SKEW AND BIAS: THE EFFICIACY OF AN INTERVENTION IN AN INTRODUCTORY STATISTICS COURSE

by

KRISTA VARANYAK

(Under the Direction of Jennifer Kaplan)

ABSTRACT

The focus of this paper is a study designed to help students overcome the lexical ambiguity associated with the terms skew and bias. In the study, an activity to address the difference between the everyday and statistical definitions of skew was developed and implemented by an instructor of introductory statistics. Open-ended definitions of the target words from students who saw the activity in class and students who did not see the activity in class were compared. The results from this study suggest that the activity did not help students to provide better complete statistical definitions of the target words. It is recommended that the activity be updated to combat misconceptions of introductory statistics students about the words skew and bias better than the activity studied.

INDEX WORDS: Statistics Education, Statistical Literacy, Lexical Ambiguity, Skew, Bias
SKEW AND BIAS: THE EFFICIACY OF AN INTERVENTION IN AN INTRODUCTORY STATISTICS COURSE

By

KRISTA VARANYAK
B.S., Monmouth University, 2015
B.A., Monmouth University, 2015

A Thesis Submitted to the Graduate Faculty of The University of Georgia in Partial Fulfillment of the Requirements for the Degree

MASTER OF SCIENCE

ATHENS, GEORGIA

2017
SKEW AND BIAS: THE EFFICIACY OF AN INTERVENTION IN AN INTRODUCTORY STATISTICS COURSE

By

KRISTA VARANYAK

Major Professor: Jennifer Kaplan

Committee: Jaxk Reeves
             Cheolwoo Park

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
May 2017
DEDICATION

I would like to dedicate this thesis to my friends, family, and mentors whose ever-present support made its completion possible. I want to dedicate this to for my mother who is the reason I will continue to pursue a career in education. Finally, I would like to dedicate to my amazing grandmother, who always supported me for pursuing my master’s at UGA, but was not able to see its completion.
ACKNOWLEDGEMENTS

First I want to recognize my advisor Dr. Kaplan; because of her tireless commitment and dedication to this project, my work has significantly improved. I would also like to acknowledge Team Kaplan (Alex, Gabe, Kristen, and Oguz) for without their mentorship and encouragement I would not have completed this thesis. From answering every little question I had about my methods to offering support and guidance to make sure I stayed on track; I cannot thank you enough. Finally, I would like to acknowledge Mr. Schlaegel and Dr. Bastian who introduced me to the wonderful world of statistics in high school and college, respectively. Their passion for teaching statistics inspired me to continue on this path. Without their teaching, I would not be here today.

This thesis was completed as part of NSF DUE Grant Number 1504013. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS ................................................................. v</td>
</tr>
<tr>
<td>LIST OF TABLES ........................................................................ viii</td>
</tr>
<tr>
<td>LIST OF FIGURES ...................................................................... ix</td>
</tr>
<tr>
<td>1. INTRODUCTION ...................................................................... 1</td>
</tr>
<tr>
<td>2. BACKGROUND ....................................................................... 5</td>
</tr>
<tr>
<td>2.1 Vocabulary Acquisition as a Goal for Introductory Statistics Students ................. 5</td>
</tr>
<tr>
<td>2.2 Defining <em>Skew</em> and <em>Bias</em> in an Introductory Statistics Course ..................... 12</td>
</tr>
<tr>
<td>2.3 Assessment and Intervention in an Introductory Statistics Classroom ............ 19</td>
</tr>
<tr>
<td>3. METHODOLOGY .................................................................. 24</td>
</tr>
<tr>
<td>3.1 Goals of the Research Study .............................................. 24</td>
</tr>
<tr>
<td>3.2 Subjects and Settings ......................................................... 24</td>
</tr>
<tr>
<td>3.3 Development of HILT Activity ............................................ 26</td>
</tr>
<tr>
<td>3.4 Instrumentation ................................................................. 28</td>
</tr>
<tr>
<td>3.5 Analysis .......................................................................... 30</td>
</tr>
<tr>
<td>4. RESULTS .......................................................................... 34</td>
</tr>
<tr>
<td>4.1 Comparing Student Definitions of <em>Skew</em> ................................... 34</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table 1: Textbook Introduction of *Skew* ................................................................. 14
Table 2: Assessment Prompts ...................................................................................... 29
Table 3: Coding Rubric for Student Definitions of *Skew* ........................................ 31
Table 4: Coding Rubric for Student Definitions of *Bias* ........................................... 32
Table 5: Definitions for *Skew* Given by Students ..................................................... 35
Table 6: Definitions for *Bias* Given by Students ....................................................... 37
Table 7: Final Exam Responses .................................................................................... 39
Table 8: Final Exam Analysis ....................................................................................... 39
LIST OF FIGURES

Figure 1: Definitions for Skew Given by Students ................................................................. 35

Figure 2: Definitions of Bias Given by Students ................................................................. 37
1. INTRODUCTION

Ordinary citizens are exposed to data daily through various mediums and should be able to assess it critically. To be considered good statistical citizens, they may need to improve their level of statistical literacy (Gal, 2002; Utts, 2003; Gould, 2010). Citizens, therefore, need a place to improve and grow their statistical literacy: for many, the place is an introductory statistics classroom. Enrollment in introductory statistics courses in the United States grew at increasingly high rates over the last decade. According to the 2010 survey report sponsored by the Conference Board of the Mathematical Sciences (CBMS, 2013), enrollment in elementary statistics at four-year universities exceeded those of 2005 by 56% in mathematics departments and by 50% in statistics departments, with over 200,000 students enrolled. As pointed out by Utts (2003) and Gal (2002) this is where many students learn statistical ideas.

The goals of an introductory statistics course should reflect the overall objective of creating statistically educated citizens. The 2016 Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report lists nine student-level goals for an introductory statistics course:

1. Students should become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted.

2. Students should be able to recognize questions for which the investigative process in statistics would be useful and should be able to answer questions using the investigative process.
3. Students should be able to produce graphical displays and numerical summaries and interpret what graphs do and do not reveal.

4. Students should recognize and be able to explain the central role of variability in the field of statistics.

5. Students should recognize and be able to explain the central role of randomness in designing studies and drawing conclusions.

6. Students should gain experience with how statistical models, including multivariable models, are used.

7. Students should demonstrate an understanding of, and ability to use, basic ideas of statistical inference, both hypothesis tests and interval estimation, in a variety of settings.

8. Students should be able to interpret and draw conclusions from standard output from statistical software packages.

9. Students should demonstrate an awareness of ethical issues associated with sound statistical practice.

As will be emphasized in Section 2.1, all of these goals require minimal fluency in statistical literacy. Additionally, the first goal is linked directly to the ultimate goal of producing educated statistics citizens. Reaching general statistical literacy through the understanding of the language used in statistics is instrumental in achieving the goals outlined in the 2016 GAISE College Report (Rumsey, 2002).

The most basic level of statistical literacy requires comprehension of the definitions statistical terms (Ben-Zvi & Garfield, 2004). Many words in statistics, however, are considered lexically ambiguous, generally meaning the terms have a definition in statistics that differs from
the definition to which students are accustomed in colloquial language (Kaplan, Rogness, & Fisher, 2009). To help students overcome this issue, instructors should help students integrate the correct statistical definitions with their prior knowledge of a colloquial definition (Bruning, Schraw & Norby, 2011).

This thesis will focus specially on a study designed to help students overcome the lexical ambiguity associated with the terms skew and bias. Skew and bias are terms in statistics that considered lexically ambiguous for several reasons. First, in ordinary English (OE), they may be used synonymously, meaning having a preference for one group over another. Additionally, each word has several other OE definitions. Finally, in statistical English (SE), skew means the state of being not symmetric and bias means the statistic does not accurately reflect the parameter. Because the words are synonymous in OE, but have contrasting SE definitions, they are lexically ambiguous and may present a problem for students who are trying to build statistical literacy.

An activity was created to help students distinguish between the OE definition of skew and its SE definition. In the study, selected students in an introductory statistics course either were or were not exposed to the activity in class. At the end of the semester, students were given an assessment in which they were asked to define skew and bias in an open-ended question. The student responses were categorized into varying levels of complete, incomplete, or incorrect definitions. The student definitions of those who were exposed to the activity and those who were not exposed to the activity then were compared to assess the efficacy of the activity.

The background in statistical literacy and the developing field of lexical ambiguity in statistics, specifically as it pertains to the words skew and bias is provided in Chapter 2. The goals of the intervention study and the methodology for data collection and analysis are described in Chapter 3. The results of the analysis of the data from this study are presented in
Chapter 4. To conclude, suggestions for instruction in an introductory statistics course and recommendations for developing further research in this field are offered in Chapter 5.
2. BACKGROUND

2.1 Vocabulary Acquisition as a Goal for Introductory Statistics Students

The primary instructional focus of an undergraduate statistics course has shifted in the last 25 years. Early statistics courses focused heavily on teaching computations, mainly due to the fact that the majority of students were practicing scientists and relied on cutting edge techniques (Utts, 2003). Over the years, however, with development of new technologies and more investigation in the field of statistics education, many researchers believe that developing statistical literacy by focusing on concepts rather than computations, should be a goal of an introductory statistics course (Cobb, 1992; Gould, 2010; Moore, 1997; Rumsey, 2002).

