2020-01-23

The Use of Story-Based Tasks in Post-Secondary Students' Learning of Statistics

Lemieux, Collette

UNIVERSITY OF CALGARY

The Use of Story-Based Tasks in Post-Secondary Students' Learning of Statistics

by

Collette Lemieux

A THESIS

SUBMITTED TO THE FACULTY OF GRADUATE STUDIES

IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE

DEGREE OF DOCTOR OF PHILOSOPHY

GRADUATE PROGRAM IN EDUCATIONAL RESEARCH

CALGARY, ALBERTA

JANUARY, 2020

Abstract

The purpose of this study is to investigate the impact of an intervention, which uses stories to explore statistics, on post-secondary students' understanding of statistics and their beliefs about the usefulness of statistics, and what features of the stories support meaningful learning.

A qualitative case study approach is used. In line with the case study approach, multiple data sources are used, which consist of student and instructor class artefacts, pre- and post-intervention written response items, and post-intervention interviews. The participants in the study are 20 students from a single first-year post-secondary business statistics course in which the intervention is implemented. Data analysis entails a thematic approach based primarily on open-coding to identify participants' understanding of statistics, their beliefs about the usefulness of statistics, and what features of the intervention supported meaningful learning.

The findings suggest that the intervention supported participants development of various types of understanding of selected topics in statistics, development of understanding of the usefulness of statistics, and personalization knowledge as part of the process of developing understanding. Further, the findings indicate that the intervention served to support positive beliefs about the usefulness of statistics. Finally, the findings suggest that the features of the intervention and, in particular, the stories that impact meaningful learning include the prompts embedded within the stories, the authentic real-world context presented in the stories, and the nature of the characters introduced in the stories.

The study contributes to the field by providing an example of an innovative intervention that supports students' learning of statistics and positive beliefs about the usefulness of statistics.

Preface

This thesis is original, unpublished, independent work by the author, C. Lemieux. The data reported in Chapters 3-4 were covered by Ethics Certificate number REB16-1399, issued by the University of Calgary Conjoint Faculties Research Board for the project “The use of story-based tasks in post-secondary students’ learning of statistics” on December 22, 2016 and by the appropriate ethics board for the university where the study was conducted on January 18, 2017 (ethics certificate number 100789).

Acknowledgements

I would like to express gratitude and appreciation to all the people who supported me in the completion of this thesis. Though there is not enough room to thank everyone, I know that getting to this point was only possible due to the community of support that I am so lucky to have.

Without the tireless support of my supervisor, Dr. Olive Chapman, this thesis would never have been completed. She helped me beyond measure and I am grateful for it.

I would also like to thank my committee and examiners. Dr. Armando Preciado Babb and Dr. Kim Koh provided exceptional feedback at all stages of this process. Dr. Claude Laflamme and Dr. Cynthia Nicol made wonderful suggestions that have improved my thesis greatly. The wisdom and thoughtfulness of all four are much appreciated.

My colleagues Brad Quiring, Martin Haworth, Dr. Indy Lagu and Dr. Egan Chernoff provided support and advice throughout this process.

To the faculty member who let me bring the intervention into his classroom and helped me in the writing the stories: I am very grateful for all of the support. The stories have life because of you.

Finally, I could not have done this without my family. I have no words to explain how much my husband, Andrew Taylor, helped and supported me. All I can say is that I am lucky to have him. My dog, Inara, never failed to provide cuddles and long walks when I needed a break or a space to think. My parents, Donna and Gerard Lemieux, have supported me endlessly throughout my education and I wouldn't be here without them.

Table of Contents

List of Tables

Chapter 1 - Introduction

This qualitative single-case study seeks to understand how a story-based teaching approach impacts post-secondary students' learning of statistics, their beliefs about the usefulness of statistics, and what features of the approach support meaningful learning for the participants. This chapter begins with the research context regarding the current state of statistics education. This is followed by the statement of the problem, the purpose, the key theoretical constructs, and the significance of the study. It ends with the researcher's background and organization of the thesis.

1. Research Context

Statistics permeates society. One would be hard pressed to find a news article or an argument on major topics of our time, such as climate change, gun control or terrorism, without being presented with some of kind of statistical information. As Konold and Higgins (2003) stated:

> At the practical level, knowledge of statistics is a fundamental tool in many careers, and without an understanding of how samples are taken and how data are analyzed and communicated, one cannot effectively participate in most of today's important political debates about the environment, health care, quality of education, and equity. (p. 193)

Not only are statistics being used to overtly persuade, they are also being used covertly. This is highlighted by the Cambridge Analytica scandal of 2018, which revealed how the company used big data collected through social media to influence elections around the world. These examples and others highlight why it has become increasingly important for citizens to be empowered with statistical literacy in today's information laden society (Gal, 2002).

Yet having statistical literacy is more than simply being able to compute statistical measures following pre-ordained recipes (Cobb, G. W., 1992; Gal, 2002). Thus, many statistics educators have advocated for reform in how statistics is taught and this movement is often called the *reform movement in statistics education*. In particular, the reform movement calls on statistics educators to focus on promoting statistical knowledge, reasoning and thinking (e.g., Ben-Zvi & Garfield, 2004; Cobb, G.W., 1992; Moore, 1997). That is, in addition to producing various statistical measures, statistically educated students should be able to interpret statistical results, understand the reasoning behind statistical concepts, understand the process of statistical investigations, know the limits of statistics, and take a critical stance when presented with the results of a statistical analysis (Everson et al., 2016; Gal, 2002; Gal & Garfield, 1997).

In addition to the cognitive skills required to understand statistics, statistics educators also recognize that belief in the usefulness of statistics in one's life and career is an important outcome for students in a statistics course (Gal, 2002; Ramirez, Schau & Emmioglu, 2012). If students leave a statistics course without seeing the usefulness of the subject, it is unlikely that they will use what they have learned (Schau & Emmioglu, 2012).

There has been significant research done on students' understanding of various concepts in statistics and probability (e.g., Cooper & Shore, 2008; Ireland & Watson, 2009) involving K-12, post-secondary and graduate students. Though these studies cover various concepts, a commonality is that students struggle with statistical concepts. For example, Mathews and Clark (2007) examined students' understanding of mean, standard deviation, and sampling distributions for post-secondary students who recently earned an A in their first-year statistics course. They found that the students confused mean with the mode and proportion; believed that the standard deviation is found by determining the distance between the data values in the sample or is the distance between the mean and one data value in the sample; and had no coherent understanding of the central limit theorem. That is, students who would be deemed to have very successfully completed a post-secondary statistics course still did not have a strong understanding of basic statistical concepts. As another example, a study conducted from 2005 to 2011, that included over 13,000 post-secondary students, found that there was no significant change in students' understanding of statistical concepts over this period (Garfield, delMas, & Zieffler, 2012). This suggests that, though the reform movement has attempted to provide best practices for the teaching and learning of statistics, there continues to be issues in how to effectively implement these practices to help students improve their understanding of statistical concepts and statistics as a discipline.

The difficulties found in students' understanding of statistical concepts also exists around students' beliefs about the usefulness of statistics. Many students begin a statistics course with a neutral view of the usefulness of statistics (Neumann & Hood, 2009; Hood, 2013; Schau & Emmioglu, 2012; Tsao, 2006). Yet beliefs about the usefulness of statistics do not necessarily improve after taking a statistics course (Carnell, 2008; D'Andrea & Waters, 2002; Gordon, 2005; Murtonen, Olkinuora, Tynjala, & Lehtinen, 2008; Schau & Emmioglu, 2012). This includes courses that are taught using innovative teaching practices (Carnell, 2008; D'Andrea & Waters, 2002). This suggests that there is still opportunity for improvement in teaching statistics in a way that helps students believe that statistics is useful in their lives.

2. Statement of Problem

Over a quarter of century ago, G.W. Cobb (1992) called for changes in how statistics was taught. More than a decade ago, the first GAISE (Guidelines for Assessment and Instruction in Statistics Education) report was published to provide guidelines on best practices in statistics education (Garfield et al., 2005). Yet there continues to be difficulties in successfully converting these guidelines to practice (Tishkovskaya & Lancaster, 2012) and many innovations in statistics education are based on intuition rather than on research (Ramirez et al., 2012). Even in classrooms that incorporate reforms, students still struggle with understanding various concepts in statistics (Chance, delMas, & Garfield, 2004) and have difficulty seeing the usefulness of statistics in their lives. Thus, though there have been changes in how statistics has been taught, there is still a need for research on how to effectively support students' understanding of statistical concepts and promote positive beliefs about the usefulness of statistics using suggestions of best practices from the reform movement.

3. Purpose of Study

The purpose of this study is to investigate the effects of an innovative teaching approach (the intervention) on students' understanding of statistics and their beliefs about the usefulness of statistics in their lives. The intervention consists of four story-based tasks and their supplemental tasks (reflection tasks and follow-up tasks), which provide authentic contexts for students to explore major topics in their first-year post-secondary business statistics course. This study investigated three research questions:

1) In what ways does the intervention impact post-secondary students' understanding of selected topics and the discipline of statistics?
2) In what ways does the intervention impact post-secondary students' beliefs about the usefulness of statistics in their everyday lives?
3) What features of the intervention support meaningful learning for the students?

Regarding questions 1 and 2, the 'intervention impact' refers to what the students' understanding and beliefs look like when they learn statistics through the intervention. Thus, this study is not trying to determine cause and effect, but is rather attempting to determine what happens when students engage in the intervention.

4. Theoretical Constructs

The key theoretical constructs that frame the study are the nature of statistics, the learning theory of constructivism, understanding, the nature of stories, and beliefs. They are briefly introduced in this chapter and are further discussed in Chapter 2.

Nature of statistics

Statistics as a discipline focuses on data, variation, and chance, and uses context to provide meaning to the analysis and interpretation (Cobb & Moore, 1997; Gal, 2002; Wild & Pfannkuch, 1999). More generally, it involves of statistical knowledge, statistical reasoning and statistical thinking, which are promoted as what should be the focus of statistics courses (e.g., Ben-Zvi & Garfield, 2004; Garfield & Chance, 2000). In this study, the focus is to investigate the impact of the intervention on the participants' understanding of specific aspects of statistical knowledge, reasoning and thinking (described later).

Learning theory

The primary learning theory related to this study is constructivism. In particular, this study uses the emergent perspective of constructivism outlined by Paul Cobb (1994). From this perspective students learn by drawing on their prior knowledge, constructing their own understanding of concepts, collaborating with peers, negotiate meaning with others, and being introduced to the cultural practices of the discipline. The story-based intervention, framed in this perspective of constructivism, provides students with opportunities to construct their own understanding of the statistics concepts.

Understanding

Within this study, understanding is defined using Skemp's (1976/1978) framework of instrumental and relational understanding. Skemp defined instrumental understanding as the ability to use an algorithmic procedure to solve a problem, but without knowing the reasons for the steps or "rules without reasons" (p. 9). Relational understanding, on the other hand, is "knowing both what do and why" (Skemp, 1976/1978, p. 9), which includes how the concept relates to other concepts and how it can be applied. This definition of understanding is widely used in mathematics education and, in this study, provided the basis to determine the types of understanding participants developed about the selected topics and the discipline of statistics as a result of the intervention.

Nature of stories

In this study, stories are defined as narratives that have a clear beginning and end, and tell a sequence of events with a character that is driving the events towards a solution to a problem or conflict (Egan, 1986). They have characters, plots, context, conflict, imagery, emotions, and humour (Carter, 1993; Roberts & Stylianides, 2013; Zazkis & Liljedahl, 2009). This interpretation of story guided the development of the story-based tasks for the intervention. Further, stories are an integral part of human culture. We are storytellers, and through the act of storytelling we connect our learning to our own lives (Clark & Rossiter, 2008). Storytelling could allow students to connect a statistical skill to "a particular human hope, intention, fear, or whatever, then [they] can embed the skill in a context that is meaningful" (Egan, 1986, p. 77). Thus, stories provide the opportunity to ground the abstract statistical concepts in a meaningful context, which connects the concepts to students' lives and future professions.

Beliefs

The works of Green (1971), Thompson (1992), and Pajares (1992) primarily informed the construct of beliefs for this study. Beliefs form one's understanding of the world and reality (Beebe, Beebe, Redmond, & Geerinck, 2011) and include one's ideas and opinions; perspectives; and truths about oneself, a domain or a social-context (Gal, 2002; Pajares, 1992). Beliefs can be held at varying levels of conviction (Green, 1971; Thompson, 1992) that affect changes to them. This perspective of beliefs informed data collection and analysis of the participants' beliefs about the usefulness of statistics resulting from the intervention.

5. Significance of Study

There is very little research on the use of stories or story-based tasks to teach statistics (Smith, 2014). However, as previously noted, stories could provide students with an innovative and engaging way to explore and meaningfully learn statistical concepts. Thus, this study has the potential to make an important contribution to the field by adding to our understanding of the ways in which the nature and use of stories could impact post-secondary students' learning and beliefs about statistics.

As the intervention being studied was designed under the guidance of the best practices for statistics education as suggested by the reform movement, this study also contributes to our understanding of how a teaching approach designed using these guidelines impacts students' understanding of statistical topics and the discipline of statistics. In addition, a primary focus in statistics education research is on the achievement outcomes of students, yet beliefs about statistics that play a role in that achievement has been researched far less (Ramirez et al., 2012). This study provides information on students' belief of the usefulness of statistics that could impact their achievement.

Finally, the results of the study will also be useful to curriculum designers and instructors who are proposing or considering the use of stories to teach statistics. In particular, it will provide them with a better understanding of the opportunities and limitations in using stories to teach statistics.

6. Background of Researcher

During my undergraduate degree in mathematics, the most common teaching style I experienced could be defined as 'chalk and talk' or, more accurately, 'chalk and talk at'. That is, my professors would write equations and proofs on the board and tell us what they were doing with little or no interaction with the students. In 2002, after completing my master's degree in mathematics, I began teaching mathematics and statistics at the post-secondary level. It quickly became clear that the teaching method of chalk and talk was not effective for my students. Instead, they wanted time to try problems in class, talk to their peers, and ask questions. They also wanted to know how what they were learning would be useful in their future careers and lives.

At the time, I had taken no education courses and was only vaguely aware of things like learning theories. All of this changed when, in 2010, I began my education degree at the University of Calgary. Through that degree and later my PhD course-work, I was introduced to learning theories, various pedagogical strategies, assessment theories, etc. By gaining this knowledge, I have come to better understand my own beliefs about learning and teaching. In particular, I believe that learning is constructivist in nature. We each actively construct our own understanding of the world by drawing on our past experiences and by interacting with others (especially knowledgeable others; Cobb & Yackel, 1996). Further, as part of actively constructing knowledge, it is important to have the opportunity to externally communicate our understanding (e.g., talking with others or writing it down) to make our learning visible, which allows us to check the coherence, consistency and completeness of our understanding both with ourselves and with others (Rossiter & Clark, 2007). With this better understanding of learning, I adjusted my teaching practices to align with them. For example, prior to my education degree, I encouraged students to collaborate during class but did so in a haphazard manner. Using what I had learned, I started to engage in more intentional practices such as using activities like think, pair, share, which allowed students to have a more well-defined way to actively construct their knowledge through engaging with others.

Even with the changes in my practice, I still felt that something was missing. In particular, for my statistics courses, though my students were actively engaging more with the content and were being assessed in multiple ways, I still felt that the course was focusing too much on calculations in isolation. That is, students were well-versed at the end of the course in producing statistical measures such as the standard deviation and confidence intervals for the mean, but if asked what the results meant either in general or in the context of a problem, they had great difficulty answering.

