw a valid conclusion.
 - He has proven that the herbicide causes higher levels of the enzyme.
 - There is evidence that the herbicide causes higher levels of the enzyme for these fish.

Items 25 to 27 refer to the following situation:

A research article reports the results of a new drug test. The drug is to be used to decrease vision loss in people with Macular Degeneration. The article gives a p -value of .04 in the analysis section. Items 25, 26, and 27 present three different interpretations of this p -value. Indicate if each interpretation is valid or invalid.

25. The probability of getting results as extreme as or more extreme than the ones in this study if the drug is actually not effective.
 - a. Valid.
 - b. Invalid.
26. The probability that the drug is not effective.
 - a. Valid.
 - b. Invalid.
27. The probability that the drug is effective.
 - a. Valid.
 - b. Invalid.

Items 28 to 31 refer to the following situation:

A high school statistics class wants to estimate the average number of chocolate chips in a generic brand of chocolate chip cookies. They collect a random sample of cookies, count the chips in each cookie, and calculate a 95% confidence interval for the average number of chips per cookie (18.6 to 21.3). Items 28, 29, 30 and 31 present four different interpretations of these results. Indicate if each interpretation is valid or invalid.

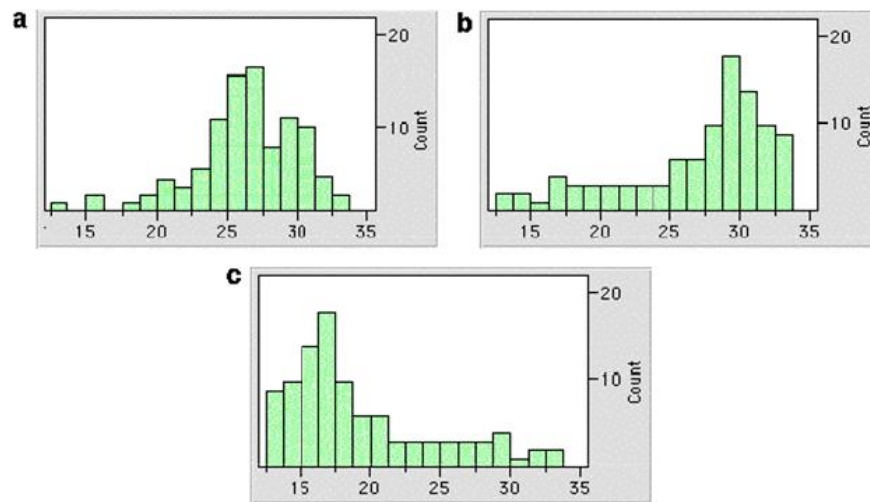
28. We are 95% certain that each cookie for this brand has approximately 18.6 to 21.3 chocolate chips.
 - a. Valid.
 - b. Invalid.
29. We expect 95% of the cookies to have between 18.6 and 21.3 chocolate chips.
 - a. Valid.
 - b. Invalid.
30. We would expect about 95% of all possible sample means from this population to be between 18.6 and 21.3 chocolate chips.
 - a. Valid.
 - b. Invalid.
31. We are 95% certain that the confidence interval of 18.6 to 21.3 includes the true average number of chocolate chips per cookie.
 - a. Valid.
 - b. Invalid.

32. It has been established that under normal environmental conditions, adult largemouth bass in Silver Lake have an average length of 12.3 inches with a standard deviation of 3 inches. People who have been fishing Silver Lake for some time claim that this year they are catching smaller than usual largemouth bass. A research group from the Department of Natural Resources took a random sample of 100 adult largemouth bass from Silver Lake and found the mean of this sample to be 11.2 inches. Which of the following is the most appropriate statistical conclusion?
- The researchers cannot conclude that the fish are smaller than what is normal because 11.2 inches is less than one standard deviation from the established mean (12.3 inches) for this species.
 - The researchers can conclude that the fish are smaller than what is normal because the sample mean should be almost identical to the population mean with a large sample of 100 fish.
 - The researchers can conclude that the fish are smaller than what is normal because the difference between 12.3 inches and 11.2 inches is much larger than the expected sampling error.

A study examined the length of a certain species of fish from one lake. The plan was to take a random sample of 100 fish and examine the results. Numerical summaries on lengths of the fish measured in this study are given.