Reform in introductory statistics began in 1992 at the Mathematical Association of America Meeting. Cobb (1992) notes the need for consistency in instruction based in education research. He compiled the first set of guidelines for teaching undergraduate introductory statistics that resulted from this meeting. These guidelines include three recommendations: 1) Emphasize statistical thinking; 2) More data and concepts: Less theory, fewer recipes; and 3) Foster active learning. These recommendations were updated and expanded most recently in 2016 in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report produced by The American Statistical Association. The 2016 GAISE College Report includes six recommendations: 1) Teach statistical thinking; 2) Focus on conceptual understanding, 3) Integrate real data with a context and a purpose, 4) Foster active learning, 5) Use technology to explore concepts and analyze data; 6) Use assessment to improve and evaluate student learning (GAISE College Report ASA Revision Committee 2016).
The guidelines make it clear that computations should not be the foundation of a statistics course. Instead, minor computations should be used as tools to aide students in developing statistical thinking. Most statistics educators now agree that statistical literacy is the baseline goal for any introductory course (Watson, 1997; Rumsey, 2002). In recent years, however, a new debate in defining statistical literacy has arisen. Rumsey (2002) summarizes the various definitions of statistical literacy, noting that although statistical literacy is not defined consistently across educators, there are two key learning objectives mentioned in the literature that she names statistical competence and statistical citizenship. Statistical competence includes the understanding of basic statistical knowledge such as data awareness, understanding of basic statistical concepts and terminology, and basic communication and interpretation skills. Statistical citizenship is defined as the ability to function in the data-driven world, which requires development of a high level of statistical reasoning and thinking. Ben-Zvi & Garfield (2004) further elaborate on Rumsey’s ideas of statistical literacy, statistical competence, and statistical citizenship. After noting again that the definitions are not consistent among educators, they summarize the ideas presented by Rumsey (2002) and others into three separate levels of student understanding: statistical literacy, statistical reasoning, and statistical thinking. Statistical literacy is defined similarly to the way Rumsey (2002) defines statistical competence, but does not include an ability to interpret data or results. Statistical reasoning involves the mastery of statistical literacy and the ability to connect concepts and interpret results. Finally, statistical thinking is similarly defined to statistical citizenship: students should not only understand how statistical studies are conducted, but more importantly, why they are done and how the main ideas of statistics guide research. Therefore, the stepping stone to achieving the highest level of
statistical understanding is to first understand the basic concepts and terminology. This belief is emphasized in the 2016 GAISE College Report recommendations.

The 2016 GAISE College Report further acknowledges that there is not one specific type of introductory course, pointing out that some courses are designed for consumers of statistics and some are for producers of statistics. In general, statistical consumers are classified as people who will not use statistical analysis in their future career or everyday life, while statistical producers are expected to use analysis and possibly computation-heavy statistics in their futures (Gal, 2002; Gould, 2010). Producers of statistics are expected to achieve statistical thinking and will carry out high level analyses in their everyday life. They are, therefore, expected to take more than one introductory level course. Since introductory courses are likely to be composed of both producers and consumers, the course ought to be structured to prepare all to function successfully in the world influenced by the understanding of data (Gal & Ginsburg, 1994). This means, regardless of a students’ perceived classification or their future plans, it is imperative that all students develop statistical literacy to be good statistics citizens (Gal, 2002; Utts, 2003; Gould, 2010).

Proficiency in statistical literacy is at the foundation of understanding the plethora of research studies that are dispersed through the media every day (Watson, 1997). Researchers work to execute and report meaningful research studies to peer-reviewed journals to ensure publications are accurate and to provide researchers a chance to defend their findings to criticism (Benos, et al., 2007). The average citizen, however, learns about current news by watching television news stations, reading the newspaper, or sharing articles on social media (Pew Research Center, 2016). Additionally, Pew Research Center (2016) found 22% of American consumers “don’t give much thought to the sources [they] get [their] news from.” Journalists
across media platforms, therefore, have the challenge of relaying research news to the general public, but are not held accountable by some consumers for the claims. Unfortunately, journalists often make incorrect report summaries, whether intentional or not (Fleishman, 2002). Dr. Ben Goldacre, a Senior Clinical Research Fellow at the Centre for Evidence Based Medicine in the Department of Primary Care in the University of Oxford, and a Research Fellow in Epidemiology at London School of Hygiene & Tropical Medicine, has even dedicated his time to educate everyday citizens through books (Bad Science), talks (TEDx), and the website badscience.net about how easy it is for journalists to make incorrect summaries about science research; and more interestingly how often it happens.

Regardless of how a person consumes news, individuals should think critically about the information presented to them. Furthermore, the first goal listed in the 2016 GAISE College Report is “Students should become critical consumers of statistically-based results reported in popular media, recognizing whether reported results reasonably follow from the study and analysis conducted” (pg. 8). To address this goal directly in the classroom, Watson (1997) suggests implementing media-based activities and assessments throughout the semester for topics like graphical representation and sampling. Tackling this issue head on in the classroom is important for students to grow as educated statistics citizens. Utts (2003) speaks to the growing problem of adults’ lack of critical evaluation skills in the context of assessing the news and states seven common topics that are most misunderstood by the general population: (1) concluding cause and effect relationships; (2) differentiation between statistical significance and practical importance; (3) differentiating between finding “no effect” and no statistical effect; (4) common sources of bias in surveys and experiments; (5) the idea that coincidence are not uncommon; (6) confusing conditional probability in one direction with conditional probability in the other
direction; and (7) understanding variability is natural, and that “normal is not the same as “average”. Utts (2003) and Gal (2002) agree that the best place to address these misunderstood topics and build statistical literacy is in an introductory statistics course. In addition, they suggest this can be done by focusing instruction on how and why research studies are conducted and subsequent results are interpreted.

While it might seem obvious that the focus of an introductory course should be on generally misunderstood topics, there is still another barrier for students to understand these concepts fully, which ultimately prevents them from becoming good statistical citizens. To assess and criticize the data around them properly, students must learn an entirely new set of vocabulary and be able to use that vocabulary fluently (Gal, 2002). Learning this new language may present a barrier to learning statistical ideas (McLean, 2002). This barrier may be partly due to the existence and interaction between three categories of definitions related to statistical terminology: Statistical English, Mathematical English, and Ordinary English. Statistical English (SE) is the vocabulary and symbols used in statistics and Mathematical English (ME), a term created by Kane, Byrne, and Hater (1974), is the specific vocabulary used in mathematics. SE and ME are different from each other and both are different from the third type of definition, Ordinary English (OE), which is the way words used in colloquial language. Rangecroft (2002) categorizes statistical terms into three categories: (i) words which have a meaning only in SE; (ii) words which occur in both OE and SE, but have a different meaning in SE from their meaning in OE; and (iii) words which occur in both ME and SE, but have a different meaning in SE from their meaning in ME. Any word in SE that has a meaning in ME and/or OE, is said to have lexical ambiguity (Barwell, 2005). In addition to these categories, there are pairs of words in SE that are synonyms within OE, but have different meanings within SE (Kaplan, et al., 2009). The
combination of OE synonyms being used to convey different statistics concepts and the newness of SE in general, poses a threat to achieving even the most basic level of statistical literacy through the achievement of understanding basic terminology and concepts (Rumsey, 2002). Therefore, language not only poses a barrier to understanding SE, but also creates difficulty in the acquisition of statistical literacy and prevents people from becoming good statistical citizens (Gal, 2003; Utts, 2003).

An introductory statistics course, where the majority of people learn about the discipline, is the best place to address lexical ambiguity in statistics. Based on Rangecroft’s categorization of SE, it can be claimed that many vocabulary words used in an introductory statistics course exhibit lexical ambiguity. Kaplan, et al. (2009) acknowledge the need to study lexically ambiguous words in an introductory statistics course by providing a list of 32 words believed to hold the characteristics associated with lexical ambiguity based on instructor experience. The list includes association, average, bias, confidence, random, significance, spread, and variance. It is important to note that these words are directly related to the seven topics established by Utts (2003). Therefore, researchers found it important to study student understanding of these words. Kaplan, Fisher, & Rogness (2010) gathered student definitions at the end of an introductory course of the words association, average, confidence, random, and spread and found that even at the end of the course; many students were not able to define these words correctly, often defining them in a colloquial context. Richardson, Dunn, & Hutchins (2013) conducted a similar study in which student tutors were asked list words they believed to be lexically ambiguous. For the most part, the list mirrored that of Kaplan, et al. (2009); it did not, however, include seven of the words. Introductory statistics students were then asked to define nine of the lexically ambiguous words in the context of different research articles: association, relationship, significant, average,
control, mean, intervention, correlation, and randomized. While there was a slight increase in the percent of students who correctly defined the words at the end of the semester, Richardson, et al. (2013) once again found that even at the end of the semester with tutor support, many students still defined words incorrectly or with a colloquial definition. Whitaker (2016) collected post-instruction definitions from Advanced Placement (AP) statistics students for the words arbitrary, model, random, range, uniform, and variable. He found that although the majority of the students from the class passed the AP exam at the end of the semester at a rate higher than the national average, many students were still unable to define the listed words properly in a statistical context. The studies listed were the only studies found with a focus on student definitions. Other methods to assess student understanding of basic terminology as well as intervention methods to combat lexical ambiguity will be described in Section 2.3.