While taking my PhD courses, thoughts about how to address this difficulty were swirling in my head. While taking the course 'Adult Learning Theories', I was introduced to narrative learning. To put it simply, narrative learning is learning through stories, which includes both hearing, telling and recognizing stories (Clark & Rossiter, 2008). As I read more about narrative learning, I became more and more intrigued about the idea of using stories in my statistics courses. The idea of using stories to teach statistics spoke to me for multiple reasons. One reason is that learning through stories fit with my beliefs about how we learn. In particular, learning through stories provides the opportunity for students to draw from their past experience to understand the story (Clark & Rossiter, 2008). Additionally, if they are given the opportunity to actively engage with the story by discussing it, writing responses to it, etc., students have the opportunity to make explicit their understanding of the concepts they are learning through the story. Thus, the idea of learning through stories aligned with my beliefs about students actively engaging in the construction of their knowledge. Another reason why stories spoke to me was that an important aspect of statistics is telling the story of the data (Pfannkuch, Regan, Wild, & Horton, 2010). That is, students not only need to produce statistical measures, but they also need to provide meaning to the results by understanding them within a context. Thus, I saw the potential for learning statistics through stories as they provide a rich and meaningful context in which students could interpret the results of the statistical analysis. Thus, the idea of learning statistics through stories aligned with my beliefs about important learning outcomes in statistics education. Finally, the idea of presenting statistical concepts through a story aligned with my beliefs about how to teach statistics. In particular, I believe it is important to teach statistics not as a set of disconnected, abstract topics but as a concrete, interconnected ideas (Friesen & Jardine, 2009). Through the narrative structure of stories, concepts can be introduced as solutions to a problem and, thus, can be presented as concrete rather than abstract (Egan, 1986). Further, the structure of stories allows for the natural progression of one concept to another, which permits for a clear connection made between concepts. Consequently, from this introduction to narrative learning, I began to design stories for use in my teaching. I first implemented them in my statistics course and that of a colleague in the 2016 Winter term. This experience provided me with important information to re-design, implement, and investigate the story-based tasks to teach statistics. I provide more information on this in chapter 3.

In summary, this study arose from my desire to address the discrepancy between the purported learning outcomes and the actual learning outcomes in my post-secondary business statistics course. After trying multiple strategies, I recognized that small changes in the classroom were not sufficient to engender the large changes I wanted. By being introduced to narrative learning, I developed the idea of using stories for students to explore statistics in a way that fit with my views on the nature of statistics, how students learn, and effective teaching practices.

7. Format of Thesis

The thesis is organized into six chapters. The current chapter, provided an overview of the study and the research questions. Chapter 2 presents the literature review for this study. This includes the key theoretical perspectives that informed the study. It also includes a review of relevant studies on students' learning of statistics, students' beliefs about the usefulness of statistics, and students' learning through stories in mathematics and statistics education. Chapter 3 provides details on the methodology of this study. This includes an overview of case study methodology, a detailed outline of the intervention, details on data collection, and the method of data analysis. Chapter 4 provides the findings of the study. Finally, Chapter 5 provides a discussion of the findings in relation to the three research questions, situates the findings within the research literature, and connects the findings to the theoretical perspectives. Chapter 6 provides the conclusion of the study, which includes a summary of the study, the implications of the work, limitations of the study, and suggestions for future research.

Chapter 2 - Literature Review

This chapter deals with two broad categories of the research literature. The first category addresses the *theoretical perspectives* of the key constructs of the study introduced in Chapter 1 while the second category focuses on relevant *empirical studies* related to statistics education. The theoretical perspectives cover statistics and statistics education, learning theories, understanding, the nature and use of stories in learning, and beliefs in mathematics education. The empirical studies cover research on students' learning of statistics, students' beliefs about statistics, and students' learning through stories in mathematics and statistics education.

8. Statistics and Statistics Education

In this section, I outline the theoretical perspectives of statistics and statistics education with a focus on the *nature of statistics* and the *reform movement in statistics education*, respectively.

Nature of statistics

Statistics, as a discipline, focuses on concepts involving data, variation, chance, and skepticism (Cobb & Moore, 1997; Gal, 2002; Wild & Pfannkuch, 1999). However, researchers have also defined statistics as a method, process, or tool and a way of knowing or thinking. For example, it is a method to "get information from [quantitative] data" (Keller, 2018, p. 1) or a "collection of methods for planning experiments, obtaining data, and then organizing, summarizing, presenting, analyzing, interpreting, and drawing conclusions based on the data" (Triola, 1998, p. 4). It is the process of exploring situations and problems by defining the contours of the exploration; collecting relevant data; summarizing, organizing and analyzing the data; and interpreting and presenting the results (Cobb & Moore, 1997).

While these definitions continue to be the focus of statistics education, researchers, such as Wild, Utts, and Horton (2018), point out that "statistics needs to be defined by the ends it pursues rather than the means statisticians have used to pursue them in the past" (p. 7). This shift is related to advancements in technologies that have and will result in changes in the means of doing statistics, but not in the goals of statistics. This focus on goals is reflected in the American Statistical Association (n.d.) definition: "statistics is the science of learning from data, and of measuring, controlling and communicating uncertainty" (para. 1).

As a tool, statistics is used to model real-world processes (Burrill, 2005), make informed decision in an "uncertain environment" (Newbold, Carlson, & Thorne, 2010, p. 2), and evaluate evidence and claims (Ben-Zvi & Garfield, 2004). It uses context to provide meaning to the analysis and interpretation (Cobb & Moore, 1997; Gal, 2002; Wild & Pfannkuch, 1999). Statistics is always grounded in a context (delMas, 2004) and understanding the context is required to provide meaning to the results of the statistical analysis (Cobb & Moore, 1997). Often, statistics is described as a way to solve real-world problems, but statistical inquiries also involve investigating problems where there is a "*knowledge-deficit* or *understanding deficit* problem" (Wild et al., 2018, p. 10).

In statistics education, statistics is also promoted as a way of knowing and thinking with a focus not only on statistical knowledge, but also statistical reasoning and statistical thinking (e.g., Ben-Zvi & Garfield, 2004; Garfield & Chance, 2000). *Statistical knowledge* (sometimes called statistical literacy) includes knowing the skills to calculate statistical measures, and the terminologies and the concepts or key terms in statistics (Ben-Zvi & Garfield, 2004). *Statistical reasoning* is the ability to understand how and why statistical processes work; to justify and explain the reasons behind choices made in the statistical investigations; to understand how the choices made relate to the conclusions drawn; and to interpret the results within the context (Ben-Zvi & Garfield, 2004; delMas, 2004). *Statistical thinking* involves the thought processes that statisticians engage in when performing statistical investigations. It is a broad mode of thinking (Moore, 1998). A statistical thinker recognizes that data are better than anecdotes; the need for properly collected data; that different representations of the data can convey new meaning (transnumeration); that variation is everywhere and in everything; that data are numbers in a context; that chance and probability play a role in statistical analysis; the limitations of statistics; and the purpose, process and logic of statistical investigations (Cobb, G. W., 1992, 2007a; Wild & Pfannkuch, 1999). Statistical knowledge, reasoning and thinking are not mutually exclusive and, in fact, they do overlap (delMas, 2002). For example, the role of context in statistics is inherent both to statistical reasoning and thinking.

For students to gain an overarching understanding of the nature of statistics, they would likely need to take multiple courses in statistics. Thus, for the business statistics course in which this study was conducted, the focus of the nature of statistics was on the usefulness of statistics in the students' everyday lives and the development of relational understanding of the core concepts covered in the course. Students can demonstrate understanding of the discipline of statistics in this study by demonstrating an understanding of the usefulness of statistics and relational understanding of the selected topics.

Reform movement in statistics education

As mentioned previously, the reform movement in statistics education calls for changes in how statistics is taught and provides broad learning goals for statistics courses. Up until the 1990s, statistics courses were often taught from an instrumental viewpoint, which focused on the passive absorption of material and rote learning. The reform movement was initiated by an email focus group of thirty-nine statisticians, funded by the Mathematical Association of America, to address issues they were observing in statistics education (Zieffler, Garfield, & Fry, 2018). Resulting recommendations summarized by G. W. Cobb (1992) include: emphasize statistical thinking; focus on broad concepts and principles instead of solely on techniques ("recipes") and mathematical theory (Everson et al., 2016, p. 7); foster active learning to allow students to "discover, construct, and understand important statistical ideas as well as to engage in statistical thinking" (ibid, p. 18); and use real data (not realistic data) that is interesting and relevant to students. Other researchers built on this work (e.g., Garfield. 1995; Moore 1997), which culminated in the American Statistical Association funding the Guidelines for Assessment and Instruction in Statistics Education (GAISE) college report (Garfield et al., 2005; Everson et al., 2016). These reports, which were developed by statisticians and statistics educators, outlined recommendations for best practices in teaching statistics and proposed learning goals for post-secondary statistics courses. The recommendations for best teaching practices align with those of G. W. Cobb (1992) and also include the effective use of technology both in aiding in calculations and in visualizations of the concepts, and the use of assessments that aid students in learning. The learning theory that informs these recommendations is constructivism (Cobb, G. W., 1992). The intervention in this study was designed to align with all but one of the recommendations. As the stories in the intervention are fictional, the intervention does not align completely with the recommendation to use real data that is relevant to the students. Instead, the data is fictional, but the context for the data was chosen with the intent of being authentic to the students. More specific details about the design of the intervention are provided in Chapter 3.

A few years before the reform movement in statistics began, a reform movement in calculus was also occurring (Rafael, 1997). Calculus was facing similar problems as statistics with students focusing on rote learning. The suggested solutions of the calculus reform movement have parallels to the statistics movement (Moore, 1997). Both movements suggest a change in focus from ‘recipes’ to conceptual understanding (Rafael, 1997). Both movements suggest changes in teaching practice that focus on active learning, the use of technology (both for doing calculations and for learning of content), and changes in assessments practice (Rafael, 1997; Steen, 1988). Finally, both movements rely on constructivism as their pre-dominant learning theory (Steen, 1988). An important difference between the two movements is that the calculus reform movement highlights the importance of students writing about mathematics (Steen, 1988). This difference is important in the context of this study because, through students engaging with the story-based tasks, they are invited to write about statistics, which is not a recommendation of best practices in the statistics reform movement. Finally, the calculus reform movement is mentioned here to highlight that the statistics reform movement is not occurring in isolation but was being developed along with other reform movements in mathematics education and these movements share similar motivations and recommendations.

In addition to recommendations on best practices for teaching statistics, the reform movement in statistics outlines common learning goals for all students who have taken a statistics course (Everson et al., 2016; Gal and Garfield, 1997). In brief, the reform movement recommends the learning goals of a statistics course to be statistical knowledge, thinking and reasoning (Ben-Zvi & Garfield, 2004) as previously defined in the section on the nature of statistics.

Over the years, there have been many interpretations of what the recommendations and learning goals mean in the classroom. In recent years, with the development of specific educational software for learning statistics, a common interpretation of the recommendations involves a focus on informal inferential statistics using randomization techniques and simulations (e.g., Tintle et al., 2014). That is, instructors promote statistical knowledge, reasoning and thinking through active learning using technology.

The learning goals of the reform movement informed the study by providing specific ways that students can demonstrate understanding of selected topics and the discipline of statistics. For example, participants could demonstrate statistical knowledge of a selected topic by producing relevant statistical measures and providing definitions of statistical terms for that topic. Additionally, they could demonstrate statistical reasoning for a selected topic by explaining the reasoning behind concepts for that topic.

9. Learning Theory

The theoretical perspective of learning related to this study involves the learning theory of *constructivism*. As this study focuses on statistics, this theory is discussed mainly from the perspective of mathematics/statistics education. Constructivism is a major learning theory in mathematics education (Cobb, P., 1994; Lerman, 1996) and is the learning theory that informed statistics educators in their understanding of how learning occurs for statistics students (Cobb, G. W., 1992; Garfield, 1995; Garfield and Ben-Zvi, 2007; Moore, 1997).

Though there are multiple perspectives of constructivism, the central tenet to all constructivist theories of learning is that knowledge is actively constructed (rather than passively received) from experiences (Merriam, Caffarella, & Baumgartner, 2007). Thus, unlike behaviourism that sees knowledge as a product, constructivism focuses on knowing as a process and the process is ongoing (Jones & Brader-Araje, 2002, p. 2). Where the constructivist perspectives disagree is on the locus of learning. For example, is it in the head (radical constructivism) or in the "individual-in-social-action" (social constructivism; Cobb, P., 1994, p.

13)? Others take a more pragmatic approach, known as the emergent perspective, which argues that the two perspectives (cognitive and social) put together can complement each other (Cobb & Yackel, 1996). This section explores these three types of constructivism, summarized in Table 2.1, and discusses which perspective is related to this study.

Table 2.1 – *Summary of constructivist perspectives*

	Locus of learning	**Pros**	**Cons**	**Role of the teacher**	**Implementation in the classroom**
Radical constructivism	In the individual’s head	Explains how learning occurs in the individual	Ignores the social and cultural nature of learning	Provide students with activities that challenge their mental structures	Discovery learning, questioning techniques, manipulatives
Social constructivism	In social interactions	Explains how social and cultural practices impact learning	Does not have a clear explanation on how learning occurs	Enculturate students in the culture of mathematics	Group and collaborative learning, peer instruction
Emergent perspective	Both in the head and in social interactions	Recognizes both the social and individual nature of learning	Attempts to meld two differing theories	All of the above	All of the above

Radical constructivism has its roots in Piaget’s work and its main proponent was Ernst von Glaserfeld (Philips, 1995). According to von Glaserfeld (1982, 1990), in radical constructivism, the locus of learning is within the individual who is actively constructing cognitive structures based on their experiences. Knowledge, for an individual, is a collection of cognitive structures that are constructed through experience and reflection. Learning for an individual is the process of constructing more viable cognitive structures that fit within the constraints of current experiences. A cognitive construct is more viable if, when applied to experiences, it is successful; that is, there are no significant conflicts or uneasiness between the structure and the experience.

While radical constructivism focuses on individual constructions, *social constructivism* focuses on the socio-cultural nature of mathematics/statistics learning (Cobb, P., 1994). Social constructivism has its roots in Vygotsky's work and activity theory (Cobb, P., 1994). Learning is seen as an "enculturation into a community of practice" where teachers and students mutually appropriate each other's offerings (Cobb, P., 1994, p. 13). That is, learning occurs through "social interaction that requires negotiation of meaning, under the direction of shared social norms for communication helps support the transformation of informal knowledge to culturally shared formal understanding" (Zieffler, Garfield, delMas, & Reading., 2008, p. 43). Here, the teacher represents the culturally agreed upon representation of statistics. Students learn by participating in cultural practices and becoming better able to participate effectively in these practices (Cobb, P., 1994).

Though these two perspectives may appear incongruent, some authors have argued that each "tells half of a good story, and each can be used to complement the other" (Cobb, P., 1994, p. 17). Thus, an alternative perspective was proposed that takes a pragmatic approach to constructivism, called the *emergent perspective*, which combines the benefits of both perspectives (O'Shea & Leavy, 2013). The goal, then, is not to determine the supremacy of one perspective, but instead to see what the combined theories has to offer (Cobb & Yackel, 1996; Simon, 1995). In the emergent perspective, knowledge is constructed both individually and socially, and "learning is a process of both self-organization and a process of enculturation that occurs while participating in cultural practices, frequently while interacting with others" (Cobb, P., 1994, p. 18). In particular, the emergent perspective recognizes that individual students construct their knowledge by making more viable cognitive structures, but that this construction happens through social interactions within a classroom community. Through social interactions, students negotiate meaning by changing their interpretations through a process of mutual adaptation (Cobb & Yackel, 1996).

The perspective of constructivism that is the best fit for this study is the emergent perspective as an aim of the story-based intervention is to engage students in constructing their understanding of statistics through both individual and collaborative learning experiences. The intervention was designed to provide students with the opportunity to construct their own individual understanding of the concepts, but this is done within "the shared knowledge in the classroom community" (Simon, 1995, p. 119). Thus, in this study, learning is understood to be individual students actively constructing statistical knowledge and understanding in response to their experiences with the story-based tasks, but they do so through collaboration with peers in groups. Students are expected to build their knowledge and understanding based on prior knowledge and exploration of new experiences or information involving statistics.

10. Understanding of Statistics

In this section, I address the theoretical perspective on what it means to understand statistics. One of the goals of this research was to investigate how the intervention impacted students' understanding of statistical topics and the discipline of statistics. As there are different ways of defining understanding in mathematics/statistics education (e.g., Hiebert & Lefevre, 1986; Skemp, 1976/1978; Sierpinska, 1994), it is important to clarify what understanding means in the context of this research. In this work, I focus on instrumental and relational understanding as defined by Richard Skemp.