Mean	26.8mm
Median	29.4mm
Standard Deviation	5.0 mm
Minimum	12.m mm
Maximum	33.4mm

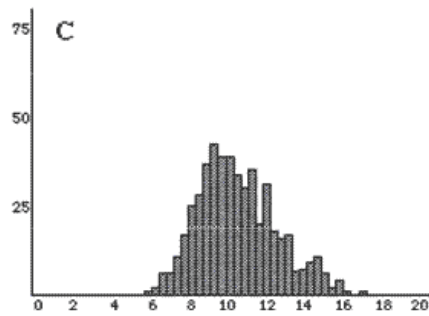
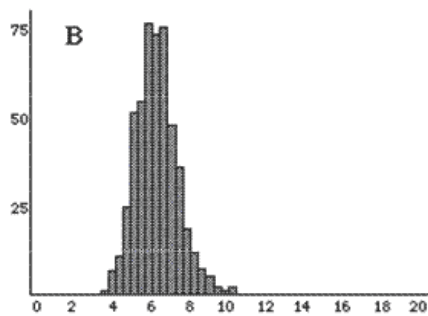
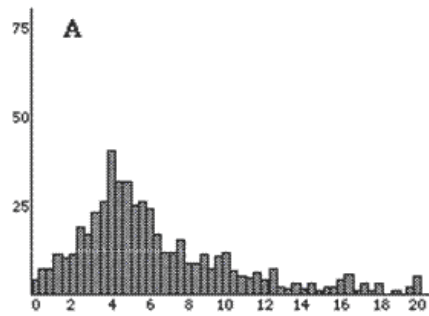
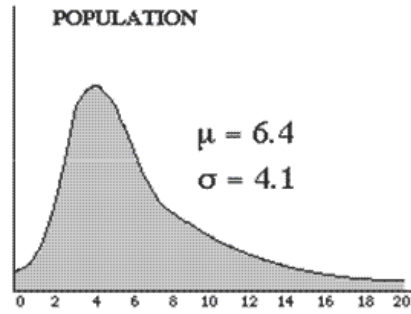
33. Which of the following histograms is most likely to be the one for these data?



- a. Histogram a.
- b. Histogram b.
- c. Histogram c.

Items 34 and 35 refer to the following situation:

Four graphs are presented below. The graph at the top is a distribution for a population of test scores. The mean score is 6.4 and the standard deviation is 4.1.



34. Which graph (A, B, or C) do you think represents a single random sample of 500 values from this population?
- Graph A
 - Graph B
 - Graph C
35. Which graph (A, B, or C) do you think represents a distribution of 500 sample means from random samples each of size 9?
- Graph A
 - Graph B
 - Graph C
36. This table is based on records of accidents compiled by a State Highway Safety and Motor Vehicles Office. The Office wants to decide if people are less likely to have a fatal accident if they are wearing a seatbelt. Which of the following comparisons is most appropriate for supporting this conclusion?

Safety Equipment in Use	Injury		ROW TOTAL
	Nonfatal	Fatal	
Seat Belt	412,368	510	412,878
No Seat Belt	162,527	1,601	164,128
COLUMN TOTAL	574,895	2,111	577,006

- Compare the ratios $510/412,878$ and $1,601/164,128$
- Compare the ratios $510/577,006$ and $1,601/577,006$
- Compare the numbers 510 and 1,601