Research on student understanding of words not listed above is lacking from the literature, so it is important that statistics educators continue to do research in this field. The remainder of this paper will focus on two words that have been identified as lexically ambiguous (Kaplan, et al., 2009; Richardson, et al., 2013), but not previously studied: skew and bias. These words can be considered lexically ambiguous because they are often used synonymously in OE. Merriam-Webster (merriam-webster.com) lists seven definitions of skew and ten definitions of bias. The definition of skew that is most common in OE on the site is “to distort especially from a true value.” The definition of bias that is most common in OE from the same site is “an inclination of temperament or outlook; especially: a personal and sometimes unreasoned judgment: prejudice.” Thesaurus.com explicitly lists bias as a synonym for skew. The statistical definitions of bias and skew will be discussed in the following section. Given that these words are synonyms in OE, but have different meanings in SE, it is no wonder that the SE meanings of
the words may cause problem for students when learning them in a statistical context. As previously mentioned, the foundation of statistical citizenship is statistical literacy, and these words pose a threat to statistical literacy. Instructors, therefore, should help students to overcome the lexically ambiguity of these words, so they can build their statistical literacy and eventually become good statistical citizens. The lexical ambiguity of *skew* and *bias* on their own and as a pair will be discussed further in Section 2.2.

2.2 Defining *Skew* and *Bias* in an Introductory Statistics Course

As stated in Section 2.1, *skew* and *bias* are used synonymously in OE, which contributes to their lexical ambiguity in SE. In statistics, *skew* is defined on merriam-webster.com as “not symmetrical” and *bias* is defined as “deviation of the expected value of a statistical estimate from the quantity it estimates.” Lexical ambiguity arises in these words because these definitions contradict the OE definition of the words and are different from each other. An efficient way to learn where misconceptions about *skew* and *bias* originate is to explore the typical curriculum of an introductory statistics course. Before we can assess how the terms *skew* and *bias* are linked to student learning outcomes of the course, we must understand how the terms are introduced a statistical context. This will be done by reviewing how they are introduced in commonly-used introductory statistics textbooks. Four textbooks were surveyed:

4. *Intro Stats, 3rd edition* by Richard D. De Veaux, Paul F. Velleman, and David E. Bock

While there is no particular mandate of topic coverage for an introductory statistic course, typical methods courses tend to address consistent student learning outcomes and use similar textbooks (GAISE College Report ASA Revision Committee 2016). Some common course objectives will be reviewed along with how the understanding of *skew* and *bias* is pivotal for fully achieving student understanding in an introductory course. Finally, an inclusive discussion of the lexical ambiguity of *skew* and *bias* on their own and as a pair will be provided along with a discussion of places in which standard instruction falls short at helping students achieve statistical literacy with the barrier of lexical ambiguity of these words.

The review of textbook definitions of *skew* and *bias* adds further insight to the lexical ambiguity of the words. In the four books that were reviewed, *skew* is introduced at least two chapters before *bias*. *Skew* is introduced in a chapter designated to cover the material associated with displaying data, mainly through graphical representations. All of the books present *skew* specifically when describing the shape of a distribution, and the word is used to contrast the definition of symmetric. Each book then defines *skewed right* or *skewed left* in varying language, although *Intro Stats* (de Veaux, et al., 2009) does not specifically use the term *skewed right* or *skewed left* (See Table 1). *Mind on Statistics* does not use the word distribution in the definition of *skew*, instead referring to “the shape of the dataset” (Utts & Heckard, 2004, pg. 28). The most comprehensive introduction to *skew* comes in *Statistics: The Art and Science of Learning from Data* in which *skew* is defined in a paragraph as, “the distribution is *skewed* if one side of the distribution stretches out longer than the other side” (Agresti & Franklin, 2009, pg. 43). The authors then provide pictures to illustrate a symmetric distribution are well as distributions that
are skewed right and skewed left. The figures are followed by a definition box summarizing skew, skewed to the left, and skewed to the right. Finally, there is an example of skew in the

Table 1: Textbook Introduction of Skew

<table>
<thead>
<tr>
<th>Textbook</th>
<th>Chapter Skew is Introduced</th>
<th>Defining “skewed right” and “skewed left”</th>
</tr>
</thead>
<tbody>
<tr>
<td>The Basic Practice of Statistics (6th edition) by David S. Moore,</td>
<td>Chapter 1: Picturing Distributions with Graphs</td>
<td>A distribution is skewed to the right if the right side of the histogram extends farther out than the left side. It is skewed to the left if the left side of the histogram extends much farther than the right side. (p 16)</td>
</tr>
<tr>
<td>William I. Notz, Michael A. Flinger, and R. Scott Linder</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mind on Statistics (2nd edition) by Jessica M. Utts and Robert F. Heckard</td>
<td>Chapter 2: Turning Data into Information</td>
<td>The shape of a dataset is either symmetric or skewed. A skewed dataset is either skewed to the left or skewed to the right. Figures 2.4 and 2.6 is called “skewed to the left” because the values trail off to the left. (p 28)</td>
</tr>
<tr>
<td>Statistics: The Art and Science of Learning from Data (4th edition)</td>
<td>Chapter 2: Exploring Data with Graphs and Numerical Summaries</td>
<td>A distribution is skewed to the left if the tail is longer than the right tail. A distribution is skewed to the right if the right tail is longer than the left tail. A left skewed distribution stretches to the left, a right skewed to the right. (p 43)</td>
</tr>
<tr>
<td>by Alan Agresti, Christine Franklin, and Bernhard Klingenberg, with contributions by Michael Posner</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intro Stats (3rd edition) by Richard D. De Veaux, Paul F. Velleman, and David E. Bock</td>
<td>Chapter 4: Displaying and Summarizing Quantitative Data</td>
<td>The (usually) thinner ends of a distribution are called the tails. If one tail stretches out farther than the other, the histogram is said to be skewed to the side of the longer tail. (p 54)</td>
</tr>
</tbody>
</table>

context of life span and income. Overall, it is clear from textbook introduction that skew in SE refers to the shape of nonsymmetrical data, where one side of the distribution stretches farther than the other creating what is known as the tail of the distribution. It is the direction of the tail that indicates the direction of the skew. Recognizing and identifying skew is, therefore, a piece of
achieving Goal 3: Students should not only be able create graphs, but also describe the shape of a distribution and discuss its significance in context (GAISE College Report 2016).

The word *bias* is introduced in a completely separate chapter from *skew*, standardly in a context that describes poorly collected samples and surveys. All of the textbooks surveyed first describe the difference between a sample and a population. *The Basic Practice of Statistics, Mind on Statistics* and *Intro Stats* then provide an example of a poorly collected sample, stating that since the sampling methods repeatedly miss the truth about the population, they are considered *biased*. These books then explain how the use of a simple random sample can eliminate *bias*. In contrast, *Statistics: The Art and Science of Learning from Data* introduces the idea of a simple random sample before defining *bias*. The explanation of simple random sampling is followed by a description of different types of *bias* that can occur if sampling and surveying are not done well. Like the first set of texts, the authors conclude that results from samples that favor one part of the population are said to have *bias* because the sample is not representative of the population. Overall, although each textbook introduces *bias* in a slightly different way, the term *bias* is linked to survey and sampling design consistently across the textbooks surveyed. Although varying language is used, the textbooks clearly state that *bias* occurs when there is systematic favoring of one side or part of the population so a sample is not representative of the population and the results from the sample do not accurately reflect the truth about the population.

Comprehension of *bias* is therefore crucial for reaching Goals 1, 5, 7, and 9 (GAISE College Report 2016) listed in Chapter 1. Students should first be able to recognize *bias* as intentional or unintentional to assess reports in the media critically. Additionally, students should know that samples that are collectedly randomly are more likely to represent the population and yield generalizable results from tests of inference.
After reviewing the textbooks, it is evident that the terms *skew* and *bias* are not taught in the same chapter because they are not linked to each other in a statistical context. *Bias* is something that can happen before data are collected or during the process, while *skew* is not caused by anything in particular, it is a feature of the distribution of the data collected. To a person very familiar with these concepts in statistics, it may not be clear why a student would have trouble providing accurate definitions and a similar case can be made for any of the words listed as lexically ambiguous. Some may believe that the terms are defined in the textbook at a level beyond the understanding beyond a typical undergraduate student. Wagler, González, and Leal (2015), however, performed a lexical analysis on 18 introductory statistics textbooks, including some of the ones examined here, and found that although the reading level of difficulty changes throughout any given textbook, overall the reading level is appropriate for undergraduate students. Therefore, the difficulty with defining *skew* and *bias*, and words like them, may lie in their lexical ambiguity attributed to students’ ideas before even entering a statistics classroom.