Skemp (1976/1978) defined instrumental understanding as the ability to perform an algorithmic procedure to solve a problem, but without knowing the reasons for the steps or "rules without reasons" (p. 9). Students who have only this type of understanding get lost when there is a change in the question and are dependent on someone else to help them solve new problems. If they forget the procedure, they cannot recreate it. In the context of statistics, students with instrumental understanding would know how to do a calculation or follow a procedure, but they would have difficulty interpreting the results. Further, they may know the definition of a term, but could not apply the term in practice or know why the definition was appropriate.

Relational understanding, on the other hand, is "knowing both what do and why" (Skemp, 1976/1978, p. 9). Students who have this type of understanding can adapt when a question is changed and can recreate a procedure if forgotten. Once they have achieved relational understanding, they can work independently on new statistics problems. They can also create connections between multiple statistics concepts. Thus, they develop deep understanding of how a concept relates to other concepts and how it can be applied. This is important in statistics as a fundamental component of the subject is understanding how to apply concepts in practice and within a context, and not just understanding the concepts. In the context of statistics, students with relational understanding would know what statistical procedure to use and why it is appropriate to use in multiple contexts; how to perform the procedure in multiple contexts; how the procedure works; how to interpret the results (including knowing what an appropriate result would look like); the limitations of the procedure; and how the various procedures are interconnected. They would also know the reasoning behind the concepts.

Both instrumental and relational understanding are relevant to this study because students can demonstrate either or both to various levels in their learning of the statistics concepts through the intervention. For example, a student may know how to use a formula to produce a statistical measure, but cannot provide a definition of statistical measure, thus demonstrating one aspect of instrumental understanding, but not all aspects. Similarly, a student may successfully interpret the results of a statistical investigation but may not be able to explain how the statistical procedure works, thus, demonstrating one aspect of relational understanding but not the other. Together, then, instrumental and relational understanding provide an appropriate, multi-faceted basis to analyze the data to determine the nature of the participants' constructed understanding of statistics as a result of the intervention.

11. Nature and Use of Stories in Learning

Of importance to this study is the theoretical perspective of the nature and use of stories in learning. Narratives or stories in education come in many forms and uses, some of which are discussed here to establish the perspective used in this study.

What are stories?

Though some authors use narrative and stories interchangeably (e.g., Polkinghorne, 1988), in this study, it is important to distinguish between them as stories have a specific meaning here. Narrative is a larger umbrella term that refers to a sequence of events that are connected to a theme, internally consistent, and present a coherent whole. A story is a specific type of narrative that has a clear beginning and end, and tells a sequence of events with a character that is driving the events towards a solution to a problem or conflict (Egan, 1986). Stories have characters, plots, context, conflict, imagery, emotions, and humor (Carter, 1993; Roberts & Stylianides, 2013; Zazkis & Liljedahl, 2009). Further, stories happen over a period of time – “it is not one single moment or a snapshot in time”, have a plot, which is understood by making connections between the events and characters, and requires interpretation by the reader (Rossiter, 2008, p. 418). Martin Luther King Jr.’s *I have a dream* speech is an example of a narrative as it connects events around the theme of equality, but is not a story as it does not have characters or a plot. Other examples of non-story narratives include essays, arguments, and reports. In this study, the story-based statistics tasks are stories rather than narratives.

Defining the statistics tasks as stories rather than narratives resulted in detailed description of the context and problem of the story; development of characters with distinct personalities; addition of dialogue between the characters; use of informal language that would fit with the personalities of the characters; physical descriptions of settings and characters; and sequential progression of events over a period of time. To illustrate the impact of personalities of the characters, in one story, one character was very lazy and indifferent to the work and another character was very prim and proper. Thus, the dialogue for the former character included more slang, while the dialogue for the latter character was more formal. If the tasks were defined as narratives, then none of these aspects would have been included. Further, as the key element of the intervention is the story-based tasks, the definition of story informed the data analysis as it provided an avenue to consider how the nature of stories impacted students’ understanding of topics and their beliefs about the usefulness of statistics.

Use of stories

Stories present content in a meaningful context (van den Heuvel-Panhuizen, van den Boogaard, & Doig, 2009), engage students at an imaginative and emotional level (Egan, 1986; Nicol & Crespo, 2005), and aid in the process of meaning making (Clark & Rossiter, 2008). Rossiter (2008) provided three general areas on how narratives can be used in adult learning: to story the curriculum, for autobiographical learning, and to teach content. The stories in this study are used to teach content. Stories to teach content involve the use of story-telling through activities such as case studies and educational drama. They can be a "potent supplement to content that is chiefly quantitative or conceptual in nature" (Rossiter, 2008, p. 421). Stories are used to teach content because stories are authentic human experiences, they are entertaining, and they induce emotions (Clark & Rossiter, 2008). Further, storytelling allows students to connect a concept or skill to "a particular human hope, intention, fear, or whatever, then [they] can embed the skill in a context that is meaningful" (Egan, 1986, p. 77).

Stories to teach content can be used in multiple ways in the mathematics and statistics classroom. Zazkis and Liljedahl (2009) outlined six ways that this can be done: stories can ask a question (usually in the form of a word problem), accompany a topic to add flavour or provide relevant anecdotes, be used to motivate and introduce a topic, be intertwined with a mathematical topic, explain a concept, and be used to introduce an activity. In each of these examples, the stories have been written by the teacher or an outside source, but not by the students. Another way stories can be used in mathematics and statistics classrooms is by asking students to write their own stories (e.g., Cho, Osborne, & Sanders, 2015; Sherwood, 2018). The stories in this intervention do not fit neatly into any of these categories. Instead, the stories in this intervention motivate and introduce a topic, explain a concept, and intertwine with a mathematical topic. In addition, they are a combination of stories presented to students and stories written by students.

Stories in mathematics education

The term stories in mathematics education is used very broadly. For example, stories can refer to word problems (Gerofsky, 2004), simple stories about sharing to understand parity (Roberts & Stylianides, 2013), and complex stories about a square living in a two-dimensional world traveling to other worlds with different dimensions (Nicol & Crespo, 2005). How the stories are told range from a teacher dressing up as Cleopatra to discuss how people may have developed the idea of shapes (Toor & Mgombelo, 2015), to students playing an adventure story video game (Giannakos, Chorianopoulos, & Jaccheri, 2012), to the teacher and students acting out the stories together (Ranatunga, Strodl, & Sorin, 2014). However, stories, as used in this study, have a clear beginning and end, and tell a sequence of events with a character that is driving the events towards a solution to a problem or conflict (Egan, 1986), are not common in mathematics classrooms (Albano & Pierri, 2016), but are used in early childhood education (e.g., Anderson, Anderson, & Shapiro, 2004; Casey, Erkut, Ceder, & Mercer Young, 2008; Dubon & Schafer, 2010). The lack of stories across educational levels is changing as educators recognize the power of stories (Albano & Pierri, 2016; Cho et al., 2015).

Stories have been used in mathematics education in traditional story form, non-traditional story form, and student written stories. Traditional story form included picture books aimed at young children, realistic stories set in the real world, and fantasy fiction set in fantasy worlds. Non-traditional story form includes cartoons and video games. Student written stories involve students writing stories in a mathematics classroom in various ways.

12. Nature of Beliefs

Finally, the theoretical perspective of the nature of beliefs is also relevant to this study. One aim of this study is to determine the impact of the intervention on the participants' beliefs about the usefulness of statistics. Beliefs can be seen "as lenses through which one looks when interpreting the world" (Philipp, 2007, p. 258). They form one's understanding of the world and reality (Beebe et al., 2011). They are highly personal (Nespor, 1987) and include one's ideas and opinions; perspectives; and truths about oneself, a domain or a social-context (Gal, 2002;

Pajares, 1992). Thus, beliefs about the usefulness of statistics, in this study, are understood to be students' viewpoints or opinions about the usefulness of statistics. While beliefs can be challenging to study, they can be inferred by what individuals say and do (Pajares, 1992; Philipp, 2007). Thus, this approach was used to frame the methodology regarding the design of the instruments used to investigate students' beliefs about the usefulness of statistics and how the resulting data was analyzed.

As this study investigated how stories impact students' beliefs about the usefulness of statistics in their everyday life, it is relevant to outline how beliefs can change. Green (1971) suggested that beliefs are either central or peripheral. Central beliefs are the strongest and are the most resistant to change, while the peripheral beliefs are not as strongly held and are easier to change. Additionally, he distinguished between evidentially and non-evidentially held beliefs. Evidentially held beliefs are supported by evidence and, thus, can be changed by evidence. While non-evidentially held beliefs are held because the belief "happens to support a belief he already holds" (Green, 1971, p. 49). In this study, the story-based tasks were designed to provide students with examples of the usefulness of statistics in everyday life with a particular focus on examples related to their future careers in business. Thus, from Green's perspective, if the participants' beliefs about the usefulness of statistics are peripheral and evidentially held, the story-based tasks have the potential to impact their beliefs by providing evidence that either further supports their beliefs or provides evidence against them. But if their beliefs about the usefulness of statistics are core beliefs or non-evidentially held, then the stories may not impact their beliefs. This perspective outlines how beliefs can change and what factors can impact that change, which will inform the interpretation of the data collected about beliefs.

In summary, the preceding five sections outlined the key theoretical perspectives for this study regarding the nature of statistics and statistics education, learning theory (constructivism), understanding, stories, and beliefs. The next three sections present the empirical research on students' learning of statistics, students' beliefs about the usefulness of statistics, and students' learning through stories.

13. Research on Students' Learning of Statistics

This section provides current empirical research about students' learning of statistics and consists of five categories. The first three categories deal with research of students' learning for three areas of statistics, respectively, that are also the selected topics for this study: descriptive statistics, sampling distributions of sample means, and inferential statistics. The fourth category presents research on the impact of technology on students' learning of statistics. Finally, the fifth category examines research on students' learning of statistics through context-rich problems. Technology and context-rich problems are included in the review as the intervention involves computer simulations and stories are a type of context-rich problem.

Students' learning of descriptive statistics

This section presents studies on three key topics of descriptive statistics, respectively: visual representations, measures of central tendency, and measures of variation.

Visual representations. Typical visual representations in a post-secondary statistics course include histograms, box plots, and graphs. Research shows that students have difficulty in understanding or creating histograms. For example, Zaidan, Ismail, Yusof, and Kashefi (2012) found that when students were asked to label histograms of the salary of individuals over the age of 40, students labelled the horizontal axis as salaries and the vertical axis as age. Students also have issues reading histograms. Kaplan, Gabrosek, Curtiss, and Malone (2014) found that students identified the mode of two data sets being the same based on the heights of the bar being the same rather than where the bars were positioned.

Students also have difficulty with box plots. For example, Lem, Onghena, Verschaffel, and Van Dooren (2013) found two issues with understanding that were unique to box plots. The first issue was that students stated that the lowest value was the first quartile. Thus, they did not recognize the whisker as going from the minimum to the first quartile. They also believed that the larger the area of the box, the more data values there were in it, rather than recognizing that all boxes contain 50% of the data. This latter issue was supported by research done by Pierce and Chick (2013) and Edwards, Özgün-Koca, and Barr (2017).

Interpretations of graphs are also important for describing the distribution of data. Arnold and Pfannkuch (2014) utilized a distribution framework to examine students' descriptions of distributions both before and after specific lessons on how to describe distributions. They focused on two students who were 14-15 years old and found that the students, after the lessons, could describe the distributions in a similar way that a statistician would. Kaplan, Lyford and Jennings (2018) used the same framework as Arnold and Pfannkuch, but examined how changing the wording of the prompts impacted undergraduate students' descriptions of distributions. In particular, they changed the wording of the prompts from generic (e.g., describe the distribution) to specific (e.g., describe the distribution, being sure to explain what the graph tells you about the context). They found that more specific prompt appeared to "cue" students to provide a more complete description (p. 97), but that most students, regardless of the prompt, only mentioned centre and shape, and failed to mention variation.

Measures of central tendency. Measures of central tendency typically include the mean, median and mode. Students often have a strong instrumental understanding of the mean (Cooper

& Shore, 2008; Dubreil-Frémont, Chevallier-Gaté, Zendrera, 2014), but have conceptual difficulties with it. For example, Mathews and Clark (2007) found that some students who received an A in a university statistics course could give a definition of the mean, while others confused mean with the mode and proportion. Zaidan et al. (2012) found that post-graduate students' difficulties included believing that if the mean was three, then the data set had to be either composed entirely of threes (e.g., 3, 3, 3) or multiples of three (e.g., 3, 3, 3 or 9, 9, 9). Guimarães, Gitirana, Marques, and dos Anjos (2010) found that elementary school students struggled with understanding that the mean is affected by each value in the data set. Cooper and Shore (2008) found that when undergraduate students were trying to find the mean using a histogram, they lost track of the context and gave unrealistic results. For example, they stated that the mean age of a restaurant patron was four years old.

Histograms continue to cause problems for students when they are determining the measures of central tendency. As students have difficulty reading histograms (as discussed above), when asked to find the mean or median of a data set represented by a histogram, they ignore the frequencies and find the mean or median of the values on the horizontal axis, or they use the heights of the bars as data values and find the mean or median of those (Cooper & Shore, 2008). When finding the mode for two histograms, students look at the height of the bars. If the heights are the same, the value of the modes are the same regardless of where the bar is positioned on the horizontal axis (Kaplan et al., 2014). When determining the relationship between the mean and median in a histogram that is skewed to the right, students ignore the shape and the frequencies to determine that the mean is less than the median (Cooper & Shore, 2008). Students also struggle with interpreting the median on a box plot as they focus on the size of the box rather than on the definition of the median (delMas, Garfield, Ooms, & Chance, 2007).

As many of the studies rely on students interpreting visual representations of data to determine students' understanding of measures of centre, it is not clear if the misconceptions relating to measures of centre outlined in the studies are due to students' actual misconceptions of measures of centre or misconceptions about the visual representations. That is, it is not clear where the misconceptions are arising from when the studies utilized visual representations to investigate students' understanding of measures of centre.

Measures of variation. Variation is one of the most important topics in statistics (Cobb & Moore, 1997; Snee, 1999). Yet, in one study, high school students found variation to be a very difficult topic (Chan & Ismail, 2013). In most statistics courses, variation is measured using range, standard deviation and interquartile range (Chaphalkar & Leary, 2014). Unlike measures of central tendency, most students struggle with even an instrumental understanding of standard deviation. Gougis et al. (2016) found that only 17% of science students in their study had a relational understanding of variation. While Mathews and Clark (2007) found that none of the eight undergraduate students in their study who got an A in a post-secondary statistics course had even a partial instrumental understanding of standard deviation. In particular, they found that students could calculate the standard deviation, but did not know what the calculated value meant and could not provide a definition for the standard deviation.

Common issues in understanding standard deviation include that it is found by determining the distance between the data values in the sample (Mathews & Clark, 2007); it is the distance between the mean and one data value in the sample (Lavy & Mashiach-Eizenberg, 2009; Mathews & Clark, 2007); the standard deviation and the mean are equal (Chan & Ismail, 2013); the smaller the standard deviation, the smaller the mean (Orta & Sanchez, 2011); and if the sample sizes for two data sets are the same the standard deviations are equal (Chan & Ismail, 2013).

When comparing variation visually (usually represented by histograms), common issues included looking at the height of the distribution rather than the width (Zaidan et al., 2012), looking at the bumpiness of the data on a histogram to determine the variation (i.e., the more bumpy it is, the more variation there is; Cooper & Shore, 2008), ignoring the frequencies and only looking at the range (Cooper & Shore, 2008; Dabos, 2014; delMas et al., 2007), and believing that the more normal the data, the smaller the variation (Dabos, 2014; Kaplan et al., 2014). Like the studies on measures of centre, many of the studies in this review also used visual representations of the data to explore students' understanding of variation. Thus, again, it is not clear whether the misconceptions arise from difficulties with understanding the visual representations or difficulties with understanding the measures of variation.

In summary, students have difficulties constructing and reading visual representations. Though they can calculate and know the definition of measures of centre, they do not have relational understanding of the concept. Finally, students may know how to calculate the standard deviation, but few know even the definition of it.