37. A student participates in a Coke versus Pepsi taste test. She correctly identifies which soda is which four times out of six tries. She claims that this proves that she can reliably tell the difference between the two soft drinks. You have studied statistics and you want to determine the probability of anyone getting at least four right out of six tries just by chance alone. Which of the following would provide an accurate estimate of that probability?
- Have the student repeat this experiment many times and calculate the percentage time she correctly distinguishes between the brands.
 - Simulate this on the computer with a 50% chance of guessing the correct soft drink on each try, and calculate the percent of times there are four or more correct guesses out of six trials.
 - Repeat this experiment with a very large sample of people and calculate the percentage of people who make four correct guesses out of six tries.
 - All of the methods listed above would provide an accurate estimate of the probability.
38. A college official conducted a survey to estimate the proportion of students currently living in dormitories about their preference for single rooms, double rooms, or multiple (more than two people) rooms in the dormitories on campus. Which of the following does NOT affect the college official's ability to generalize the survey results to all dormitory students?
- Five thousand students live in dormitories on campus. A random sample of only 500 were sent the survey.
 - The survey was sent to only first-year students.
 - Of the 500 students who were sent the survey, only 160 responded.
 - All of the above present a problem for generalizing the results.
39. The number of people living on American farms has declined steadily during the last century. Data gathered on the U.S. farm population (millions of people) from 1910 to 2000 were used to generate the following regression equation: Predicted Farm Population = $1167 - .59(\text{YEAR})$. Which method is best to use to predict the number of people living on farms in 2050?
- Substitute the value of 2050 for YEAR in the regression equation, and compute the predicted farm population.
 - Plot the regression line on a scatterplot, locate 2050 on the horizontal axis, and read off the corresponding value of population on the vertical axis.
 - Neither method is appropriate for making a prediction for the year 2050 based on these data.
 - Both methods are appropriate for making a prediction for the year 2050 based on these data.

40. The following situation models the logic of a hypothesis test. An electrician uses an instrument to test whether or not an electrical circuit is defective. The instrument sometimes fails to detect that a circuit is good and working. The null hypothesis is that the circuit is good (not defective). The alternative hypothesis is that the circuit is not good (defective). If the electrician rejects the null hypothesis, which of the following statements is true?
- a. The circuit is definitely not good and needs to be repaired.
 - b. The electrician decides that the circuit is defective, but it could be good.
 - c. The circuit is definitely good and does not need to be repaired.
 - d. The circuit is most likely good, but it could be defective.

CAOS 4 ANSWER KEY

1. D	11. B	21. C	31. A
2. B	12. A	22. A	32. C
3. C	13. B	23. B	33. B
4. A	14. A	24. D	34. A
5. D	15. B	25. A	35. B
6. C	16. C	26. B	36. A
7. D	17. A	27. B	37. B
8. A	18. A	28. B	38. A
9. D	19. B	29. B	39. C
10. C	20. A	30. B	40. B

Demographic and Academic Questionnaire

Screening questions

1. Are you over 18 years of age Yes/No
2. Are you repeating Statistics 2023 Yes/No

Demographic and Academic Questions

1. What is your age? (enter age in years)
2. What is your gender? Male/Female/Other
3. Which race best describes you (please choose only one)
 - American Indian or Alaskan Native
 - Black or African America
 - Hispanic
 - White/Caucasian
 - Multiple ethnicity (please specify)
4. What is your mother tongue (first language spoken)?
 - English
 - Spanish
 - Chinese
 - Russian
 - Vietnamese
 - Others
5. What is your class standing? Freshman Sophomore Junior Senior
6. Are you a first-generation college student? Yes/No
7. Are you a full time or part time student? Full time/Part time
8. What is your major? (enter major)
9. Do you receive financial aid? Yes/No
10. What was your high school GPA?
11. What is your current GPA? (enter GPA)
12. What is the highest math class you took in high school?
 - Pre-algebra
 - Algebra I

- Algebra II
- Pre-Calculus
- Analysis of Function
- Calculus
- Other (please specify)