It is no secret that students are likely to have opinions about a course and knowledge of the material in the course prior to the first day of the semester (Gal & Ginsburg, 1994). Students may look at reviews on sites like *Rate My Professors* or ask peers who previously took the course for their opinions. In addition to the ideas they may have about the course and professor in general, students come to a course with their own understandings and intuitions of concepts taught in the course (Garfield, 1995). Some of these understandings will be correct, but more often than not, students’ notions prior to course instruction conflict with theory (Bruning, et al., 2011; Konold, 1995). Incorrect student notions are prevalent in student ideas about concepts taught in an introductory statistics course and are relatively resistant to change (Konold, 1995;
Garfield, 1995). Namely, many of the students have difficulty with vocabulary in a statistics course because definitions of SE are different from the OE definitions to which students are accustomed (Barwell, 2005; Rangecroft, 2002).

Since preconceived ideas are difficult to combat (Bruning, et. al., 2011; Konold, 1995; Garfield, 1995), prior knowledge may make it difficult for students to learn the correct meaning of lexically ambiguous words. Students are very familiar with OE definitions of skew and bias because they are accustomed to using and hearing them. Prior to instruction, Kaplan, et al. (2009) collected student definitions of SE terms that are considered lexically ambiguous. Many students in this study defined the words with an OE definition. These findings support the idea that students have preconceptions about statistical topics prior to enrolling in a statistics course. Even after the correct definition of a term is introduced, however, the students still had difficulty defining the word in the statistical context and often defined them in a colloquial context (Kaplan, et al., 2010; Richardson, et al., 2013; Whitaker 2016). This may be because students struggle to reconfigure their memory to fit the new definitions (Bruning, et al., 2011).

To delve into the lexical ambiguity associated with the words skew and bias, an exploration of the definitions of these words in OE is also necessary. According to merriam-webster.com, there are four definitions of the word skew. Although skew is not as common compared to askew, skew is defined in OE on merriam-webster.com as “to take an oblique course.” Furthermore, in a high school geometry course (ME), students are introduced to skewed lines defined as “a pair of lines neither parallel nor intersecting.” Both of these definitions are very different from the statistical definition of “not being symmetrical.” Nevertheless, the fact that students must distinguish between three uses of skew is troublesome in itself and causes skew to be classified in categories ii and iii defined by Rangecroft (2002). Bias, on the other
hand, is an idea introduced to students in middle school or even earlier at home alongside lessons on prejudice. The commonly used definition of bias is “to give a settled and often prejudiced outlook to” (merriam-webster.com). This is only one of seven definitions attributed to the word. Additionally, the subtle difference between the OE definition and the statistical definition of bias, a systematic favoring of certain outcomes in the population, adds to the lexical ambiguity from category ii (Rangecroft, 2002). It is evident that the words skew and bias on their own possess a great deal of lexical ambiguity because students must distinguish between preexisting definitions they have for the words and the new statistically correct definitions.

Skew and bias have lexical ambiguity on their own because each word has several different correct definitions in OE, SE, and ME that student must negotiate. Unfortunately, skew and bias have an additional level of lexical ambiguity as a pair because the words skew and bias can be used interchangeably in OE. Rangecroft (2002) did not include this type of vocabulary as a category of SE, but this added categorization of SE is likely the reason this pair of words and other pairs of SE are lexically ambiguous. One merriam-webster.com definition of skew seems to contribute mostly to the interchangeability of skew and bias in OE: “to distort especially from a true value or symmetrical form.” Although it seems like it falls in line more with the SE definition of not being symmetric, the problem with this definition comes from the emphasis on “distort.” An example sentence of skew in this context from merriam-webster.com is “He accused them of skewing the rules in their favor.” By using skew in the context of distorting to be in favor of something, skew automatically becomes replaceable with bias. Most likely because of that understanding of the definition, thesaurus.com lists bias as a synonym for skew. The converse, however, is not true. Skew is not listed as a synonym for bias and the example sentences on merriam-webster.com do not imply bias is caused by a distortion. Instead, bias is
referred to more as a feeling about a particular group of people, where the feeling is mostly negative. Nevertheless, the connectedness of these terms in OE probably makes it even more difficult for students to distinguish between the two words and their correct definitions in SE.

Although instructors may believe that the textbook definitions and lectures are clear enough to teach the ideas of *skew* and *bias* and other vocabulary effectively, students’ definitions from prior experience may be too strong to be overcome from instruction alone (Rangecroft, 2002). Instructors, therefore, should be aware of notions students have before course instruction and how, when coupled with lexical ambiguity of SE, they may cause a block to students achieving statistical literacy (Albert, 2003; Kaplan, Rogness & Fisher, 2014). If instructors are aware of such issues, intervention methods can be developed and implemented to aid students in achieving a mastery of statistical literacy so they can be on their way to becoming statistical thinkers (Kaplan, et al., 2014). In the next section methods for collecting students’ prior knowledge on course content and interventions to address lexical ambiguity in SE will be discussed.

2.3 Assessment and Intervention in an Introductory Statistics Classroom

The first step in combating lexical ambiguity in a statistics classroom is to inform instructors of the issue (Rangecroft, 2002; Albert, 2003). Only if instructors recognize it as an obstacle to a students’ ability to achieve statistical literacy, will they even begin to design and implement lessons to help students overcome lexical ambiguity (Garfield, 1995). As described in the previous sections, researchers found lexical ambiguity to be prevalent in statistics courses (Kaplan, et al., 2009 & 2010; Richardson, et al., 2013; Whitaker 2016). Additionally, although statistics courses may differ, students enrolled in the course tend to have similar preconceptions (Gal & Garfield, 1994). Since student understanding of statistical terminology is of concern, it is
important to gather insight in this area through assessments (Konold, 1995; Garfield & Chance, 2000). While there is a great deal of literature on misconceptions about inference processes that are typically taught in an introductory statistics course (Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007), there is not much recorded research on misconceptions of students’ definitions of terms alone. The research that exists on preconceptions in the form of student definitions does not include an expansive list of terms from the course (Kaplan, et al., 2010; Richardson, et al., 2013; Whitaker, 2016). Therefore, further exploration into what students actually know prior to instruction regarding statistical terms is imperative and will aid in creating lessons that help students correct their false notions about statistical vocabulary.

The 2016 GAISE College Report recommends using assessment to improve and evaluate student learning. Just as there are several assessments of content learning that can be administered throughout a course (Garfield & Chance, 2000), there are several methods of collecting student prior knowledge in the form of a pretest questionnaire. Kaplan, et al. (2009) and Whitaker (2016) ask students to define, provide a synonym, and/or write a sentence for the word of interest. Richardson, et al. (2013), however, claim that asking students to define word without context is not realistic and instead provides students with research article extracts in which a word was bolded then asks students to define the word in context. Both approaches for collecting student definitions of terms in SE pre-instruction are clear ways to gather open-ended information prior to instruction and allow for students to write freely without the restrictions of a multiple choice format. Furthermore, both methods are likely better than not collecting any preconceptions, if an instructor would like to create lessons that address students’ prior knowledge that potentially contradicts course content. In addition to the possibility of using students’ prior knowledge to guide lesson planning, pre-instruction definitions can be compared
to post-instruction definitions to see how student definitions change after instruction. A similar questionnaire can be administered at the end of the semester to gather this information. To specifically address the difference between statistical definitions and colloquial definitions, Kaplan, et al. (2010) ask students to give two definitions, one statistical and one colloquial, for each word at the end of the semester. Whitaker (2016) and Richardson, et al. (2013) give the same prompts that were given at the beginning of the semester, respectively. Collecting and comparing definitions of SE terms pre and post instruction are important for assessing the effectiveness of course instruction in relation to students’ understanding of SE.

To gather student conceptual understanding of statistics one can administer a multiple choice assessment. In an introductory statistics course, the most commonly used measure of conceptual growth is through the Comprehensive Assessment of Outcomes in Statistics 4 (CAOS 4). This 40-question assessment was designed by statistics education researchers to assess conceptual understanding over computation understanding (Delmas, Garfield, Ooms, & Chance, 2007). By administering the CAOS 4 pre instruction, instructors can see where students have difficulty with concepts associated with lexically ambiguous words in order to plan lessons targeting the issues. Then the CAOS 4 can be given at the end of the semester to see growth or enduring misconceptions to potentially alter lessons for the following year. Overall, the CAOS 4 is simply a different way to survey students’ knowledge on course content without the pressure of a graded exam or assignment.

After instructors are made aware of student preconceptions, they can alter the structure of the course to directly address misconceptions (Rangecroft, 2002). Since students are not blank slates, instructors should specifically aim to integrate new content with prior knowledge and directly contradict prior knowledge that does not fit with the new course content (Bruning, et al.,
Instructors should also enact an engaging classroom environment through active learning to help better students retain new content (Prince, 2004). In a statistics classroom it is recommended that instructors foster active learning to engage student in statistical thinking (2016 GAISE College Report). A research study using an action research paradigm following this recommendation was done to address the lexical ambiguity of the word *random* (Kaplan, et al., 2014). The Zebra vs Hat activity developed by the author and presented in the form of a PowerPoint, takes into account prior knowledge of students and directly addresses misconceptions about the random processes in statistics. First, students are asked to define *random* in a sentence with the word used colloquially then statistically. They are given five multiple-choice options that were created based on student responses from previous studies (Kaplan, et al., 2009; 2010). Then students are shown the correct definition for both sentences with a picture for each. A picture of three people dressed as colorful zebras on a street accompanies the colloquial meaning of random to affirm that in everyday language random means something is weird or haphazard. A picture of a hat accompanies the statistical definition to affirm that every participant has an equal chance of being selected in a random process. This enacts a Zebra vs. Hat mnemonic that students can easily store in their memory and refer back to when needing to distinguish between the two meanings. Finally, students are given a full statistical definition of the word random accompanied by other pictures of objects that are also used in equally likely scenarios, like dice and cards. This slide allows students to take what they know about random from every day use, it has to do with uncertainty, and connect it to statistics, which clarifies that uncertainty in the short run equates to a distributional structure in the long run. Post-instruction student responses from this study were compared to the previous study which had no intervention (Kaplan, et al., 2010). After the intervention fewer students gave
incorrect definitions and fewer students defined random in a colloquial context. The results from Kaplan, et al. (2014) demonstrate the importance of using students’ prior knowledge to continue to develop activities that directly address lexical ambiguity in statistical language.