Students' learning of sampling distributions

Sampling distributions are the bridge between descriptive statistics, probability, and inferential statistics. To understand sampling distributions, students need to understand distributions, variability, normal distributions, and sampling (Chance et al., 2004). Additionally, sampling distributions provide the foundation for formal statistical inference (Sotos, Vanhoof, Van den Noortgate, & Onghena, 2007). Thus, students' understanding of sampling distribution is needed for hypothesis testing and confidence intervals, but to have this understanding they need an understanding of descriptive statistics, probability and sampling.

Due to the complexity of sampling distributions, it is not surprising that many students struggle with understanding the concept (Cobb, G.W., 2007b). There are three main properties of the sampling distribution: centre, variation and shape. There are issues with understanding of all three of these properties, both with how to apply the sampling distribution and how the properties relate to the central limit theorem.

Students often confuse the terms related to sampling distributions. For example, students confuse the terms standard deviation and sampling error (Reaburn, 2010); the multiple notations for the different variations (Stevens & Palocsay, 2012); and the terms sample distribution, population distribution and sampling distribution (Chance et al., 2004). Further students do not understand that the sampling distribution is a distribution of statistics (Chance et al., 2004). Based on the confusion of notation and terminology, this suggests that students do not easily distinguish between the concepts of sample, population, and multiple samples from a population.

One of the key concepts to understanding variation in the sampling distribution is that as the sample size increases, the statistic being measured in the random sample will likely get closer to the parameter (i.e., the law of large numbers). Due to this, as the sample size increases, the variability between the statistics (i.e., sampling variability) in multiple samples will decrease. Many students have difficulties making this connection. For example, some students believe that as the sample size increased, the variation of the sampling distribution also increased or stayed the same (Chance et al., 2004). Brown and delMas (2018) found that by having students explore the concepts of swamping (i.e., as sample size increases, outliers have less impact on the sample mean) and heaping (i.e., sample means concentrate around the population mean), they developed a stronger understanding of sampling variability. This suggests that certain learning tasks can have a positive impact on students' understanding of sampling variability.

The central limit theorem states that as the sample size increases, the sampling distribution of the sample means approaches a normal distribution. Students misapply this theorem to the population distribution and thus believe that as the sample size increases, the sample becomes more normal (Stevens & Palocsay, 2012). Or they believe that the central limit theorem, as stated above, applies to any statistic including proportion (Sotos et al., 2007). They also believe that the sampling distribution approaches the population distribution for larger sample sizes (Chance et al., 2004). Finally, students also confuse the law of large numbers and the central limit theorem (Chance et al., 2004).

When applying the concepts of the sampling distribution, students often exhibit instrumental understanding. For example, when presented with the population distribution, only 44.2% of 719 undergraduate students could identify the associated sampling distribution (delMas et al., 2007). When asked to make inferences using the sampling distribution, conflicts in their understanding became apparent. For instance, when asked to determine if a sample mean was considered an outlier, they used reasoning based on the population distribution rather than the sampling distribution (delMas et al., 2007). If presented with a histogram of the sampling distribution, 82.9% of 718 students could not determine if a sample mean was unlikely if the sample mean was not explicitly in the sampling distribution (Weinberg, 2014). Another study found that students who were 16-18 years old could articulate the basics of the sampling distribution, but could not apply the concepts (Jacob & Doerr, 2014). For example, the study found that students produced a sampling distribution for how a Monopoly house lands when tossed using 32 tosses, but when presented with the sampling distribution for a Monopoly hotel that had been tossed 200 times, they didn't see how they could use it to determine if the probability of landing upright was the same as that of the house. Instead they felt they needed to create a sampling distribution that was the same as the house (i.e., only had 32 tosses rather than 200).

In summary, students have difficulty understanding most aspects of sampling distributions, which is not surprising given that they had misconceptions about the underlying concepts of sampling distributions (i.e., centre, variation, distribution).

Students' learning of inferential statistics

Inferential statistics often forms the backbone of a statistics class and is considered "the cornerstone of statistics" (Makar, Bakker, & Ben-Zvi, 2011, p. 154). That is, the end goal of many statistics courses is for students to learn how to do inferential statistics. There are two types of inferential statistics: formal and informal. Formal statistical inference refers to statistical inference that uses clearly defined and agreed upon statistical procedures such as confidence intervals and hypothesis tests (Makar & Rubin, 2009). The definition of informal statistical inference is more ambiguous. What is agreed upon is that it is a generalization about the population based on sample data that does not use formal inferential techniques (Makar & Rubin, 2009; Zieffler et al., 2008). This section outlines research on students' learning of formal inferential statistics and of informal inferential statistics.

***Research on formal inferential statistics*.** Formal inferential statistics in post-secondary education is often taught through confidence intervals and hypothesis tests. Thus, this section examines research on confidence intervals and hypothesis tests.

Confidence intervals. As students are prone to making incorrect conclusions based on hypothesis tests, it has been suggested that students use confidence intervals instead (Fiddler, 2006; Kalinowski, 2010). Though students tend to make more appropriate conclusions using confidence intervals (Fiddler, 2006), many students' understanding of confidence intervals is instrumental. The most prevalent interpretation of a confidence interval iss that it contains a percentage (equivalent to the confidence level) of data values from the population (Canal & Gutiérrez, 2010; delMas et al., 2007; Fiddler, 2006; Reaburn, 2010; Stevens & Palocsay, 2012). This interpretation is perhaps explained by the misconception that the confidence interval was constructed from the population distribution (Canal & Gutiérrez, 2010). Another misconception iss that the confidence interval contains a percentage of possible sample means (delMas et al., 2007; Reaburn, 2010). This suggests that perhaps these students were understanding that the confidence interval was constructed from the sampling distribution, but were incorrectly interpreting what that means for a confidence interval. Deeper issues students have are incorrectly interpreting the confidence level as the probability that the confidence interval contains the sample mean or not understanding that the confidence interval contains the sample mean (Canal & Gutiérrez, 2010). Canal and Gutiérrez's (2010) study was particularly interesting as the researchers asked students and experts to answer a series of questions on the confidence intervals. Though experts did significantly better than students, around 25% of experts got the interpretation of the confidence interval questions wrong. This led the authors to conclude that students' misconceptions are "often born of their own teachers' misconceptions" (p. 3).

An interesting interpretation of a confidence level by students is the chance that the interval contains the parameter being estimated (Jacob & Doerr, 2014; Reaburn, 2010). Students with this interpretation have the correct understanding that the confidence interval is estimating a parameter, but have an incorrect interpretation of the confidence level. I would suggest that students with this interpretation have a relational understanding of confidence intervals for a first-year student because why this is incorrect is quite subtle and may involve multiple courses to see why.

Hypothesis testing. Students' issues with understanding hypothesis tests start with their understanding of why a hypothesis test is conducted and continue to the end of the hypothesis test when they interpret the results.

A common misconception with why hypothesis tests are performed is that it was to determine if the null hypothesis is true or false (Krishnan & Idris, 2014; Sotos, Vanhoof, Van den Noorgate, & Onghena, 2009; Stevens & Palocsay, 2012). I believe that this issue arises because students did not understand the difference between *assuming* the null hypothesis is true and *proving* that the null hypothesis is true. A minority of undergraduate students (6 out of 15) believed that the goal of a hypothesis test was to reject the null hypothesis test (Krishnan & Idris, 2014). There are also issues with determining the alternative and null hypotheses. When determining the hypotheses, some students struggle with determining the appropriate inequality to use (Sotos et al., 2009; Stevens & Palocsay, 2012) and the tail of the test (Stevens & Palocsay, 2012). Some believe that the hypotheses could refer to either statistics or parameters (Sotos et al., 2009).

Understanding what the *p*-value and the level of significance α represent are difficult for many students. Common misconceptions students have with understanding the *p*-value include that the *p*-value is the probability of committing a type I error, the *p*-value is the probability the null hypothesis is true, and the *p*-value is the probability that the null hypothesis is true assuming the observed data (Reaburn, 2010; Sotos et al., 2009). The understanding of the *p*-value as a conditional probability is missed in the second situation and there os confusion about the conditions in the last situation. Similar issues with the understanding of the level of significance occur. For example, students state that the level of significance is the probability of rejecting the null hypothesis; it is the probability that either the null or alternative hypothesis is true; and it is the probability that the null hypothesis is true assuming the null hypothesis has been rejected (Sotos et al., 2009). The latter, once again, results from confusion about the conditions of the probability.

In summary, from the review of research, many students could properly calculate confidence intervals and *p*-values. Many students could also interpret confidence intervals, but have difficulty explaining what a confidence interval was. Few students could explain the process and goals of hypothesis tests as parts of statistical investigations.

Research on informal inferential statistics. Unlike most of the other studies in this review, the studies that examine informal inferential reasoning often focus on what students learned rather than on their misconceptions. In addition, most of the research suggests that using informal inferential methods in a statistics course aid students in developing relational understanding of the topic of inference. For example, Makar et al. (2011) investigated the development of informal inferential reasoning and found that students' informal inferential reasoning supported concept development. This is reinforced by a study done on a course designed around randomization tests, for university students, which resulted in an increase in students' conceptual understanding of statistics (Tintle et al., 2014) and the greater retention of the material four months after the course compared to those who took a traditional course (Tintle, Topliff, VanderStoep, Holmes, & Swanson, 2012). For example, the study found that in the area of data collection and design, students in the randomization class improved their results on the Comprehensive Assessment of Outcomes in a First Statistics Course or CAOS test (a common research instrument used to investigate students' understanding of key statistical concepts;

delMas et al., 2007) by 18.2% from the beginning to the end of the course, while students in the traditional class only improved their results by 1.6% (Tintle et al., 2014).

Yet not all research had positive results. Castro, Garcia and Sanchez (2018) found that teaching a course from the perspective of informal inferential statistics resulted in similar misconceptions as was found when courses focused on formal inferential statistics. In particular, they found that high school students had difficulty defining the rejection region, calculating the empirical *p*-value, and believed that the inferences were proof of truth.

Impact of technology on students' learning of statistics

Inferential statistics relies heavily on computer simulations, thus the relevance of studies on technology. Researchers have looked at various ways that technology can be used in learning statistical topics including sampling variability, sampling distributions, and inferences. Research into the use of technology includes how the use of computer simulations influence reasoning about statistics and examines the development of students' understanding.

A possible advantage of simulations is its power of visualizing abstract concepts. Budgett and Wild (2014) investigated how Visual Inference Tools (VIT – an educational software designed to explore the concepts of statistical inference) impacted how students visually reasoned about inference. They found that VIT helped students integrate inferential concepts, that the visualizations became part of student thinking, and that students linked the steps of inference visually. In another study, Pfannkuch & Budgett (2014) found that not only did VIT promote visual reasoning, it also promoted versatile thinking.

In an effort to promote statistical thinking, Garfield et al. (2012) used TinkerPlots 2.0 (another type of educational software designed to explore concepts of statistical inference) and a curriculum focused on randomization tests and bootstrapping. They found that, after the course, students demonstrated the beginnings of statistical thinking by recognizing the need for a model and collecting data, and understanding how to evaluate results using *p*-values.

Researchers have also investigated how technology impacted student understanding of statistical concepts in general. In particular, they were concerned with whether better conceptual understanding was demonstrated by students in a traditional class versus in a technology driven class. Pfannkuch, Arnold, and Wild (2015) investigated how the use of the computer program VIT influenced students' understanding of sampling variability. A feature of the program allows for the creation of many box plots superimposed on each other to visually demonstrate sampling variability. It can also show the same results based on different sample sizes to demonstrate how sampling variability changes in relation to increasing sample size. The researchers found that, through the use of learning activities that used VIT, students could start to understand that samples are used to make conjectures about the population, that sampling variability exists, how sampling variability changes depending on sample size, and how that affects making decisions.

As the CAOS test is a common assessment of students' conceptual understanding, it is often used in studies to determine the impact of technology on conceptual understanding. Garfield et al. (2012) found that students who were taught a course using randomization tests and bootstrapping did better on the CAOS test when compared to those in a traditional class. In fact, they found that those in the experimental group did well even on topics that they were not explicitly taught. As mentioned previously, Tintle et al. (2014) and Tintle et al. (2012) investigated how a course in randomization affected students' conceptual understanding compared to a traditional class. Once again, those in the randomization class did significantly better overall than the traditional class based on the CAOS test. In particular, they did better on the topics of data collection and study design, probability, and hypothesis testing, but the traditional class did better in the area of descriptive statistics (Tintle et al., 2014). Tintle et al. (2012) investigated retention of conceptual understanding using the CAOS test. Four months after the course was completed those who had taken the randomization class showed significantly more retention of the course material than those in the traditional course. In fact, in the areas of bivariate data, sampling variability, and hypothesis testing the students in the randomization class did slightly better, on average, four months later than they did at the end of the class. Chance et al. (2004), Watson (2008), and Zieffler, delMas, Garfield, and Brown (2014) research provided further support that simulations help students' understanding of concepts in statistics.

Though most of the research focused on simulations, Kreetong (2014) used a computer game called Ship Odyssey to have students explore the law of large numbers and measures of centre to find treasure by sending out rats who came back with noisy data on where the treasure was located. Kreetong found that playing the game multiple times helped students understand sampling variability and, by having them use varying numbers of rats, students explored how more rats lead to better estimates of where the treasure was.

All of the studies did not produce positive results though. Saldanha and McAllister (2014) had mixed results when they used bootstrapping to explore inference. They found that some students had difficulty reading the results of the simulations and, due to this, could not make inferences. Yet those who could read the simulations were successful in making inferences. This is perhaps linked to the issues in understanding histograms outlined above. Ireland and Watson (2009) had similar results involving simulations and probability.

Watkins, Bargagliotti, and Franklin (2014) demonstrated how simulations could lead to issues in understanding sampling distributions. They found that when practicing teachers explored sampling distributions using simulations, they believed that the mean of the sampling distribution of the mean (SDM) approached the mean of the population as the sample size increased (as opposed to equaling it). Watkins et al. realized that the issue was that the simulations produced an empirical SDM and, due to this, the mean of these SDMs were different from the population mean, but they did get closer to it as the sample size of the empirical SDMs increased, which lead to issues in understanding.

In summary, there appears to be evidence that the use of technologies can help to support students' learning of many statistical concepts. But there are potential pitfalls to students' learning when clear connections are not made between experimental and theoretical results.

Role of context-rich problems in students' learning of statistics

Context-rich problems present meaningful situations that are authentic to the students (Langrall, 2010), use real-life or realistic data (Neumann, Hood, & Neumann, 2013), simulate the unstructured problems that statisticians encounter (Kim, Alberts & Thatcher, 2014), have an affective component (Pratt et al., 2011), and require more than a couple of sentences to answer (Pfannkuch et al., 2010). Within the literature, examples of context rich problems include instructional case studies (e.g., Pariseau & Kezim, 2007; Hiedemann & Jones, 2010), service learning projects (e.g., Phelps & Dostilio, 2008), statistical consulting (e.g., Kim et al., 2014), computer simulations of real-world problems (e.g., Pratt et al., 2011), and context-rich problems embedded within the curriculum (e.g., Lock & Meng, 2010). This section examines research on the role context-rich problems play in students' learning of statistics. Stories are a type of context-rich problem, but as they are a major part of the intervention, their role in learning is discussed separately from context-rich problems later in this review. Thus, this section focuses on context-rich problems that are not stories.

Students in statistics classes that use context-rich problems throughout the course perceived their learning to be better (Nowacki, 2011) and that they would be more likely to remember the material (Neumann et al., 2013). Students who learn using instructional case studies do better on comprehensive final exams compared to those in a traditional course and there is less variation in the final exam results for those who learn using case studies (Pariseau & Kezim, 2007). Students who participate in statistical consulting see benefits by learning a topic in-depth, applying knowledge, and developing proficiency (Kim et al., 2014). They also see the greatest improvement in their skills in the areas of data collection, interpretation of data, and knowledge synthesis (Kim et al., 2014). Context-rich problems also help students delve deeper into statistical concepts (Hiedemann & Jones, 2010) and integrate the various concepts (Nowacki, 2011). They also help students understand the relationship between statistics and scientific inquiry (Nolan & Temple Lang, 2015), and make connections between their statistical knowledge and the problems presented (Dierdorp, Bakker, Eijkelhof, & van Maanen, 2011). Pariseau and Kezim (2007) found that students who used instructional case studies had significantly higher perceptions of their ability to communicate about statistics and make decisions using statistics compared to the students who learned in a traditional manner.