13. Have you taken other statistical classes before? Yes/No

APPENDIX E: ADDITIONAL RESULTS TABLES

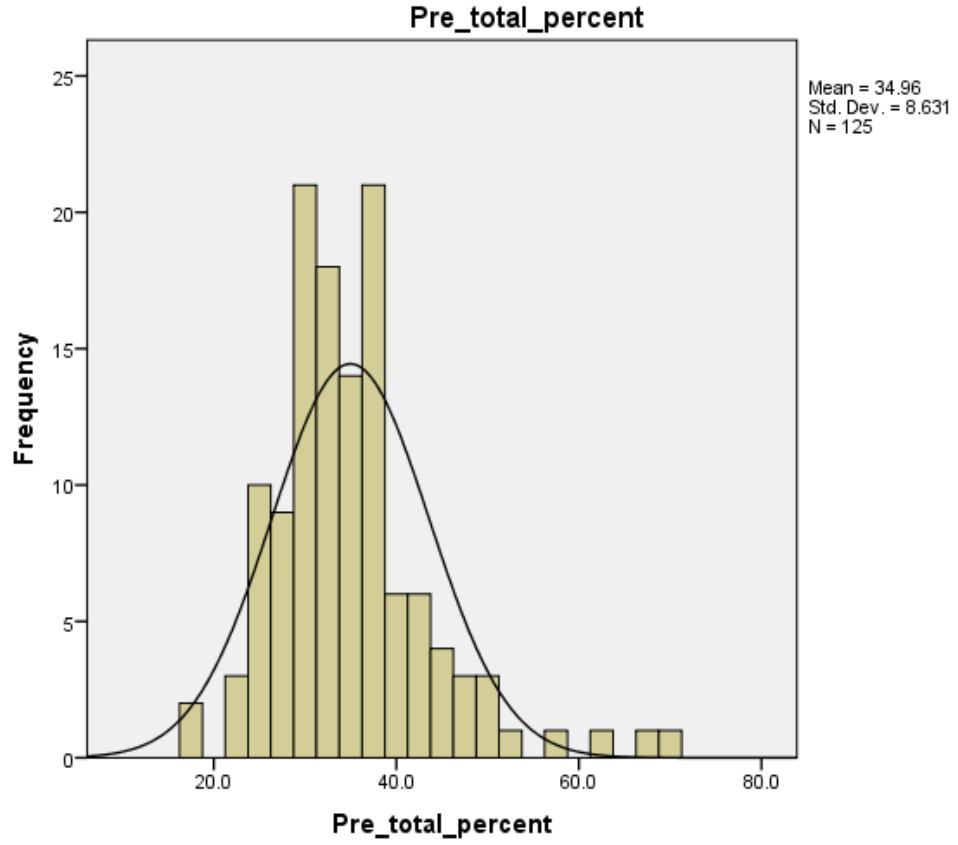


Figure E1: Histogram of Pre-Total CAOS Percent Score

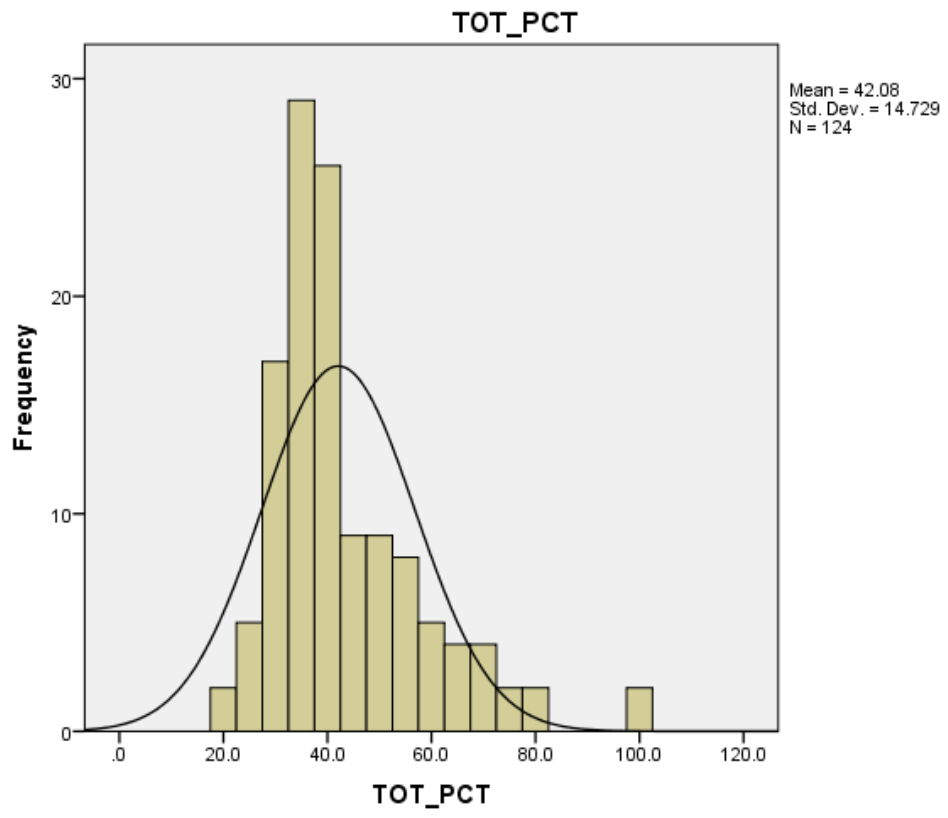


Figure E2: Histogram of Post-Total CAOS Percent Score

Table E1

Full Table Regression for Regression 1

Model	B	Std. Error	Beta	t	Sig.	Collinearity Statistics	
						Tolerance	IF
(Constant)	16.990	5.011		3.391	.001		
Group	10.664	2.345	.360	4.547	.000	.998	1.002
Pre CAOS Percent	.544	.135	.318	4.017	.000	.998	1.002
(Constant)	15.431	5.627		2.742	.007		
Group	11.048	2.365	.373	4.672	.000	.983	1.017
Pre CAOS Percent	.514	.145	.301	3.552	.001	.877	1.141
Gender	1.952	2.484	.063	.786	.434	.967	1.034
First language spoken	2.750	2.497	.092	1.101	.273	.895	1.117

a. Dependent Variable: Post CAOS Percent

Table E2

Full Table for Stepwise Regression

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	35.147	1.876		18.734	.000		
Group	11.746	2.467	.400	4.762	.000	1.000	1.000
(Constant)	19.493	5.408		3.605	.000		
Group	11.224	2.390	.382	4.697	.000	.995	1.005
Pre CAOS Percent	.457	.149	.250	3.072	.003	.995	1.005

a. Dependent Variable: Post CAOS Percent

Table E3

Full Table for Enter Method Regression

Model	B	Std. Error	Beta	t	Sig.	Tolerance	VIF
(Constant)	11.499	11.520		.998	.320		
Group	11.851	2.515	.404	4.712	.000	.882	1.133
Pre CAOS Percent	.412	.173	.225	2.379	.019	.723	1.383
Gender	.985	2.593	.033	.380	.705	.884	1.132
First language spoken	1.210	2.712	.041	.446	.657	.767	1.303
Age	.067	.205	.029	.329	.742	.829	1.207
First generation college student	-2.324	2.631	-.078	-.883	.379	.834	1.199
Full or part time student	3.264	2.971	.102	1.099	.274	.750	1.334