Since publications about lexical ambiguity in statistics were rare prior to 2009, instructors may not be aware of the issue and therefore may not recognize the need to address it. The first step towards helping achieve a level of statistical thinking may be for instructors to be aware that students enter the course with their own definitions of SE terms that are often based on OE definitions (Albert, 2003; Rangecroft, 2002). Typically students’ prior knowledge conflicts accurate definitions that are used in the course (Bruning, et al., 2011). The next step would be to collect students’ pre-instruction definitions in whatever fashion is most efficient for the class. Instructors can then enact interventions like the Zebra vs. Hat activity to help students overcome their preexisting OE definitions of SE terms. Finally, instructors should evaluate the intervention techniques to insure they are effective. With help from instructors, students may be able to get clear understanding on the SE terminology, and then they are more likely to understand the big picture ideas of statistics and ultimately take a step closer at achieving statistical thinking.
3. METHODOLOGY

3.1 Goals of the Research Study

Understanding of basic statistical terminology is crucial for students to achieve statistical literacy and ultimately become educated statistics citizens. Additionally, the importance in the understanding of the terms skew and bias is evident in reference to the goals defined for an introductory statistics course from the 2016 GAISE College Report. Furthermore, interventions that address lexical ambiguity directly are useful in aiding students in this process. The efficacy of a High Impact, Little Time (HILT) activity, which is an intervention designed to help students understand the difference between the statistical meanings of the words skew and bias and their colloquial meanings will be assessed in this paper. The efficacy of the HILT activity will be measured by comparing definitions from students who saw the activity in class with those from students who did not see the activity in class.

3.2 Subjects and Settings

The data for this study were collected from introductory statistics students over three semesters from two different institutions: Grand Valley State University (GVSU) and the University of Georgia (UGA). Data were collected as part of the HILT-LAS NSF DUE Project Number 1504013. This study specifically addresses the goals of the HILT-LAS project with respect to the words bias and skew: (1) develop a set of research-based HILT activities for addressing issues in student learning of statistics related to language use; and (2) generate evidence of the effectiveness these activities have on student learning in statistics.
The Department of Statistics at GSVU offers approximately 55 sections of their introductory course, STA 215, each semester, with about 30 students in each section. In addition to traditional classroom meetings, students meet once a week in a computer lab. This is a service course for a variety of disciplines and covers the following topics: data collection, study design, descriptive statistics, confidence intervals and hypothesis testing for one-sample and two-sample proportions and means, correlation and simple linear regression, two-way tables and the chi square test of independence, and one-way analysis of variance. The Department of Statistics at UGA offers a similar introductory course, STAT 2000, with about 7 sections of about 250 students in each section. Similar to the course at GVSU, students meet three hours a week in a lecture classroom and one hour a week in a computer lab in groups of 28 students. The course content at UGA is identical except one-way analysis of variance is not covered. In addition, different textbooks are used at the two institutions and this leads to a different organization of the course material.

The course is a service course at both institutions. While the two institutions have different Carnegie classifications, we will none-the-less assume students at the beginning of the course at each institution have similar content knowledge. Therefore, the pre-instruction data collected from UGA during the first two weeks of the semester will be used to represent pre-instruction data for all students from both institutions.

Six instructors were recruited to be part of the HILT-LAS project. These instructors volunteered to be part of the two-year project and are called the HILT instructors. All six HILT instructors have extensive experience teaching STA 215: five are affiliate faculty member with sole expectations in teaching, principally STA 215. The other HILT Instructor is a long-term adjunct instructor of STA 215. One goal of this study is to gather evidence of the effectiveness of
HILT activities, so Comparison instructors were chosen in addition to the HILT instructors. The choice of the Comparison instructors was based on factors including teaching experience in STA 215 (e.g., the number of sections of STA 215 previously taught) and student performance (e.g., course average GPA) to have the Comparison instructors mirror the HILT instructors as much as possible. The Comparison instructors were asked to teach STA 215 as they typically would and to not use any of the HILT activities. HILT instructors implemented at least one HILT activity for a lexically ambiguous word in their class per semester. There were no HILT instructors at UGA so all instructors at UGA are Comparison instructors.

3.3 Development of HILT Activity

Each HILT instructor was responsible for developing a HILT activity based on a word the instructor considered lexically ambiguous in the context of introductory statistics. To aid in the creation of these activities, the Project Team conducted a workshop with the HILT Instructors during summer 2015. During the workshop, instructors learned about the concept of and the research behind lexical ambiguity. Then instructors saw a demonstration of the Zebra vs. Hat activity. Finally, they brainstormed other words that are used in introductory statistics that might possess lexical ambiguity.

HILT instructors created their activities in fall 2015. To allow for collaborative feedback on the development of the activities, the HILT instructors met twice a month during the fall 2015 semester with the Project Team. The meetings typically lasted 45 minutes to an hour and began with a discussion of the implementation and/or discussion in class of the Zebra vs. Hat activity and data collection activities, followed by a discussion of the student results for the instructors’ target words. By the end of the semester, instructors were sharing draft and revised activities for
target words. These newly developed activities were implemented during the spring 2016 semester by their respective developers.

One instructor chose to create an activity that would target the lexical ambiguity of *skew* and *bias*. The activity was implemented during the initial instruction of *skew* in the course topic of graphical displays of data. The first few drafts of the activity were filled with statistical content, with very little contradiction to OE definitions that leads to the lexical ambiguity of the words. Students would be shown the dictionary definition of *skew* accompanied by pictures of a chicken skewer as well as the mathematical definition of *skew*, “Nonparallel lines in space that do not intersect.” Then students would be shown the statistical definition of *skew* as it related to the graphical representation of a distribution through pictures and words. Then students would write a minute paper with the prompt “What did you learn about the word *skew* today as it applies to statistics?” The activity then leads into a discussion of *bias* in everyday language and closes with a lesson on kurtosis.

The final draft of the activity mirrored the Zebra vs. Hat activity. Students were given the prompt “That person’s point was skewed,” on the first slide of the activity. The class is then polled to select the best meaning for the word *skew* from the options: (a) A bias toward one particular group or subject; (b) Twist or turn or cause to do this; (c) The state of not being symmetrical; and (d) Sudden change in direction or position. Students were required to select a response and were only able to vote for one response. The distribution of student responses was then drawn on the board. The second slide had the prompt “The shape of the distribution (graphic or histogram) is skewed,” and again were polled to define the word *skew* with the same options. After students were polled form both prompts, the instructor acknowledges that *skew* can mean several different things, especially there are different meanings in a statistical sense
and everyday language. The final slide in the activity compared the statistical use of the word *skew* with the common use of the word *skew*, accompanied by pictures of statistical *skew* and a cartoon for common-use *skew*. The instructor explained that in statistics *skewed* means the state of not being symmetrical. He went on to say that a *skewed* distribution is “distorted in some way.” The instructor then contrasted the statistical definition with the colloquial definition of *skew: a bias* towards a group. The HILT activity closed as the instructor reminded students that the definition of the word *skew* in SE is different than the definition in OE. The activity does not address a difference between *skew* and *bias* in SE. Additionally, there is no HILT activity taught that specifically addressed the lexical ambiguity of *bias* on its own.

3.4 Instrumentation

Student responses from open-ended assessments were collected to evaluate student understanding of the words *skew* and *bias*, both pre and post instruction. Students were asked to define *skew* and *bias* after reading a prompt in a question that were asked as part of a homework or optional assignment. Several prompts were used over the course of data collection. The prompts and number of students who responded to each prompt are listed in Table 2. Allowing students to define the words instead of selecting from multiple-choice responses provided the researchers an opportunity to see into a student’s reasoning and collect student misconceptions. Categorizing these responses, however, is much more difficult than coding a multiple-choice assessment because responses are not simply correct or incorrect: there are varying levels of correct and incorrect definitions, along with misconceptions. Student responses were, therefore, categorized based on the misconception used in the definitions. The categorization process is described in Section 3.5.
The course final exam also had a multiple-choice question to see if students understood the concept of skew in a contextual situation to show a deeper level of statistical reasoning. The final exam had four questions specifically created to assess the lexically ambiguous words that had a HILT activity that were added to the bank of standard questions. Instead of asking students about the definition of the words, they were given a context and needed to use a deeper level of statistical literacy. The four HILT questions were randomly assigned to each student. A total of 203 answered the question about skew in spring 2016. The final exam question was “The registrar at a university wants to construct a graphical display showing the ages of students who are currently enrolled (where most students are in the 18 – 22 range and there are some students who are older). What general shape should the graph have?” Students selected from the following choices: (a) A strong negative relationship; (b) Left- or negatively-skewed; (c) Relatively symmetrical; and (d) Right- or positively-skewed.