Context-rich problems also aid students to learn specific statistical concepts. In particular, context-rich problems help students understand the importance of aggregate, recognize trends in data, understand variability in a context, and understand how data can be used as evidence for making inferences (Dierdorp et al., 2011). Students make inferences that were more realistic from the data when context was provided (Dierdorp et al., 2011).

Gil and Ben-Zvi (2011) investigated how students explained results using informal inferential reasoning for context-rich problems. They believed that a good explanation of statistical inference includes “an account of the *why* and *how* of the inference” (p. 91). They found that students provided descriptive (or how) explanations (i.e., referred to the results of the data analysis) or abductive (or why) explanations (i.e., provided a reason for the inference based on context) when answering context-rich problems. The study found that context-rich problems allowed students to explain more fully their statistical inferences.

Context also helps students have a more critical disposition when evaluating the results of a data analysis. For example, when Langrall (2010) asked students what might be limitations to the data presented about video games sales, they pointed out missing games, wondered about price of games and console, and whether the sales included used games.

Context-rich problems are also beneficial for a diversity of learners. When students choose their own topic for a final research project, the weaker students are able to do the project with guidance, which provide them with a sense of accomplishment. The stronger students go beyond what is asked, and develope stronger statistical thinking and understanding of statistical methods (Forster & MacGillivray, 2010).

Pfannkuch’s (2011) results on the benefits of context-rich problems were mixed. She found that data-context helped students learn how to make abstractions about concepts; understand the purpose of the study; make connections between the statistical concepts and the interpretation of the story the data was telling; and “enculturated” students into the statistical investigation (p. 43). But the context could, on occasion, be a distraction and took students’ attention away from learning the theory. Wroughton, McGowan, Weiss, and Cope (2013) found that students who had strong opinions about a topic tended to provide more irrelevant or personal information in their analysis particularly when the results of the study did not agree with their opinions. That is, confirmation and belief bias impeded their ability to successfully engage in statistical thinking. Yet there appears to be a conflict in the literature about what is deemed irrelevant information. Wroughton et al. (2013) considered an explanation based on the price of PCs versus Macs to be irrelevant when evaluating results of a study, but Langrall (2010) considered students to be using a critical disposition when they used the price of video games in evaluating results. This discrepancy could be explained by the age of the students. Langrall's (2010) study was on young children while Wroughton et al.'s (2013) was on university students. The level of explanation may be expected to be higher in the latter study.

In summary, research suggests that context-rich problems appear to help students retain information, be more confident in their understanding, make more appropriate inferences, develop critical dispositions, and more. But context-rich problems can also distract students from the theory and, if a student has strong opinions about a context, it may impede their ability to engage in statistical thinking.

Summary of students' learning of statistics

From the review of the literature, we can see that there are many studies that investigate students' learning of statistics. Often, the focus on the studies is on students' learning of specific statistical topics or on examining the role of an intervention (e.g., technology or context-rich problems) on students' learning. Most studies on students' learning focus on where students have misconceptions (with the exception of the studies done on informal inferential statistics). Further, many of the studies use quantitative methods to investigate student learning. For example, there are multiple studies that compared the results of the CAOS test from the beginning and the end of the term to determine the impact of an intervention. Finally, though there are studies on the use of technology in statistics and studies on context-rich problems in statistics, there was only one study found that covers both (see Blackburn (2016) below). Thus, there is room within the field of statistics education for research that focuses on students' understanding in a statistics course (rather than their misconceptions), more qualitative studies on students' learning, and studies that examine the intersection between technology and context-rich problems.

14. Research on Students' Beliefs about Statistics

This section provides an overview of the empirical research on students' beliefs about statistics. Compared to students' learning of statistics, students' beliefs about statistics is studied less frequently. In a recent review of statistics education research, Petocz, Reid, and Gal (2018) examined 653 articles on statistics education that were published between 2010 and 2014. They found only 45 articles on students' attitudes towards statistics compared to the 392 articles that discussed statistical knowledge, reasoning or thinking (p. 85). Another important aspect regarding research on students' beliefs about statistics is that researchers rarely distinguish between attitudes and beliefs, and often use the terms interchangeably. For example, Schau and Emmioglu (2012) used the Survey of Attitudes Towards Statistics (SATS-36) to examine students' attitudes towards statistics, but when summarizing their results, they described students' beliefs about the usefulness of statistics. Further, the terms value and worth are also used to describe usefulness. Thus, this review examines research on students' beliefs, perceptions and attitudes about the value, worth, and usefulness of statistics.

Belief in the usefulness of statistics is one of the learning goals of statistics education (Gal, 2002). For students to be motivated to learn a subject, they need to appreciate the usefulness of what they are learning (Ambrose, Bridges, DiPietro, Lovett, & Norman, 2010). Motivated students may persevere through more difficult problems (Gal, Ginsburg, & Schau, 1997). In fact, if students do not see the usefulness of statistics, they are less likely to study the course material (Murtonen et al., 2008). If students leave a statistics course without seeing the usefulness of the subject, it is unlikely that they will use what they have learned (Schau & Emmioglu, 2012). Students who do not see the usefulness of statistics are also more likely to have anxiety towards statistics (Onwuegbuzie, 1997).

Many students begin a statistics course with a neutral view of the usefulness of statistics (Neumann & Hood, 2009; Hood, 2013; Schau & Emmioglu, 2012; Tsao, 2006). Attitudes towards the usefulness of statistics do not necessarily improve after taking a statistics course (Carnell, 2008; D'Andrea & Waters, 2002; Gordon, 2005; Murtonen et al., 2008; Schau & Emmioglu, 2012). In fact, Gordon (2005) found that only 7% of students thought that the statistics was useful in life near the end of a first-year university statistics course. Based on this research, Kerby and Wroughton (2017) investigated when during a course students' beliefs about the usefulness of statistics began to become more negative. They found that students' beliefs were at their lowest in the middle of the term.

Some researchers have examined how different delivery methods may impact attitudes. For example, Neumann and Hood (2009) examined whether having students write a wiki rather than a traditional report impacted students' attitudes about the worth of statistics. They found that those who wrote the wiki had a slight improvement in their attitudes from the beginning of the term to the end, while those who wrote a traditional report had no significant change. Carnell (2008) did a similar study, but instead of a wiki, she had students design their own data collection projects. She found that doing the student-led project had no significant impact on students' attitudes towards the usefulness of statistics.

Most of the studies described above use a Likert-scale type survey to investigate students' attitudes and beliefs. Songsore and White (2018) used a qualitative approach to examine what students perceived as the relevance and value of statistics in their future. They found that students' perceptions of possible areas where they would find statistics useful fell into three general areas: 1) usefulness in terms of practical life skills, 2) usefulness in future career, and 3) usefulness in specific discipline (p. 134). The most common future use was in the area of practical life skills.

Research on beliefs also examines the impact beliefs have on students' achievement. The research suggests that there is a positive correlation between attitude/belief and achievement in a statistics course (Carlson & Winquist, 2011; Murtonen et al., 2008; Ramirez et al., 2012; Zimmerman & Austin, 2018). That is, if students have positive attitudes and beliefs towards statistics, then, in general, they re successful in their course. For example, Zimmerman and Austin (2018) found that students' positive attitudes about the worth of statistics had a positive correlation with students' final exam results in both an online and face-to-face introductory statistics course. On the other hand, Williams (2015) found that students with negative attitudes, beliefs and anxiety towards statistics tended to avoid statistics courses and, when they were in them, avoided doing course work. Hagen, Awosoga, Kellett, and Damgaard (2013) found that nursing students who felt that the statistics course had no relevance to their career were resentful at the beginning of the term that they had to take the course.

Some studies also examine the role technology plays on students' beliefs about the usefulness of statistics. Garfield et al. (2012) found that students in a course that used randomization tests and bootstrapping perceived the usefulness of statistics. Swanson, VanderStoep and Tintle (2014) found different results. Their study examined students' attitudes in a randomization class versus a traditional class. They found that students' attitudes towards the usefulness decreased in the randomization course. Thus, from these two studies, it is unclear if using technology can improve students' beliefs about the usefulness of statistics.

Other studies have examined how context-rich problems impact students' beliefs about the usefulness of statistics. Hiedemann and Jones (2010) found that using instructional case studies and service learning projects yielded favourable views of the applications and necessity of statistics. The results were more significant in the service learning course. Dierdorp et al. (2011) found that authentic problems engaged students in the material, helped students see the usefulness of regression, and helped students understand how practitioners in multiple professions would use regression. Neumann et al. (2013) found that the majority of students found the use of real-life data provided examples of the real-life usefulness to the learning of statistics. About a third of the students' commented that the use of real-life data provided motivation to learn and engaged the students into looking deeper into the problems and concepts.

D'Andrea and Waters' (2002) study has the most relevance to the current study as it examined how the use of short stories impacted students' statistical anxiety and attitudes about the usefulness of statistics. The authors wrote a series of short stories that involved solving crimes. The students were asked to solve the crime by following instructions provided within the story. For example, one story had the students investigate blood alcohol levels to determine if a driver was drunk at the time of a car accident (p. 2). The researchers found that, though levels of anxiety decreased, students still had difficulty believing in the relevance of statistics to their everyday lives. As the students only read the stories and did not reflect on the uses of statistics presented, this study provided evidence that students simply being presented with the uses of statistics is not sufficient to cause their beliefs about the usefulness of statistics to change.

In summary, though there are few studies on students' beliefs about the usefulness of statistics, the research does appear to suggest that taking a statistics course does not necessarily improve students' beliefs about the usefulness of statistics. In fact, statistics courses can have a negative impact on their beliefs. This is discouraging as other studies have found a positive correlation between students' beliefs about the usefulness of statistics and their achievement in a course. From this review, there is only one study (D'Andrea & Waters, 2002) that examined the impact of stories on students' beliefs about the usefulness of statistics. This suggests that further research into interventions (including interventions that involve stories) that may impact students' beliefs about the usefulness of statistics is warranted.

15. Research on Learning Through Stories in Statistics/Mathematics Education

From the review of the literature, I found only three studies that examine students' learning statistics through stories: Smith (2014), Blackburn (2016), and Sherwood (2018). The latter two involved undergraduate students. Though D'Andrea and Waters' (2002) study involved stories, it did not examine students' learning.

Blackburn's (2016) study is particularly relevant as his study looked at an undergraduate business statistics course, which is the same context as this study. In Blackburn's study, he created a story about a fictionalized fish farm in an effort to contextualize concepts like the sampling distribution for proportion in a business setting. The stories have characters such as "Freaky Fish" who explained statistical concepts (p. 3). To present the story, Blackburn used an eLearning platform. To investigate the impact of the intervention, Blackburn compared the results of students on a multiple-choice test at the beginning of term versus the end of term. The test at the end of term also included open-ended questions about their experiences in the course. From the results, Blackburn found that students reported that they were engaged in the learning through the intervention and believed that they would retain the material longer than if traditionally taught. But based on the quantitative analysis of the pre- and post-tests, Blackburn did not find that learning through a story resulted in better performance on the post-tests.

Sherwood (2018) examined the use of storytelling by undergraduate students in an economics program to develop their understanding of the normal and sampling distributions. In Sherwood's study, students wrote two stories: one about the normal distribution and the other about sampling distributions. Both stories had to use the same student-created context, reflect the key features of the topic, and have characters (p. 3). After completing the stories, students were interviewed about their experiences writing the stories. Sherwood found that students went through three stages in writing the stories. In the first stage, students struggled with "breaking down the abstract [concepts] into something simple" (p. 5). In the next stage, they realized their lack of conceptual understanding and their focus on surface learning. Finally, in the last stage, they began to make personal connections to the abstract concepts and develop a deeper understanding of the concepts. Sherwood argued that through the act of storytelling, "students' higher order thinking skill of creativity can be activated" (p. 5).

Smith (2014) developed a storytelling-questioning approach in an effort to aid young children in making informal statistical inferences from stories. Smith found that students made informal inferences using the realistic stories and the questioning technique: "The stories did what traditional mathematical word problems could not; that is, elicit a wide variety of data use across a variety of supporting rationales" (p. 112). Smith found that a feature of the story-telling questions that led students to engage in informal inferential reasoning and statistical thinking was when the prompts specifically asked students to do so. This result in supported by research done by Elia, van den Heuvel-Panhuizen, and Georgiou (2010), who found that, though picture books have the capability to elicit mathematical thinking, often appropriate prompting is needed to generate learning. Additionally, Smith (2014) also found that, unlike with realistic stories, the use of imaginative fantasy stories was not always successful. The students' informal inferential reasoning relied less on appropriate inferences than they did in the realistic stories. For example, in the stories with magic, inferences based on magic were used rather than data. This suggests that, at least for young children, the nature of the reality can influence what is deemed to be acceptable reasoning.

While other studies on use of stories in statistics education were not found in the review, there are studies that investigated the use of stories for other mathematics topics. For example, Albano and Pierri (2016) investigated how stories could promote mathematical literacy as active problem-solvers in real-world settings. The intervention that they researched presented students with the story of a journalist who is presented with the problem of graphs being misrepresented in an election. Each of six situations in the story involved the student acting out a part of the story. For example, in the fifth situation, the students communicated the results of their interpretation of graphs in the form of a journalistic piece. The pilot project for the intervention suggested that students' knowledge about different graphical representations increased, but there was no significant change in the answers on PISA style questions from before to after the intervention. All of the twenty-seven students in the pilot found the story intervention to be at least moderately engaging.

Some studies focus on young children. For example, Anderson et al. (2004) found that there were multiple benefits observed from using picture books as stories to learn mathematics. They found that children who read these picture books demonstrated the ability to filter extraneous information and use information over multiple pages to answer mathematical questions that arose in the reading of the text. As another example, when students read the books multiple times, they had the opportunity to explore and build on their initial notions in a natural way (van den Heuvel-Panhuizen et al., 2009). Casey et al. (2008) found that learning increased for all students in a culturally diverse classroom when picture books were used. In the case of Dubon and Shafer's (2010) study, oral storytelling was used to add meaning to the process of making patterns with snap cubes for children in kindergarten. The researchers found that "unlike the traditional approach to pattern instruction, storyboards enable students to use cubes to represent patterns in their stories, to use representations to guide their storytelling, and to create patterns by attaching meaning to the pattern elements" (p. 328).

Other studies emphasize the use of fantasy stories and imagination. For example, Nicol and Crespo (2005) used an abbreviated version of Abbot's *Flatland* novel, which is about a square living in a two-dimensional world that travels to the one-dimensional and three-dimensional worlds, to explore space and shape. Ranatunga et al. (2014) used Australian Aboriginal creation stories to explore mathematics. Capraro and Capraro (2006) used the fantasy book series of Sir Cumference to explore geometry. Using imaginative stories improves students' engagement in the material (Nicol & Crespo, 2005; Ranatunga et al., 2014). Students' achievement on tests improves using stories (Capraro & Capraro, 2006). Students pose questions amongst themselves about the stories that arises organically out of curiosity and wonder, which suggests intellectual engagement (Nicol & Crespo, 2005).

Balakrishnan (2008) also used multiple fantasy stories to explore multiple topics in mathematics for students in grades 8-10. For example, in Skull Island, students were presented with the story of Bloodhead the pirate who hid a cursed treasure on Skull Island. The students used clues to find the location of the treasure on a map. The clues used the Cartesian plane and linear equations. Balakrishnan found that the stories helped in the building of mathematical understanding. For example, concepts that were seen as abstract and foreign, became approachable and they helped the students see the big ideas of mathematics. Further, in

Balakrishnan's (2008) study, students were asked to write their own stories. The study found that this allowed the students to begin the expansion of the concept. For example, in a story about exponents, students changed the base from the original story they heard. This was done naturally and not through prompting. Dubon and Shafer (2010) used storyboards to introduce stories to snap cubes. They found that when students created their own stories using the snap cubes, it made it easier for students to discover errors in their patterns.