STEM Major	-2.247	2.465	-.078	-.912	.364	.896	1.116
Current GPA?	2.780	2.609	.090	1.066	.289	.912	1.096
Statistics class?	3.177	3.114	.089	1.020	.310	.856	1.169
Fin Aid	-3.732	2.634	-.124	-1.417	.159	.840	1.191
Hispanic	-2.457	3.587	-.063	-.685	.495	.766	1.305
Highest Math Class	3.296	2.814	.109	1.171	.244	.750	1.333

a. Dependent Variable: Post CAOS Percent

VITA

ERIC HERNANDEZ

Born, Stillwater, Oklahoma

- | | |
|-----------|---|
| 1995-2002 | B.S., Geophysics Engineering
Simon Bolivar University
Caracas, Venezuela |
| 2001-2002 | Junior Geophysics Engineer
Veritas Geophysics
Carcass, Venezuela |
| 2003-2004 | Mathematics Instructor
American Academy HS
Miami, USA |
| 2004-2008 | Professional Math Teacher
Booker T. Washington SH
Miami, Florida |
| 2006-2008 | M.S., Mathematics Education
Florida International University
Miami, Florida |
| 2008-2011 | Professional Math Teacher
Coral Way Middle School
Miami, Florida |
| 2009-2014 | Educational Specialist, Curriculum and Instruction.
NCATE Accredited
Florida International University
Miami, Florida |
| 2014-2018 | Doctoral Candidate
Florida International University
Miami, Florida |

2011-Present

Professor
Miami Dade College
Miami, FL

PUBLICATIONS AND PRESENTATIONS

Hernandez, E (2017, April). Bootstraps: An Intuitive Approach to Confidence Intervals.

Presenter. 15th Annual Mathematics Retreat of MDC. Miami, FL.

Hernandez, E (2012, November). Randomization Test for Hypothesis Testing.

Presenter. AMATYC Annual Conference. Jacksonville, FL.

Hernandez, E (2010, August). PISA, TIMSS, and PIRLS Database Seminar.

Presenter National Center for Education Statistics. Washington, DC.