Table 2: Assessment Prompts

<table>
<thead>
<tr>
<th>Word</th>
<th>Prompt</th>
<th>Pre/ Post</th>
<th>Semester</th>
<th>Number of Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Skew</td>
<td>The researcher reported that the data were skewed. What does skewed mean in this context?</td>
<td>Pre</td>
<td>Spring 2015</td>
<td>642</td>
</tr>
<tr>
<td>Skew</td>
<td>Provide a definition for the meaning of the word &quot;skew&quot; as it is used in the discipline of statistics. Also, use the word &quot;skew&quot; in a sentence to demonstrate your definition.</td>
<td>Post</td>
<td>Spring 2016</td>
<td>95</td>
</tr>
<tr>
<td>Bias</td>
<td>The researcher reported that the data were biased. What does biased mean in this context?</td>
<td>Pre</td>
<td>Spring 2015</td>
<td>541</td>
</tr>
<tr>
<td>Bias</td>
<td>Provide a definition for the meaning of the word &quot;bias&quot; as it is used in the discipline of statistics. Also, use the word &quot;bias&quot; in a sentence to demonstrate your definition.</td>
<td>Post</td>
<td>Spring 2016</td>
<td>108</td>
</tr>
<tr>
<td>Skew and Bias</td>
<td>One statistician reported that his data were &quot;skewed&quot;. Another statistician reported that her data were &quot;biased&quot;. Explain the meanings of the words &quot;skew&quot; and &quot;bias&quot; as used by the two statisticians.</td>
<td>Post</td>
<td>Spring 2016</td>
<td>85</td>
</tr>
</tbody>
</table>
3.5 Analysis

*Categorizing Student Definitions*

In order to assess the improvement in student understanding from the definitions, rubrics were created to categorize student responses from the prompts. These coding rubrics were created after the responses from the spring 2015 prompts were collected. Similar responses were grouped together, ultimately leading to categories of the responses. Some categories from the rubrics for *skew* and *bias* overlap because *skew* and *bias* are linked so closely in OE, so students tended to confuse the words in similar ways. Ultimately, two separate rubrics were designed and used since the goal of the study was to gauge misconceptions about *skew* and *bias* separately. The final rubrics for *skew* and *bias* are listed in Table 3 and Table 4 along with example student responses from each category.

After the rubrics were agreed upon, the responses were hand coded by the research team. Two researchers coded the responses independently according to the rubrics then the two researchers met to discuss the few disagreements in scoring. During this process, minor adjustments were made to the rubrics to clarify or expand some of the categories. In every case of different scoring, the researchers reached an agreement for coding.

Due to the high volume of student responses, only the pre-instruction data from spring 2015 were hand coded. Subsequent responses were coded using machine learning algorithms in R. Through this process, the hand-coded student definitions and their codes from spring 2015 acted as a training set. The training set was used to construct models that would accurately predict codes for new, un-coded student definitions. Therefore, the data from spring 2016 were entered into R and assigned the code that has the highest probability based on the models created from training set.
<table>
<thead>
<tr>
<th>Number</th>
<th>Description of Category</th>
<th>Example Response from Spring 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>(Correct STATISTICAL definition of <em>skew</em>) Responses that correctly identify <em>skewed</em> data as having a tail when graphed. Response must make coder believe the student has a correct picture in mind. To be categorized into this category it must be clear that the response refers specifically to the physical shape of the data. Otherwise a response is categorized as 3 (see below).</td>
<td>“If the data is skewed, the tail of the data is longer than the other side.”</td>
</tr>
<tr>
<td>2</td>
<td>(Correct STATISTICAL definition of <em>bias</em>) Responses that connect <em>bias</em> to problems with data, samples or results that have occurred as a result of an experimental design flaw; some examples include (but are not limited to) the wording of questions, response <em>bias</em>, and not taking a random sample. Not representative is a minimal response in this category.</td>
<td>“Not a complete ley (sic) faithful representation of the population”</td>
</tr>
<tr>
<td>3</td>
<td>(Correct EVERYDAY definition of <em>bias</em> in statistical context) Be one-sided or lean in one direction (without clarification, for example, reference to right or left; to favor one group, be unfair or based on opinion</td>
<td>“Skewed in this context means that one side of the data was favored over the other.”</td>
</tr>
<tr>
<td>4</td>
<td>(Correct EVERYDAY definition of <em>skew</em> in statistical context) Untrue, incorrect, different or having a difference, unexpected, and anything having to do with malicious intent</td>
<td>“to make bias or distort in a way that is inaccurate”</td>
</tr>
<tr>
<td>5</td>
<td>About the shape, either physical graph or by using something related to the distribution but incorrect or incomplete. For example, “not evenly distributed” or “not a normal distribution.” Responses that state facts about <em>skewed</em> distributions, such as “the mean and median are not the same.” Responses about distributions that have poor or incorrect communication.</td>
<td>“Skewed can mean that the mean is typically greater than the median or the mean is typically less than the median, depending on whether the data is skewed right or left.”</td>
</tr>
<tr>
<td>6</td>
<td>Responses that indicate outliers create a problem</td>
<td>“Skewed means that there were outliers in the data that made the results less than symmetric.”</td>
</tr>
<tr>
<td>7</td>
<td>(Correct Everyday Definition of <em>SKEW</em>): suddenly change direction or position; non-parallel.</td>
<td>“A sudden change in direction or position.”</td>
</tr>
<tr>
<td>8</td>
<td>Focus on variability. For example, having a lot of variability – “all over the place” is a naïve expression of this category, or variability causing issues</td>
<td>“Skewed means spread in this context.”</td>
</tr>
<tr>
<td>9</td>
<td>Defined using the word <em>skew</em> (with nothing to help categorize)</td>
<td>“In this context this means that the data were <em>biased</em> in one way or another.”</td>
</tr>
<tr>
<td>10</td>
<td>Define using the word <em>bias</em> (with nothing to help categorize)</td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Coding Rubric for Student Definitions of *Bias*

<table>
<thead>
<tr>
<th>Number</th>
<th>Description of Category</th>
<th>Example Response from Spring 2015</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Correct and complete statistical definition including not representative of the population</td>
<td>“The data from the sample does not accurately express what the assumed data from the population the sample represents would express.”</td>
</tr>
<tr>
<td>2</td>
<td>(incomplete statistical definition) Data are flawed due to an experimental design issues, such as the data were collected incorrectly, for example, “<em>skewed</em> due to non-random sampling” or from the way the question was worded, for example, “the question might have been asked in a leading way.”</td>
<td>“<em>Biased</em> means that it was a bad sample or I did not collect the sample correctly.”</td>
</tr>
<tr>
<td>3</td>
<td>Data or results do not represent the sample (might be confusing sample and population)</td>
<td>“<em>Biased</em> means that the information was skewed (sic) in a way that does not accurately represent the data.”</td>
</tr>
<tr>
<td>4</td>
<td>(Correct EVERYDAY definition of <em>bias</em> in statistical context) Be one-sided or lean in one direction; to favor one group or side, be unfair (without reference to an experimental design issue)</td>
<td>“<em>Biased</em> means that the data was more favorable to one side of the experiment or argument.”</td>
</tr>
<tr>
<td>5</td>
<td>Manipulated to match the researcher’s intent, pre-conceived notion or purpose, clear malicious intent on the part of the researcher</td>
<td>“Having a <em>biased</em> set of data means that the researcher views a certain outcome over another, hence the data is swayed in the researcher's own view.”</td>
</tr>
<tr>
<td>6</td>
<td>(Correct statistical definition of <em>skew</em>) When graphed the data are not symmetric, they have a tail to one side.</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>(Incorrect EVERYDAY definition of <em>bias</em> in statistical context) Untrue, incorrect, different or having a difference, unexpected, effected by an outside factor (not clearly manipulated by the researcher) also no statistical context</td>
<td>“<em>Biased</em> means that the data reported has been influenced by other factors.”</td>
</tr>
<tr>
<td>8</td>
<td>data are subjective, not objective, based on subjects’ opinions</td>
<td>“The data was not based on facts only but also personal opinion.”</td>
</tr>
<tr>
<td>9</td>
<td>Defined using the word <em>bias</em></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Defined using the word <em>skew</em> without enough information to classify elsewhere.</td>
<td>“<em>Biased</em> means that the data is skewed.”</td>
</tr>
<tr>
<td>99</td>
<td>Off topic answer</td>
<td>“<em>Biased</em> means that they are connected (sic) to each other.”</td>
</tr>
</tbody>
</table>
The distribution of definitions for the words bias and skew given after instruction by students who did and did not experience the HILT activity for skew, were compared to each other and to the distribution of definitions given by students prior to statistics instruction. Descriptive statistics of these distributions are discussed in Chapter 4.