Finally, the use of cartoons and video games to tell a story is also used in some studies. For example, Cho (2012) had students read cartoons found in newspapers and online to examine how cartoons that explained a concept can affect students' intrinsic motivation, interest and mathematics anxiety. Cho found that when the activities and the students' abilities aligned, motivation and interest increased while anxiety decreased. If they did not align, students became frustrated. The teacher in the study reported that students showed greater perseverance, were more willing to participate, and were more focused during the lessons involving cartoons. One concern about the study was that it was unclear which cartoons were actually used in the intervention. In another study, Cho and colleagues (2015) had pre-service teachers create their own cartoons to represent their mathematical thinking. The cartoons allowed the pre-service teachers not only to explore a mathematical concept, but also how mathematical reasoning can be used in a meaningful way and in a real situation. It was suggested that problem posing through cartoons allowed students to take on a different authority than traditional word problems.

In the case of Giannakos et al. (2012), they investigated how an adventure story video game could improve mathematical skills. They purposefully designed a video game to have a narrative structure. They found that overall students showed significant improvement from pre- to post-video game test on their mathematical skills. In particular, students who did poorly on the pre-video game test showed the most improvement. It was unclear in the study what mathematical skills were incorporated into the game or were tested for.

Summary of students' learning through stories

In general, stories have been used in a variety of ways in research in mathematics education to support and determine the impact on students' learning. They are shown to be effective in increasing engagement (e.g., Nicol & Crespo, 2005) and improving learning (e.g., Dubon & Shafer, 2010). But not all studies produced favourable results. In particular, two of the three studies that focus on stories in statistics did not have entirely favourable results. Blackburn (2016) found no significant change in the learning of statistical concepts. Though Smith (2014) found that realistic stories led to students making appropriate statistical inferences, he also found that fantasy stories led to students making less appropriate inferences than in realistic stories.

Even though Pfannkuch et al. (2010) connected stories and statistics by arguing that part of statistical thinking is unveiling the stories that the data has to tell, I found few studies that used stories to teach statistics. Blackburn (2016) used stories set on a fictional fish farm to have students explore various statistical concepts. Smith (2014) used a storytelling-questioning method as a way to use stories to have young children explore informal inferential reasoning. Sherwood (2018) had students write their own stories to explain the normal and sampling distributions. Though this list is short, it is encouraging to see that two of the three studies were done in a post-secondary environment.

I found only one study that used a somewhat similar format of story used in this study. Albano and Pierri's (2016) intervention provided the students with the characters, context, problem, and beginning of the story that are written by the teacher. The resolution of the problem and end of the story is handled by the student, in Albano and Pierri's case, by completing tasks for the story and, in my case, by writing dialogue within the story. A significant difference between the format of the story is that, in Albano and Pierri's stories, the students respond at the end of the story while, in this study, the students' responses are integrated within the story. This suggests that the story model that I am using is innovative and relatively unique in mathematics and statistics education.

16. Chapter Summary

This chapter began with an overview of the theoretical perspectives related to this study. This included an overview of the nature of statistics and the reform movement in statistics education; constructivism; instrumental and relational understanding; beliefs; and the nature and use of stories. Following the overview of the theoretical perspectives, a review of relevant empirical studies was presented. The review focused on studies on students' learning of statistics by examining students' learning of the topics of descriptive statistics, sampling distributions, and inferential statistics (both informal and formal). Additionally, studies on the role of technology and context-rich problems in students' learning were presented. The review also provided research on students' beliefs on the usefulness of statistics and students' learning through stories in mathematics education.

Statistics education is a wide field that has at least four journals dedicated to the area (Journal of Statistics Education, Statistics Education Research Journal, Teaching Statistics, and Technological Innovations in Statistics Education). In addition to these journals, other mathematics education and statistics journals also publish articles on statistics education (e.g., Mathematical Thinking and Learning, and The American Statistician). Due to the wide array of publications on statistics education, it is not possible for this literature review to cover all areas. For example, students' learning about regression and multi-variate statistics were not discussed.

Instead the focus was on topics that are covered in the intervention. Further, the focus of this literature review was on the learning of statistics, therefore areas such as supporting teacher education in statistics both at the prospective teacher and practicing teacher level were not included. Collaborative learning in statistics, statistical literacy for the general public, assessment in statistics, and curriculum assessment are some areas that were not covered in this literature review but do show up in the literature.

From this review of the literature, it can be argued that this study adds to the field of statistics education. Most studies on student learning in the review focused on misconceptions. As this study explores students' understanding of statistics through the intervention, this study provides a different focus than most studies on understanding in statistics. That is, this study focuses on how students understood the concepts rather than on their difficulties and misconceptions. Additionally, Roberts and Stylianides (2013) stated that research on how narratives can be used to promote learning and mathematical thinking is in its infancy (p. 454). Based on my review of the literature on the use of stories in statistics education, I would also apply this statement to statistics education. Thus, this study also adds to the limited current research on how stories impact the learning of statistics. Finally, most studies in statistics education covered in this review are quantitative and, thus, rarely mention the theoretical perspective on learning guiding their work. This suggests that there is room for research that is both qualitative and done within a specific theoretical perspective on learning. Thus, this study adds to the literature by being a qualitative study and related to the learning theory of constructivism.

Chapter 3 - Methodology

This chapter presents the methodology for addressing the research questions in this study. This includes descriptions of case studies, the business statistics course, the story-based intervention, the participants, methods of data collection, the data analysis process, the validation process, and the ethical considerations. The research questions that this study investigated are as follows:

1. In what ways does the intervention impact post-secondary students' understanding of selected topics and the discipline of statistics?
2. In what ways does the intervention impact post-secondary students' beliefs about the usefulness of statistics in their everyday lives?
3. What features of the intervention support meaningful learning for the students?

17. Perspective of Research Methodology - Case Study

This study employed a qualitative case study to investigate the research questions. Although there are multiple definitions of a case study, there is consensus that a case study involves an in-depth description of the case, is done in a real-life context, is a bounded system, and involves multiple methods of data collection (Cohen, Manion, & Morrison, 2011; Merriam, 2009; Stake, 2005; Yin, 2014). Possible examples of a case are a person, a group, a community, an event, or an organization (Merriam, 2009). The case in this study was a "group", that is, a class of post-secondary business students.

Case studies are useful in studying educational innovations (Merriam, 2009), in researching 'how' and 'why' questions, and when the context and the variable being studied are difficult to separate (Yin, 2014). They are useful as they can improve understanding of the case being studied; focus on experiential knowledge, context and activities of the case; and gain credibility by triangulating the description and interpretation (Stake, 2005, pp. 443-4). Thus, a case study was appropriate in this study as it was an exploratory study of an innovative intervention in the teaching and learning of statistics.

There are also different types of cases as suggested by Merriam (2009). She defined the type of case study based on the intent of the researcher: descriptive, interpretive and evaluative. In a descriptive case study, the intent of the study is to describe. It is not to generate theories, hypotheses or generalities. It is useful when exploring a hitherto minimally researched area. In interpretive or analytical case studies, like descriptive studies, the researcher describes the case, but the intent is to then interpret, analyze, and theorize about those descriptions within a theoretical framework (Merriam, 2008). An evaluative case study is also descriptive, but the intent is to explain and judge. In particular, this type of case study can be used to evaluate an intervention, innovation, curriculum or program as in this study.

A qualitative case study is appropriate for this research as it provides a basis to deeply explore the students' understanding and beliefs of statistics with the potential to produce insight into the students' learning of statistics. It is a case study because the focus is on a group of students, in one class, in one term, that is, it is bounded by a specific time (Winter 2017), place (specific class) and activity (participating in this statistics class; Creswell, 1994). It is an appropriate methodology to use for the evaluation of the intervention (Guba & Lincoln, 1981) by employing multiple types of data collection such as class artefacts, written response items, and interviews to increase the rigor and validity of the results.

Possible limitations of case studies

Common issues raised about case studies are that they cannot be generalized and they are prone to bias (Cohen et al., 2011, Flyvbjerg, 2006). Merriam (2009) argued that thinking that values generalizations (and thus dismisses case studies) misses the point of case study research. The goal of a case study is to provide a vivid story that the reader learns from and applies in practice. Further Yin (2014) argued that case studies lend themselves to analytic rather than statistical generalizations.

Regarding bias, Flyvbjerg (2006) argued that all research has a level of bias. For example, if someone is implementing an experimental design, then their choice of what to study, what variables to look at, and who to include in the study are all forms of bias. For a case study, the researcher needs to be aware of their own intentional and unintentional biases, but that is true of all research. My primary bias was my desire to show that the intervention is successful. At the beginning of this study, I believed that the use of stories has the potential to help students learn statistics and to change their belief in the usefulness of statistics. Though this bias has the potential to colour the results of my research, I mitigated this by having multiple data sources and by actively considering how this area of bias may be colouring my interpretations, which made it difficult for the results to confirm by belief.

18. The Business Statistics Course

The course in this study is a multi-section algebra-based business statistics course that is intended for first-year business students at a university in southern Alberta, Canada. The course was chosen for the study because of my experience teaching it for multiple years and finding it challenging to engage students in developing relational understanding of the course content. This led me to consider and try alternative teaching approaches and eventually to design and investigate this intervention in an effort to address these challenges.

The course ran four hours per week during a thirteen-week semester, with two classes per week. It was taught in a computer lab where each student had access to a desktop PC. The section of the course used in this study in the Winter 2017 semester was not taught by myself to satisfy ethics requirements. The instructor was a tenured faculty member with over fifteen years of experience teaching business statistics. He worked with me in piloting an initial version of the story-based tasks, discussed later, thus, he had experience teaching with such tasks and did not require special training for the study.

There were four units in the course: sampling techniques and descriptive statistics; probability; formal statistical inference; and simple linear regression. The story-based tasks were designed only for the first three units. The unit on sampling techniques and descriptive statistics served as an introduction to the course. It covered terminologies (such as sample, population and data types), the importance of a good sample, methods of data collection, and how to summarize that data through visual representations, measures of centre, measures of variation, and measures of location. The unit on probability covered the concept of probability, the rules of probability, probability distributions (specifically the binomial and the normal distributions), informal inferential statistics, sampling distributions, and the central limit theorem. The unit on formal statistical inference covered confidence intervals and hypothesis testing for the one mean (for both small and large samples) and one proportion. Implementation of the story-based tasks to teach some of these topics is described later after discussing the development and nature of these tasks in the next section.

19. The Intervention: Story-based Tasks and Supplemental Tasks

As previously discussed, the reason for the intervention lies in addressing a gap in the literature involving implementation of recommendations of the reform movement in statistics education. My interpretation of these recommendations led me to the idea of using stories as a vehicle for exploring statistical concepts. Stories have the potential to engage students in meeting the learning goals of the reform movement in statistics education by encouraging the development of statistical knowledge, reasoning and thinking.

This section has three parts. First it discusses the development of the intervention (including the pilot study for the intervention), then describes the intervention specifically used in this study by describing the story-based tasks and the supplemental tasks, and finally provides details on how the intervention was implemented in the study.

Development of intervention

This section presents the background on how the intervention and, in particular, the story-based tasks were developed. It begins by outlining how the theoretical perspectives presented in Chapter 2 framed the intervention. Then a description of how the stories were piloted is presented. Finally, the modifications of the intervention from the pilot study to the current study are outlined and explained.

Theoretical perspectives that framed intervention. The theoretical perspectives discussed in Chapter 2 framed the intervention. Based on the recommendations of the reform movement in statistics education and the related emergent perspective of constructivism, the intervention was developed to have a focus on:

1. active learning through the collaborative, interactive nature of the story-based tasks, the writing of dialogue to complete the stories, and self-reflection on learning;
2. individual and group inquiry/explorations of the statistical concepts through the stories, and prior knowledge and experiences;
3. effective use of technology through both the use of an Excel macro (called MegaStat) as a data analysis tool and the use of computer simulations to explore statistical topics;

4. assessments that promote the learning goals of the course and help students learn through out-of-class assignments involving the story-based, reflection and follow-up tasks; and
5. rich, meaningful, and authentic contexts that had the potential to motivate and engage students, and stimulate real-world statistical applications and investigations.

These ways of grounding the intervention also offer students the opportunity to experience the usefulness of statistics in a way that could serve as evidence to support changes to their beliefs about statistics.

In order to develop meaningful stories, I participated in a workshop on writing short stories conducted by a local storyteller, Kate McKenzie. I learned about writing stories in one scene, ensuring dialogue was consistent with the character, and providing more sensory details. Another key idea I learned that really helped in writing stories was the idea of show versus tell; that is, instead of telling the reader what is happening, show the reader what is happening by having details, internal thoughts, and feelings.

Pilot of stories as statistics tasks. My original idea for using stories in my teaching of statistics was to create them for the topics of sampling techniques, descriptive statistics, rules of probability, probability distributions, sampling distributions, confidence intervals, and hypothesis tests. These initial stories were designed to pose a problem for the students to solve and to be a vehicle for motivating and developing statistical concepts. The students were required to respond to the following questions at the end of the story:

1. What is the story the data is telling?
2. What is missing from the data?
3. What is your response?
4. What are the limitations of your response?

For each of the questions, additional sub-questions were provided to guide the students on how to answer these questions for the specific story task.

The stories were implemented in my class and that of a colleague, who had been collaborating with me on the story idea, in the Winter 2016 term. We obtained ethics approval from our institutional ethics board to investigate our use of the stories as a pilot study. Table 3.1 provides an overview of the major stories used in the pilot.

Table 3.1 - *Stories used in pilot*

Story	**Statistical topic**	**Synopsis of problem presented**
Barry's Big Problem	Sampling techniques, and descriptive statistics	Barry is a manager at an oil and gas company and wants to determine a more equitable way to distribute bonuses at his firm. Write a memo that explains his idea and support the rationale with statistical evidence.
Did She Do It? The Trial of Shirley Ingram	Rules of probability	Shirley Ingram is accused of murdering her two baby boys. The prosecution's case hinges on the probabilistic evidence given by a pediatrician. Write a news article that evaluates the evidence provided.
Can Dolphins Communicate?	Informal inferential statistics and probability distributions	A couple argues over whether dolphins can abstractly communicate. To overcome the skeptical position of the boyfriend, the girlfriend uses the evidence of two dolphins correctly communicating 15 out of 16 times. Write an argument that explains why or why not that is sufficient evidence that the dolphins are communicating.
The Dragon Lady	Sampling distributions	A company that produces microprocessors for self-driving cars has a contract that requires that any random sample of 20 processors react, on average, in 18ms or less. Use a bootstrapping technique to find the empirical sampling distribution, to determine whether the company can confidently meet this expectation.

It's Impossible to Score	Confidence intervals	Mike Babcock, a coach in the NHL, claimed that it is impossible to score nowadays. Compare the total number of goals scored per game in the 1985-86 season with this year's goals scored to determine if Mr. Babcock is correct.
Tell your own story	Hypothesis test	Design a quantitative study that answers a question of interest to you. Write a report that answers the question.

From the pilot, two major issues arose. The first issue was that the context of the stories was often lost in the students' responses. This was considered problematic as it suggested that the connection for the students between the meaningful context and the statistical concept might not be occurring. The second issue was that students appeared to be having difficulty connecting the story-based tasks to the rest of the course. That is, students did not appear to be connecting the concepts in the stories with the concepts in the class. This was considered problematic as it suggested a disconnect between the tasks and the learning goals of the course. These two problems suggested that the tasks were not fully aligning with the recommendations of the reform movement in statistics education. To address these problems, we made significant changes to the stories that then became the story-based task for this study.

Modifying original stories. The original piloted stories were modified to address the issues that emerged regarding loss of context and loss of connection.

Two changes were made to address the loss of context. The first change was updating the contexts to be more accessible to students. For example, in the pilot, *The Dragon Lady* was about the manufacturing of hardware for self-driving cars. The context was chosen as self-driving cars are a new and innovative area in business. But the focus of the story was on the speed of a microprocessor, which quickly confused and bored the students. The context was updated to be about the speed of scooters, which was believed to be a more accessible context as students would be more familiar with the speed of scooters instead of microprocessors. By making the contexts more accessible, students may connect more with the stories and, thus, better understand the usefulness of statistics in their everyday lives.

The second change involved the location of the prompts. In the pilot of the story-based tasks, the prompts were at the end of the story, but for this study, the prompts were embedded in the story. That is, the story was left intentionally incomplete and the students were invited, through the prompts, to fill-in the story. By embedding the prompts within the story, it was hoped that the context would not be lost in the students' responses.