Comparing Multiple Choice Final Exam Responses

Final exam responses were marked as correct or incorrect. A Chi- Square test of homogeneity was performed using R to test the null hypothesis that the proportion of students who answered the question correctly is the same for both HILT students and Comparison students. The test was performed at a significance level of $\alpha = 0.05$. 
4. RESULTS

4.1 Comparing Student Definitions of *Skew*

The second goal of the HILT-LAS project is to generate evidence of the effectiveness of the HILT activities on student learning in statistics. The differences between categories of collected student definitions of the word *skew* students who were exposed to the HILT activity and those who were not exposed to the HILT activity are presented in this section. Additionally, pre-instruction definitions are compared to both groups in post-instruction. A Chi-Square test of test significant was not performed on these data because the conditions for the test were not met.

The percent of student definitions of *skew* that were in each category as defined by the rubric in Section 3.5 are listed in Table 5. Figure 1 is a visual representation of this table, with categories with small percentages removed. Overall, the percent of students who defined *skew* with an OE definition in a statistical context was lower post instruction, but 2% of HILT students (n = 2) compared to 0% of Comparison students defined *skew* in a completely colloquial context. Additionally, overall the percent of students who correctly defined *skew* in a statistical context increased post instruction. Although the 59% of HILT students (n = 54) with a correct statistical definition of *skew* was lower than the 64% of Comparison students (n = 56) in that category, 14% of HILT students (n = 13) used a colloquial definition in a statistical context compared to 17% of Comparison students (n = 15). With regard to the lexical ambiguity of *skew* and *bias* as a pair, no students post instruction wrote a definition of statistical *bias*, nor defined *skew* as *bias*. Seven percent of HILT students (n = 6), however, defined *skew* with an everyday definition of *bias*. 
Table 5: Definitions for *Skew* Given by Students

<table>
<thead>
<tr>
<th>Definition Category</th>
<th>Pre-Test</th>
<th>Comparison</th>
<th>HILT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 642)</td>
<td>(n = 88)</td>
<td>(n = 92)</td>
</tr>
<tr>
<td>1: Complete Statistical</td>
<td>33%</td>
<td>64%</td>
<td>59%</td>
</tr>
<tr>
<td>2: Statistical <em>Bias</em></td>
<td>3%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>3: Everyday Definition of <em>Bias</em></td>
<td>7%</td>
<td>1%</td>
<td>7%</td>
</tr>
<tr>
<td>4: Everyday Definition with Statistical Context</td>
<td>41%</td>
<td>17%</td>
<td>14%</td>
</tr>
<tr>
<td>5: Incomplete About Shape</td>
<td>8%</td>
<td>9%</td>
<td>12%</td>
</tr>
<tr>
<td>6: Outliers</td>
<td>5%</td>
<td>9%</td>
<td>7%</td>
</tr>
<tr>
<td>7: Everyday Definition with no Statistical Context</td>
<td>0%</td>
<td>0%</td>
<td>2%</td>
</tr>
<tr>
<td>8: Variability</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>10: Defined with <em>Bias</em></td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Proportions of Each Response Type Observed by Group

![Proportions of Each Response Type Observed by Group](image)

Figure 1: Definitions for *Skew* Given by Students
4.2 Comparing Student Definitions of *Bias*

This section describes how categories of collected student definitions of the word *skew* differed between students who were exposed to the HILT activity and those who were not exposed to the HILT activity. Recall, however, that the HILT activity did not directly address the lexical ambiguity of *bias* on its own and briefly discussed the lexical ambiguity of *skew* and *bias* as a pair. Additionally, pre-instruction definitions are compared to both groups in post-instruction. A Chi-Square test of test significant was not performed on these data because the conditions for the test were not met.

The percent of student definitions of *bias* that were in each category as defined by the rubric in Section 3.5 are listed in Table 6. Figure 2 is a visual representation of this table, with categories with small percentages removed. Overall, the percent of students who defined *bias* with an OE definition in a statistical context was lower post instruction, and no students defined *bias* in a completely colloquial context. Additionally, overall the percent of students who correctly defined *bias* in a statistical context increased post instruction. Although the 25% of HILT students (n = 24) with a correct statistical definition of *bias* was lower compared to 32% of Comparison students (n = 31), 17% of HILT students (n = 16) used a colloquial definition in a statistical context compared to 24% of Comparison students (n = 24). Additionally 13% of HILT students (n = 12) defined *bias* with an incorrect everyday definition in a statistical context, compared to 15% of Comparison students (n = 15). With regard to the lexical ambiguity of *skew* and *bias* as a pair, no students post instruction wrote a definition of statistical *skew*, defined *bias* with an everyday definition of *skew*, and no HILT students defined *bias* as “*skew,*” while two percent of Comparison students (n = 2) defined *bias* as “*skew.*”
Table 6: Definitions for Bias Given by Students

<table>
<thead>
<tr>
<th>Definition Category</th>
<th>Pre-Test (n = 541)</th>
<th>Comparison (n = 98)</th>
<th>HILT (n = 95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1: Complete Statistical</td>
<td>11%</td>
<td>32%</td>
<td>25%</td>
</tr>
<tr>
<td>2: Flawed Data Collection</td>
<td>9%</td>
<td>11%</td>
<td>29%</td>
</tr>
<tr>
<td>3: Does not Represent Sample</td>
<td>2%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td>4: Everyday Definition in Statistical Context</td>
<td>29%</td>
<td>24%</td>
<td>17%</td>
</tr>
<tr>
<td>5: Manipulated Results</td>
<td>19%</td>
<td>10%</td>
<td>11%</td>
</tr>
<tr>
<td>7: Untrue, Incorrect</td>
<td>18%</td>
<td>15%</td>
<td>13%</td>
</tr>
<tr>
<td>8: Subjective Data</td>
<td>9%</td>
<td>5%</td>
<td>5%</td>
</tr>
<tr>
<td>10: Defined with Skew</td>
<td>3%</td>
<td>2%</td>
<td>0%</td>
</tr>
<tr>
<td>99: Off Topic</td>
<td>1%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>

Figure 2: Definitions of Bias Given by Students
4.3 Comparing Final Exam Question

The results of the final exam question that assessed students’ statistical reasoning of skew in the context of the expected shape of the distribution of ages of university students are presented in this section. Exam results for the particular question were compared across students who were exposed to the HILT activity and those who were not exposed to the HILT activity. There are no pre-instruction data to compare for this question.

The distribution of responses across groups of students is list in Table 7. About half of all students answered this question correctly. The distribution of correct and incorrect responses across groups is listed in Table 8. The result of the Chi-Squared test of homogeneity to test for an association between whether a student saw the HILT activity or not and whether they answered the final exam question about skew correct or incorrect yielded a test statistic of $\chi^2 = 0.19$ and a p-value of 0.66. There is, therefore, no evidence of a relationship between whether a student was exposed to the HILT activity and whether they answered the question correctly. Furthermore, no HILT students answered that the shape of one variable had a negative relationship, while 2% of Comparison students (n = 2) selected that answer. Forty percent of HILT students (n = 34) responded that the shape would be symmetric, while 33% of Comparison students selected that answer (n = 39).
Table 7: Final Exam Responses

<table>
<thead>
<tr>
<th></th>
<th>(a) A strong negative relationship</th>
<th>(b) Left- or negatively-skewed</th>
<th>(c) Relatively symmetrical</th>
<th>(d) Right- or positively-skewed (Correct)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison (n = 118)</td>
<td>2%</td>
<td>15%</td>
<td>33%</td>
<td>50%</td>
</tr>
<tr>
<td>HILT (n = 85)</td>
<td>0%</td>
<td>14%</td>
<td>40%</td>
<td>46%</td>
</tr>
</tbody>
</table>

Table 8: Final Exam Analysis

<table>
<thead>
<tr>
<th></th>
<th>Correct</th>
<th>Incorrect</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comparison (n=118)</td>
<td>50%</td>
<td>50%</td>
</tr>
<tr>
<td>HILT (n = 85)</td>
<td>46%</td>
<td>54%</td>
</tr>
</tbody>
</table>
5. DISCUSSION

5.1 Summary of Results

The goal of the research study was to assess the efficacy of the HILT activity for *skew* and *bias* by comparing definitions from students who saw the activity in class with those from students who did not see the activity in class. Overall, the results indicate that the HILT activity did not help students to provide more complete statistical definitions of the words. While the percent of students who defined *skew* with an OE definition in a statistical context was lower post instruction, a higher percentage of HILT students defined *skew* in a completely colloquial context. Furthermore, the HILT activity did not address *bias* at all, so the results from *bias* probably cannot be attributed to the intervention. Therefore, the HILT activity for *skew* should be updated to better address misconceptions and a HILT activity should be created for *bias*.

5.2 Limitations

This study did not randomly assign students to treatments. This is an observational study, in which all HILT students came from the same instructors. This is a limitation of the study because the observed difference cannot be solely attributed to the HILT activity. Additionally, the HILT instructors volunteered to be a part of the study and possibly have motivation for improving statistics instruction in their classrooms. The results are therefore confounded by instructor.

This study did not use the same prompt across semesters to assess the student understanding of the target words. In fact, an additional prompt was used in fall 2015: “Provide a synonym or definition for the meaning of the word *skew/bias* as it is used in statistics.” Data
collected using this prompt were unusable in the analysis to compare with the prompts from other semesters. When coding this prompt for both words, it was found that about half of the students provided a complete definition, while the other half provided a synonym without context. For example, if a student wrote "partial" as a synonym for *bias*, it is not clear if they mean the sample was partial, the statistic was partial, the response was partial, the data was partial, etc. It was therefore impossible to categorize their responses to compare to responses from prompts other semesters. While collecting data that would represent the population across universities was attempted, the inconsistency of prompts did not allow this to happen. Additionally, the results are confounded by prompt.

Finally, the subjects in this study are not independent. Post-instruction students come from the same university, use the same textbook, and follow the same course content. Therefore, these results might not be generalizable to other populations, even those from similar universities.

5.3 Implication of the Results for Teaching

The results of the analysis of student definitions indicate that students still have difficulty providing a complete correct statistical definition of *skew* and *bias* post instruction, even with exposure to the HILT activity intervention. HILT students, however, were less likely to provide an OE definition in a statistical context. This could mean that the HILT activity did well at distinguishing between OE and SE, but there was not enough emphasis on the correct statistical definition, or the concepts of *skew* and *bias* are too difficult for students to define. It is important therefore, that instructors are aware of the lexical ambiguity of *skew* and *bias* and continue to develop interventions to address it. A recommendation for updating this HILT activity to help
students better combat the lexical ambiguity of *skew* and *bias* will be discussed in the next section.

Students were generally able to specify that *skew* referred to the shape of the distribution of the data and did not confuse it with as OE definition of *bias*. Although some students did not give a complete definition, they stated correct, but incomplete characteristics of *skewed* data. For example, one student wrote, “*Skewed* meaning the data was not normal or in the middle.” While it is true that *skewed* data are not normal, in order to be a complete answer the student should have specified that the data are not symmetric and have a tail. To help students address this problem, instructors must be very clear and specific about the language they use. Additionally, students should be exposed to other distributions, so they are aware that the normal distribution is not the only symmetric distribution.

Additionally, fewer students were able to provide a complete, correct statistical definition of *bias*. Incomplete definitions can be categorized into two groups: students who wrote what causes *bias* and students who wrote that *bias* causes incorrect results. The former can be further divided into two groups: those who discussed sampling methods and those who discussed opinions of researchers or participants. This incomplete categorization of responses is likely due to how and where *bias* is taught in the course. Since *bias* appears exclusively in the chapter about sampling methods, students link the two concepts together, and therefore miss the true definition of *bias*. While it is important that students recognize the various causes of *bias*, instructors should take time to make the implications of *bias* clear. If the goal is for students to assess data critically in the real world, students need not only to recognize when a sample is not *biased*, but also should know that the results cannot be generalized to the population. Students who fell into the second category of incomplete definitions almost made this distinction; however, they
instead conclude that biased data are inaccurate or wrong. Another distinction instructors should make is that neither biased data, nor results from biased data, are incorrect, instead biased data are not representative of the population and results cannot be generalized to the population.

Another concern across all student definitions was the lack of specificity. Students may be on the right track with a definition, but often wrote too generally and did not completely explain what they were trying to say. For example one student wrote, “Bias means over estimating something or under estimating something because it is what the researcher is looking for,” when defining bias. It is not clear what the student means by “something.” The student could be referring to a population parameter, which would be a correct definition. On the other hand, they could be referring to sample statistic, which could indicate more trouble. Notice also that this student attributes the problem to the colloquial bias of the researcher. Another student wrote, “Skewed refers to data leaning toward one side rather than another,” to define skew. Again, the student is not clear what means by “leaning”. Is the student referring to the distribution of the data or are is “leaning” used as a synonym for bias? In general, since the word “lean” can be synonymous for skew or bias, it should be removed from the language of a statistics classroom. Writing and language are important for communicating and understanding statistics. The introductory courses in this study do not have extensive writing assignments, nor many open-ended questions on assignments, so students might not know the best way to write specifically about statistical concepts. Instructors should integrate more writing assignments into the curriculum to help students develop the skills they need to clearly communicate their thoughts.

The results from the final exam question indicate that the HILT activity did not help students better identify skew in a context. From teaching experience, students do not have
difficulty identifying skew, or the direction of skew when given a graphical representation of data. Students have difficulty not only identifying skew, but also the direction of skew, when given the written description of a variable, as evident in the results of the final exam question. To address this problem, instructors should provide more examples of skewed data in both graphical displays and written descriptions. Instructors can also help students discover general rules about skewed distributions (e.g. variables that are bounded are likely skewed). Additionally, instructors should allow students to think about the distribution of a variable before telling them what the shape will be. If students are given an opportunity to think about the variable they may be able to build skills to better identify skew from a written description.

5.4 Directions for Future Research: Updating the HILT Activity & Collecting Student Definitions

The HILT activity as it is now, attempts to provide the difference between the SE definition and OE definition of skew. The results of this study indicate that the HILT activity did not help students to provide a more complete SE definition of skew. To better address the needs of the students, the HILT activity should warn against using the word lean, a commonly used synonym of skew. One response choice in the HILT activity should state “leaning a certain way.” This way, the instructor can address the lack of specificity in that definition directly since many students used lean in their definition of skew in this study. Students should then see that lean is not a good word to use to describe the shape of a distribution, as it is lexically ambiguous itself. Furthermore, the instructor should not say that statistical skew means the data are “distorted in some way” in the summary of the activity. Again, the word distort is lexically ambiguous. The instructor should use the definition as he does, then further elaborate that the data are shifted, which is the explanation in the textbook. By making these changes to the HILT activity, student
misconceptions and lack of specific language is immediately targeted. Students, therefore, should be more likely to refrain from using such language and provide a complete definition.

Furthermore, the HILT activity does not address bias whatsoever. In fact, it had the potential to create a greater link between skew and bias in a statistics course, since the correct colloquial definition was “A bias toward one particular group or subject.” This might explain why 7% of HILT students (n = 6) defined skew in the everyday context of bias compared to 1% of Comparison students (n = 1). The instructor, therefore, might consider adjusting that response. More importantly, a HILT activity should be made exclusively for the word bias. Since many student defined bias by listing ways bias occurs, the activity should focus more on the implication of bias. A recommendation for a HILT activity is “Data are what the Data are.” In this activity, short descriptions of several different versions of the same research study are presented to the class. The description includes the way data were collected and statistics that were calculated. Each description has the same research question and overall same method of collection but each has varying levels of implicit bias, including malicious intent of the researcher and respondents favoring one of the results, as students wrote in the definitions. The instructor then asks students which study has the correct data; students are given a short time to discuss in groups. Groups may be hesitant to answer because they may have the misconception that all of the studies having bias means the data are incorrect for all studies. On the other hand, students may not have that misconception and decide to identify the various forms of bias. The instructor then discusses various forms of bias in the different studies. This addresses misconceptions that bias is often malicious or on purpose or only comes from the selection of the sample because the descriptions will have various forms of bias. Additionally this is an opportunity to contradict the OE definition by clarifying that a “favoring bias” could then cause
statistical bias. The instructor then concludes that the data are not wrong, the “data are what the data are.” This explicitly addresses the misconception that data are inaccurate if data are biased. The instructor then selects one research description and states several different conclusions about the population parameter and asks students which is correct. Some responses might be “this study will accurately predict the population parameter”; “this study will over/under estimate population parameter”; “we do not know anything about the population parameter from this study.” After students are given time to discuss and answer, the instructor concludes that the study will over/under estimate the population parameter; then summarizes that there are many forms of bias (unintentional or intentional), “the data are what the data are,” when data are biased, we can still calculate statistics but they will not represent the true population parameter, they will be an over/under estimate. This activity addresses the most common misconceptions in the student definitions from this study, which ideally would mean more students would be able to provide a complete statistical definition of the word bias.

When an instructor decides to implement an intervention, such as one of the HILT activities, it is ideal to collect pre- and post-instruction student definitions and compare to groups without the intervention to measure the effectiveness of the activity. This study used open-ended prompts for this assessment, which provides the best opportunity for students to write freely about what they know. In future studies however, it is recommended that the same prompt be used throughout all levels of data collection. Furthermore, it is not useful to ask students to provide a synonym of the word, as it will be without context and make it difficult to categorize. Therefore, in order to get definitions that are specific and easy to code on the same rubric, the prompt for these words should be specific such as: “The researcher said data were skewed/biased. What is the statistical definition of skewed/biased in this sentence?” By using this prompt, all
students are given an example to consider and definitions will clearly talking about the data whether they explicitly state it or not. A complete response to this question demonstrates their statistical literacy of the words. Finally, with a different prompt and a different student population, future researchers might need to update the coding rubric based on the same procedure described in Section 3.5.

It is clear from this study, that even at the end of the semester with an intervention, just over one quarter of students were able to define skew properly in SE and about half of students were able to define bias properly in SE. By collecting student definitions, however, researchers should be able to create activities that address common misconceptions that students have. Overall, it is important for researchers and educators to continue to study lexical ambiguity in statistics and how an intervention can be updated or created to help students better define terms in SE.
6. REFERENCES


Pew Research Center, July, 2016, “The modern news consumer”