Due to the prompts being embedded in the story, the nature of the prompts changed as well. To illustrate, in the pilot, the prompts were very open-ended, which allowed the students more choice in how they could approach the problem. But once the prompts were embedded in the story, they had to be written in a way that the rest of the story continued to make sense regardless of the students' responses. Thus, the prompts became more directed. For example, in the pilot, a prompt that asked students to produce descriptive statistics and to interpret them was written as "What story does the above data tell you about the current distribution of salaries at this firm?" while in the revised version the prompt detailed what type of descriptive statistics to produce. Further, as the prompts were embedded within the story, the students' responses were also embedded within the story. To ensure that the flow of the story was maintained, the students were prompted to write dialogue for the characters. These changes had the potential to maintain connection between the students' responses and the context of the story.

To address the problem of loss of connection between the concepts in the stories and the concepts in the class, two changes were also made. The first change involved focusing the story-based tasks on the major topics in the course and providing students with the opportunity to explore the major topics in more depth. Thus, instead of seven story-based tasks, as used in the pilot, it was decided to use only four in order to better focus on the major topics of the course. Students could then spend more time on these tasks rather than quickly moving from one task to the next every couple of weeks. In addition, two new supplemental tasks were included for each story task: a reflection task and follow-up task. In the reflection task, intended for engaging with the concepts more deeply, students were asked to consider what they understood about the concept, what they did not understand, and how the statistical concepts could be used in their everyday lives. In the follow-up tasks, set up more like a traditional assessment, students had the opportunity to explore the topic in more depth by applying their learning in a new and different context.

The second change involved considering how best to connect the stories and the course content. The result was to create two types of stories. The first type, a *comprehensive story*, was used for statistical topics that were comprised of multiple statistical concepts. For example, the topic of sampling techniques and descriptive statistics involves multiple concepts, which includes various statistical measures such as mean, standard deviation and box plots. Thus, comprehensive stories were used at the end of the unit on the topic to provide the students the opportunity to consider how the separate concepts could be used together to address one problem. The second type, an *introductory story*, was used for statistical topics that covered one significant and complex statistical concept. For example, the topic of sampling distributions of sample means is abstract and involves multiple pieces that build on each other. Thus, introductory stories were used at the beginning of the topic to motivate the learning of the concepts and to provide a concrete reason for the learning of the abstract topic. These changes impacted the role of the characters in the story. In particular, as one character provided explanations of the statistical concepts or guided the process of the statistical investigation, each story required an expert character who was well-versed in statistics and who could make suggestions of what to do next. As a foil to this character, a novice character was also included who could ask questions of the expert and who could voice common misconceptions of the statistical concepts. These changes had the potential to help students see a closer connection between the course content and the story-based tasks.

In summary, the intervention was developed to have stories that engaged students in applying their learning of statistical concepts in a meaningful context. From the pilot of the studies, it was revealed that the stories did not entirely engage students as intended. To address these issues, changes were made to the intervention which included updating the contexts to be more accessible for the students, embedding the prompts within the story, having students write dialogue for two types of characters (expert and novice), reducing the number of stories, adding a reflection and follow-up task, having the stories bring together multiple concepts, and having explanations of concepts embedded in the story.

Intervention for the study

This section presents the specific story-based tasks, following modification, and supplemental tasks used in this study. It begins with a general description of the story-based tasks and is followed by each of the four stories presented in detail. Finally, the supplemental tasks are outlined.

Overview of the story-based tasks. The four stories in the intervention used in this study were pre-dominantly written by myself in conjunction with the instructor for the course. In particular, we wrote the context, problem, characters, and the plot of the story. The stories were written as short stories and were around 10-12 pages long. Each story focused on one major statistical topic and had a unique context in which the students explored the topic through the story. As the course in which the intervention was implemented was a business statistics course, the contexts were business related. The stories were fictional, but set in realistic situations.

Table 3.2 summarizes each of the stories by outlining the statistical problem for each story, the statistical data provided for the story, the key statistical concept covered in the story, and the names of the main characters.

Both comprehensive and introductory stories, as previously described, were used. Each story had a problem that would be resolved through some form of statistical analysis. The stories were left intentionally incomplete. Within the story, the students were prompted to write dialogue between the characters. The students could choose to write their dialogue as one would normally see dialogue in a story or they could write it like a script. There were between 12 and 14 prompts per story. The prompts had the students produce and interpret statistical measures, draw conclusions from the statistical analysis, explain their reasoning for why they choose specific statistical measures, and explain aspects of the statistical concepts covered in the stories. Thus, the students did not simply passively read the stories but were invited to actively engage with the story. As such, the stories were written both as a teaching tool (teacher as storyteller) and as a cognitive tool (student as storyteller/narrator; Roberts & Stylianides, 2013). The portion of the stories where students wrote dialogue for the characters are the story-based tasks. Each story also had an expert and novice character. That is, there was at least one character who was well-versed in statistics and at least one character who was new to the statistical topic presented in the story.

Table 3.2 - *Summary of stories used in intervention*

	Problem presented in the story	**Statistical data in the story**	**Key statistical concept(s)**	**Expert character(s)** **Novice character**
Bob's Bikes	Determining if an inventory system is undervaluing items by on average more than $12	Difference between inventoried price and actual price	Sampling techniques and descriptive statistics	Jolene, Franca Bart
Can Dolphins Communicate?	Whether the dolphin Aries can understand an oral communication from the dolphin Daphne	Number of times Aries got the right answer	Informal inferential statistics (with a focus on informal hypothesis testing) and the binomial distribution	Emily Sam
The Dragon Lady	Determining if electric scooters are meeting the contractual obligations	1) Peak speed of individual scooters. 2) Mean peak speed of 30 scooters	Sampling distributions of sample means including the central limit theorem and normal distribution	Reema Jed
Can They DIG It?	Determining if company should go forward with business expansion plan	1) Percentage of waste that is recyclable, 2) Percentage of waste that is compostable, 3) Number of buildings with more than 15% waste that is recyclable, 4) Number of buildings with more than 15% of waste that is compostable.	Hypothesis testing and confidence intervals for mean and proportion	Kate and student Leor

The students did not have complete say in how the stories evolved. As the portions where students wrote dialogue were embedded within the story, the stories had to be structured so that they still made sense even after the students wrote their portion. Thus, the prompts were written to direct the students to consider specific things. For example, in *Bob’s Bikes*, the students were prompted to consider visual descriptive measures, measures of centre, and measures of variation and, thus, could not choose their own way to explore the problem, but they did have choice in which measures to use in their final analysis. Thus, the students had some control over the resolution of the story, but not complete control.

Though the story-based tasks covered many of the topics covered in a first-year algebra-based statistics course, they did not cover all topics in the course. Approximately 85% of the course topics were covered in the story-based tasks. The topics that were not covered by the story-based tasks were probability rules, independent events, sampling distribution of sample proportions, and simple linear regression. All of these topics were minimally covered in the course and as such were not included in the stories. The following sections provide details about each of the four stories included in the study. They are presented in the same order as they were presented in the class.

Bob’s Bikes – Sampling techniques and descriptive statistics. The story of *Bob’s Bikes* is that of three accounting articling students, Jolene, Franca and Bart, who are asked to determine whether an inventory system should be repaired immediately or if the cost could be deferred to a later date. In writing the story, the criteria for investigating the inventory system was determined in consultation with faculty members in accounting. In the story, the characters’ boss promises them the remainder of the week off if they can determine whether the inventory system requires immediate repair or not. This means that the fewer days they spend at the store assessing the inventory system, the more days they have off. Thus, they want to use a sample of items from the store rather than the whole store. The first part of the story has the characters describing sampling techniques, which they then implement. At the end of the day, each of the three characters has collected their own sample. Each character then makes an argument as to why they should choose a specific sample or group of samples. The students then had to choose which of the three arguments they think is best and look at that sample. Once they had their sample, students found various descriptive statistics that they interpreted and then considered as a whole to determine whether the system should be repaired or not. Depending on the choice for the sample, they arrived at a different conclusion (e.g., the system needs immediate repair, the system can be repaired later, or the data is inconclusive) and the story was slightly different. In responding to the story, the students explored sampling methods, finding and interpreting descriptive statistics, and coming up with a conclusion based on the sample. *Bob's Bikes* is a comprehensive story that was given to the students at the end of the unit on sampling techniques and descriptive statistics. Jolene and Franca are the expert characters, while Bart is the novice.

Can Dolphins Communicate? – Informal inferential statistics and binomial distribution. The students explored informal inferential statistics and the binomial distribution using the story *Can Dolphins Communicate?* The premise of the problem presented in the story arose from an example presented in Tintle and colleagues (2015) textbook *Introduction to statistical investigations*. The type of informal inferential statistics that is focused on is informal hypothesis testing and utilizes a randomization test. The story involves the head of marketing, Sam, wanting to change the dolphin show at an aquarium. He approaches a scientist, Emily, who suggests an exhibit that demonstrates that dolphins can communicate abstractly. Sam is skeptical that dolphins can communicate. Emily uses an experiment where two dolphins need to communicate to successfully receive food to demonstrate evidence of abstract communication. The trial is repeated twenty times and the dolphins successfully communicate seventeen of those times. The story had the students explore what is sufficient evidence to support the claim. To do this, they repeated the experiment using a simulation many, many times, under the assumption the dolphins were not communicating, and determined the experimental probability that the dolphins had successfully communicated at least seventeen times. Thus, the students applied their understanding of probability, were introduced to the binomial distribution, and used informal inferential statistics to make a decision. *Can Dolphins Communicate?* is an introductory story used in the unit on probability. Emily is the expert character and Sam is the novice.

The Dragon Lady - Sampling distributions. In this story, a company has secured an investment from the reality TV show *Dragon's Den* for their solar powered scooters, but that investment is contingent on a consultant coming in to find ways of cutting costs. The 'dragon lady' Reema finds that the company is spending a lot of money testing individual scooters as part of quality control, but the actual requirements of their contract is that the *mean* peak speed a batch of scooters at 50 kmph give or take 2 kmph. This leads the consultant to suggest looking at the sampling distribution of sample means. The students used the sampling distribution of sample means to determine the probability that the mean peak speed of the scooters met the contractual obligations. This story explores normal distributions and sampling distributions of samples means, including the central limit theorem. *The Dragon Lady* is an introductory story that is used to introduce the topics of sampling distribution and the central limit theorem in the unit on probability. The expert character is Reema and the novice character is Jed. This story is provided in its entirety in Appendix A.

Can They DIG It? - Formal inferential statistics. In this story, students were introduced to a problem being experienced by a real-life company called Do It Green or DIG based in Calgary. DIG provides environmentally friendly waste management solutions to events like the Calgary Stampede. *Can They DIG It?* was written with the guidance of the co-founder of DIG, Leor Rotchild. DIG's business model at the time of writing the story was seasonal, as most of their work came from summer festivals. They were looking at expanding their business by providing their waste management strategies to buildings. The story is about a statistical consultancy firm that has been hired to determine whether there would be enough buildings in Calgary that would benefit from DIG offering their services. Thus, the goal was to determine if the expansion of their business would be viable. The story was used to explore confidence intervals and hypothesis testing in the unit on formal inferential statistics. Unlike the other story-based tasks, the response to the original story was written as a report and the follow-up task involved students writing dialogue. The expert characters are the student and Kate, and the novice character is Leor. It is a comprehensive story.

Supplemental tasks. In addition to the four story-based tasks, the intervention consisted of two additional tasks: a reflection task and a follow-up task. Students were assigned these tasks after they completed each story-based task. In the *reflection task*, students were asked to choose one concept that they understood and to present their dialogue to the instructor; to choose dialogue for their prompt that they did not understand and to revise their dialogue; and to consider how the concepts covered in the story-based task could be used in their own lives or careers. Solutions were provided to the story-based tasks to help students determine what they did and did not understand for the first two prompts. The prompts for the reflection tasks are included in Appendix A.

The *follow-up tasks* provided the students with the opportunity to apply their understanding of the statistical topics covered in the story in a different context. Similar to the story-based task, a context and a problem were presented to the students. Unlike the story-based tasks, the follow-up task did not require the writing of dialogue but rather involved a series of prompts that guided the students through the process of solving the problem and explaining key statistical concepts. The exception to this is the fourth story-based task, *Can they DIG it?*, which had dialogue written in the follow-up task. As an example, the follow-up task to *The Dragon Lady* is provided in Appendix A.

Implementation of intervention

As previously mentioned, the intervention was implemented in one section of a business statistics course in the Winter 2017 term. This section provides an outline of how the intervention was implemented in this section of the course. The details of the implementation was determined through discussions and feedback from the instructor for the course and classroom observations, as consented by him.

The instructor posted he story-based tasks on the course website as a Word document. The key statistical concepts covered in the stories were provided on the first page of each task, followed by instructions and an outline of how the students would be assessed. The students were asked to read the first portion of the story-based tasks up to the first prompt prior to the class.

The course was taught in a computer lab where each student had a desktop PC to work on with access to the story-based task and, when provided, the Excel sheet with data. This allowed students to write their responses directly in the story-based tasks and to use Excel and the macro MegaStat to produce appropriate statistical measures when prompted. To differentiate between students' written dialogue and the story, students wrote their answers in a different color from black.

For each of the four story-based tasks, the instructor would begin the class by providing discussion items for the students to frame their reading of the stories. For example:

1. Describe the characters in the story.
2. Describe the central conflict/problem.
3. Based on your answer to question 2) above, write your research question.

The students worked in small groups to come up with answers, which was followed by a whole-class discussion to arrive at consensus about the problem presented in the story, the characters, and the research question. Through the discussion, the instructor would establish the roles the two types of characters should play in the dialogue. For example, after establishing that Jed was the novice character in *The Dragon Lady*, he reminded the students that they are like Jed – they are also learning the material and to use Jed to voice their own questions when writing dialogue.

Students then worked on the story-based tasks at their own pace. They were given the remainder of the two-hour class and a portion of the next two-hour class to work on the task. They mostly worked independently while the instructor provided support by walking around and answering questions. If the instructor felt that multiple students were having the same problem, he would address the problem with the whole-class. For the most part, the instructor did not provide direct instruction for the story-based tasks. The story-based tasks were due shortly after the second class. The reflection and follow-up tasks were completed outside of class time and after the story-based task was due.

For three of the four story-based tasks, students could work in groups, but for the *Can*

Dolphins Communicate? story, students were asked to submit their tasks individually. Students chose their own groups of two or three, but some students chose to work individually although the instructor often promoted group work.

For the story-based tasks, the students were assessed on whether they completed dialogue for all prompts and whether they engaged in writing dialogue between the characters. That is, they were not assessed on the correctness of their answers, but rather on their level of engagement with the task. After the deadline for submission of the story-based task, detailed dialogue for a possible solution to the story were posted, which included common errors and alternative solutions. For the reflection and follow-up tasks, students were assessed on the correctness and completeness of their answers.

20. Participants

The participants were 20 of the students enrolled in the business statistics course. The majority of students were in the first-year of their business programs with a few being in their second- and third-year. The course is a pre-requisite for other courses required for some degrees. This is the only mathematics course required for the business degree and the only pre-requisite for it is grade 12 mathematics. Thus, the majority of students did not have prior experience with the majority of content covered in the course.

At the beginning of the semester, I recruited participants for the study; first by posting the consent form on the course website, then, three days later, going to the class and inviting students to participate. Of the forty students registered in the course section at the beginning of the term, twenty-one students agreed to participate in the study. Data were collected only from students who signed a consent form. No participant withdrew from the study but one did not submit any data. Therefore, data was collected from only twenty participants. Based on ethics requirement, participants had the choice of allowing access to their course work, submitting answers to pre- and post-intervention written response items, and/or participating in an interview. Table 3.3 provides a breakdown of the number of participants who agreed to do specific aspects of the study, and the number of participants who actually completed the aspect they consented to do. For the post-intervention written response items, of the six participants who completed it, five also answered the pre-written response items.

Table 3.3 - *Overview of student participation in study*

Instrument	**# of participants who agreed**	**# of participants who participated**
Pre-written response items	14	10
Post-intervention written response items	14	6
Interview	3	2
Student work	20	19

21. Data Collection

Merriam (2001), Stake (1995) and Yin (2014) suggested multiple and extensive sources of data for a case study. Common sources of data in case studies include interviews, observations, and documents. In this study, the focus was on written documents and interviews, which provided multiple sources of data. Specifically, these sources were various class artefacts, pre- and post-intervention written response items, and post-intervention student interviews. The class artefacts were collected both from the instructor and the student-participants throughout the term. Participants responded to the written response items both before and after they were exposed to the intervention. The interviews were conducted at the end of the course.

Table 3.4 outlines the timing of the data collection and which research questions the data predominantly addressed.

Table 3.4 - *Summary of research instruments*

Data collection instrument		**Time given or collected**	**Research question addressed**
Class artefacts	Student produced material	Throughout course	Participants' understanding and beliefs, and features of intervention (question 1, 2 & 3)
	Instructor produced material	Throughout course	Participants' understanding (question 1)
Pre-intervention written response items		Within first month of the course	Participants' beliefs (question 2)
Post-intervention written response items		Within the last week of the course	Participants' beliefs and features of intervention (question 2 & 3)
Student interviews		After the last day of classes	Participants' beliefs and features of intervention (question 2 & 3)

All of the audio-recorded data were transcribed by a professional transcriber. All of the data were made anonymous by assigning numbers to each participant or group of participants. Minor edits were made to the text of participants' written response items and story-based responses for clarity. For example, if an abbreviation was used for the name of the character, it was changed to the full name. Further, obvious spelling errors (e.g., trail instead of trial) were corrected. These changes did not change the meaning of the participants' work.

Each source of data in Table 3.4 is described next. In particular, details on the collection of the artefacts (both student and instructor) are presented. Then, the written response items (both pre and post) and interviews are presented by first describing the development of these instruments and then describing how these instruments were used to collect data.

Artefacts collected from students and instructor

Student artefacts. Class artefacts were collected from the student-participants to gather evidence related the first and second research questions regarding their understanding of statistics and the discipline of statistics, and their beliefs about the usefulness of statistics they demonstrated in their responses to the intervention. This evidence was also used to investigate what features of the intervention supported meaningful learning (the third research question). These artefacts consisted of their written work submitted for all of the tasks related to the intervention, that is, their written dialogues as responses to the story-based tasks, and their responses to the prompts for the reflection and follow-up tasks. These were collected shortly after the due date for each task. As some tasks were done in groups, only submissions where all group members agreed to be part of the study were collected. Though some participants stayed in the same group throughout the term, other participants changed groups. Table 3.5 provides a summary of how many submissions were collected for each task and the number of submissions per task since not all participants submitted all three tasks.

Table 3.5 - *Details regarding the number of submissions included in the study per assignment*

Title of story-based task	**Story-based task (Part 1)**	**Reflection task (Part 2)**	**Follow-up task (Part 3)**	**Details**
Bob's Bikes	6 groups	6 groups	6 groups	All groups of 2
Can Dolphins Communicate?	19 participants	15 participants	15 participants	Done individually
The Dragon Lady	8 groups	8 groups	8 groups	Three groups of 1 and five groups of 2
Can They DIG It?	7 groups	8 groups	7 groups	Three groups of 1 (with one additional group of 1 for the reflection task), two groups of 2, two groups of 3

Instructor artefacts. Class artefacts collected from the instructor provided information about the course and other resources he provided students in addition to the intervention. This information was useful to offer additional factors that may have impacted the understanding demonstrated by the participants in the intervention. The artefacts consisted of the course outline, the textbook for the course, handouts and course notes, assessments, and any document posted on the course website, which were obtained at the end of the course. The instructor was asked for any personal lesson plans or notes but he did not have any.

Data collection instruments development

In this section, I describe the two data collection instruments developed for this study: the written response items and the interview protocol.

***Written response items*.** The pre- and post-intervention written response items (see Appendix B) were developed to collect data related to the second and the third research questions. In particular, they were to inquire into participants' beliefs about the usefulness of statistics both prior to and after being exposed to the intervention. The post-intervention written response item also inquired about participants' experiences learning through the stories.

As beliefs can be inferred by what individuals say and do (Pajares, 1992), open-ended questions were chosen for the written response item. These questions also allowed participants to provide more information and insights than close-ended questions regarding their beliefs and experiences with the stories. The pre-intervention written response item consisted of three questions that asked participants about their initial beliefs regarding the usefulness of statistics and to provide examples about how they have used statistics in their lives. The post-intervention response item consisted of seven questions, four dealing with their beliefs (three of which were the same as the pre-intervention written response item questions) and three addressing their experiences learning from the stories. Table 3.6 summarizes the focus of the seven questions.

Table 3.6 - *Summary of questions on written response items*

Question Number	**Question subject**	**Pre/Post**
1	Beliefs about usefulness of statistics	Pre and post
2	Examples of how statistics is used in their lives	Pre and post
3	Examples of how statistics is used in their future careers	Pre and post
4	How the stories impacted their beliefs	Post only
5	Most memorable story	Post only
6	Positive impact of stories on learning	Post only
7	Negative impact of stories on learning	Post only

The questions were given to three experts in the field of mathematics and statistics education to review and provide feedback. Some concerns were raised regarding the clarity of the questions and whether the questions would elicit responses about students' beliefs regarding the usefulness of statistics and the impact of stories on those beliefs and their learning. Based on the responses, I revised the questions accordingly. For example, for a pre-written response item, the original wording included "Please provide an example of how you have either recently used statistics in your everyday life or at work." It was suggested that this question may be too narrow as the participant may not have direct experience with using statistics, but they may know someone who has. Thus, the question was revised to "Please provide an example of how you or someone else has either recently used statistics in everyday life or at work." As there were neither substantial changes suggested for the questions nor were there contradictions between the suggestions from different experts, they were only asked to provide feedback once.

Interview protocol. The interviews were used to collect data related to the second and the third research questions, that is, participants' beliefs about the usefulness of statistics, their beliefs about how the intervention impacted their learning, and their experiences with the intervention. A semi-structured approach was used to provide the opportunity to delve more deeply into participants' experiences or beliefs; that is, general questions forming the interview protocol were developed for what I wanted to explore, but the wording, timing and inclusion of the questions was determined as the interview unfolded (Merriam, 2001).

To develop the interview protocol, I started with the questions used in the pilot study, and modified and added to them to address limitations that arose in capturing the participants' thinking. For example, the pilot study participants' responses were often short and did not provide insight regarding their learning or beliefs. To illustrate, for the question from the pilot "Can you give me an example of how what you've learned in this course has been or could be useful in your everyday life?", I added the following follow-up questions aimed at eliciting more details:

- How has your view of statistics changed during the course?
- The stories were written to provide examples of how statistics is used. How, if at all, did they change your beliefs about the usefulness of statistics?

I also utilized the theoretical perspectives of this study to add to the questions asked. For example, as beliefs about the usefulness of statistics are understood to be participants' viewpoints or opinions, I added additional questions that asked about their views and opinions (e.g., "How do you think statistics will help you in your future career?" and "How do you think statistics helps businesses?"). The complete interview protocol is provided in Appendix B.

Administering data collection instruments

Written response items. The participants were invited to respond to the pre-intervention written response item the fourth week of the term (prior to being exposed to the intervention) and to the post-intervention written response item in the last two weeks of the term. For convenience and to maintain confidentiality among participants, the written response items were provided to the participants electronically using Google Forms. A link to the forms was sent by email to the participants who agreed to respond to these items with instructions to include their names to allow for comparison of the pre- and post-intervention responses. The responses were downloaded at the end of the term.

Of the 14 participants who agreed to respond to the written response items, ten submitted responses to the pre-intervention written response item and six submitted responses to the post-intervention written response item. In total, only five participants submitted both the pre- and post-intervention written response items.

***Conducting interviews*.** I conducted the interviews with individual participants after the last day of classes in an office at the university where the study was conducted. Three participants had agreed to be interviewed, but only two followed through and set up a time for the interview that suited them. Each interview took about 45 minutes and was audio-recorded and later transcribed.

The interview began by asking the participants some basic questions, not on the interview protocol, about themselves and the course. Based on their responses, I then asked them a question from the interview protocol in an order that made sense based on the flow of the interview. During the interview, I also asked follow-up questions to clarify their responses or to delve deeper into the response.

As part of a respondent validation process, the interview participants were sent a transcript of their interview and asked if there were any comments they made that they wished to clarify or modify. Neither participant responded with any changes.

Classroom observation

While classroom observation was relevant to this study and was initially planned as another source of data, constraints beyond my control made it impractical for it to be carried out. The purpose of the classroom observations was to gain insight into how the instructor's actions impacted the implementation of the stories and how the students engaged with the story-based tasks. Through the ethics approval process, concerns were raised about the potential to inadvertently observe students who did not agree to be participants and about efforts to avoid observing non-participating students resulting in loss of confidentiality of participants by focusing only on them in class. Due to this, it was decided to observe only the instructor when he was engaging with the whole class. This, however, did not provide any information regarding how the instructor's interaction with students impacted the intervention since most of the relevant interactions were with individual or small groups of students, which were not allowed to be observed. The idea of creating and observing groups of participants working on the story-based tasks out of class after the end of the course was also abandoned partly because of lack of interest by the participants and, more importantly, because this would have changed the context of the study from being naturalistic (i.e., a real classroom setting) to being experimental (i.e., a simulated learning context that lacked features of the real classroom). However, the field notes collected from observing the classes did provide useful information regarding the implementation of the intervention on an instrumental level as described above in the implementation section.

22. Data Analysis

Creswell (2013) outlined a general spiral process for qualitative data analysis, which included organizing the data; reading and memoing; describing, classifying and interpreting data into codes and themes; and representing and visualizing the data. The process of analyzing data qualitatively is “recursive and dynamic” (Merriam, 2009, p. 169). Specifically, for a qualitative case study, Stake (1995) envisioned the analysis as a process of classifying and interpreting the data to make the unfamiliar familiar by revisiting the data until the meaning becomes clear. This can be done by *direct interpretation* (looking at a single instance) and *categorical aggregation* (looking at multiple instances and interpreting them as a whole). Stake also argued for *naturalistic generalization* in case studies where the generalization is not necessarily about all similar cases, but rather that the readers of the case can learn from it. In this study, the focus was on establishing patterns or themes regarding how the intervention impacted participants’ understanding of statistics and their beliefs about the usefulness of statistics, and the features of the intervention that supported meaningful learning.

The general approach I used in analyzing the data involved focusing on one research question at a time, reading through a relevant data source and making notes (or memos) on what I read. Table 3.7 outlines the data sources used for each research question. I then examined the notes for categorical aggregations (Stake, 1995) in an effort to generate codes. Once codes were established, I re-read the data sources and coded the data. When the analysis of specific data sources was complete, additional analysis was done to look across all relevant data sources to establish themes (i.e., patterns and correspondence, as per Stake, 1995). I then re-read the data sources to determine whether the themes I established were supported by multiple data sources.

Table 3.7 - *Sources of data for research questions*

Research question	Sources of data
In what ways does the intervention impact post-secondary students' understanding of selected topics and the discipline of statistics?	Class artefacts
In what ways does the intervention impact post-secondary students' beliefs about the usefulness of statistics in their everyday lives?	Pre- and post-intervention written response items, transcripts of interviews, class artefacts (student responses to reflection tasks)
What features of the intervention support meaningful learning for the students?	Class artefacts (student responses to story-based tasks), transcripts of interviews, post-intervention written response items

As was indicated in Table 3.5, some of the participants' responses to the story-based tasks were done as a group while others were done individually. During the analysis, I did not distinguish between responses from groups versus from individuals. Thus, throughout the description of analysis, the term 'participants' response' is used to refer either to a response from a group or from an individual. While there are limitations to this, discussed later, the study was not about individual students as individual cases, but about their collective learning and thinking, which is a combination of their individual and group experiences and thinking. From this perspective, this treatment of the data is appropriate for this study. Further details of the analysis process are provided next organized by categories for the research questions.

Participants' understanding of selected statistics topics

Regarding research question 1, this section describes the process for choosing the selected topics and the data analysis process for determining participants' understanding of the selected topics.

Table 3.8 provides the *selected statistics topics* chosen as the basis to determine the participants' understanding as a result of the intervention. These topics were chosen by a review of the story-based tasks and the course outline (i.e., the instructor class artefacts). This review was guided by my knowledge of the course, the learning goals of the reform movement in statistics education and the literature review. I chose one statistical topic per story-based task that covered multiple aspects of the course and included the majority of the core concepts covered in the course. Once the topics were chosen, I examined both the story-based and related follow-up tasks to identify what the students were asked to do in relation to these topics.

Table 3.8 – *Selected statistics topics for the study and related concepts*

Story-based task	**Topics covered in story**	***Selected topic* and related concepts covered in story-based tasks**
Bob's Bikes	Sampling techniques and descriptive statistics	*Descriptive statistics* • Visual descriptive statistics including histograms and box plots. • Measures of centre including mean, median and mode • Measures of variation including range, standard deviation, interquartile range, and coefficient of variation • Outliers • Conclusion made based on descriptive statistics
Can Dolphins Communicate?	Informal inferential statistics and binomial distribution	*Informal inferential statistics* • Assumption made in evaluating evidence • Finding and interpreting *p*-value in context • Defining an unlikely event and level of significance • Making a decision based on *p*-value
The Dragon Lady	Normal distribution, properties of sampling distributions of sample	*Properties of sampling distributions of sample means* • Constructing a sampling distribution of sample means

	means (including central limit theorem).	• Difference between population and sampling distributions • Sampling variability • Properties of sampling distribution (centre, variation, shape, central limit theorem)
Can They DIG It?	Formal inferential statistics (confidence intervals and hypothesis testing)	*Confidence intervals* • Definition of confidence interval • Interpretation of a confidence interval in the context of the story • Definition of a confidence level

The analysis to determine the *participants' understanding of these selected topics* focused on the students' class artefacts and specifically students' written work associated with the intervention (i.e., the four story-based and follow-up tasks). Additionally, the participants' responses were compared to the instructor-provided resources to provide context for the understanding demonstrated. For example, participants' definitions of key terms were compared to the definitions provided in the textbook to determine if they were effectively the same (which would suggest understanding was not demonstrated) or if they were different (which would suggest understanding was demonstrated).

The analysis of the participants' responses consisted of three stages. The first stage involved open-coding of this data for common ways participants demonstrated understanding (if any) for a concept. For example, I identified the different ways that participants explained how the sampling distribution of sample means is generated and how it differs from its parent population. An example of a common way was a detailed explanation of the process of resampling to create a sampling distribution and making a clear distinction between the type of data in each distribution. This stage of analysis provided an initial idea of the participants' understanding.

The second stage of analysis involved coding the common ways that the participants responded to the story-based tasks based on my disciplinary knowledge and Skemp's (1976/1978) framework for relational and instrumental understanding.

Table 3.9 provides samples of the data (from groups 6 and 4 participants) that were coded as representing relational understanding or minimal instrumental understanding.

Table 3.9 - *Samples of relational and instructional understanding data*

Sample of relational understanding - group 6	**Sample of instrumental understanding - group 4**
Jed: So a sampling distribution is used when the entire population is unknown, so we will take our sample of 320 and randomly select a sample of 30 from this larger sample and measure the mean. We will put these 30 back into the large sample and we will then randomly take another sample of 30 from the 320, measure the mean, and continue on until we have enough means to create a sufficient enough sampling distribution from all of our sample means. You said this process is called bootstrapping. *Reema*: Right on, it is important to note that there is a known difference between the data in the sample and the data on the sampling distribution. The sample of the 320 scooters collected were randomly sampled every 15 minutes, and then tested for peak speed which was then recorded and plotted on the curve. This is the actual sample containing raw data. The data collected from the	*Reema*: To make sure that we get all possible samples we will use the empirical sampling distribution method. In this method we take a parent sample of 320 scooters and then take a sample from that parent sample and we will re-sample until we have sampled all of the sample. For example our sample size will be 30 and the only variable that will change will be the speed of the scooter.

process of bootstrapping, is the same data, except when we look at the sampling distribution, this is strictly made up from the sample means derived from bootstrapping. So it is the means of the 30 empirically sampled scooters.	

Group 6's response demonstrated a thorough understanding of the process of generating a sampling distribution from the parent sample and clearly described the differences between the two types of distributions. Though their answer contained some errors (e.g., they incorrectly suggest that all sampling distributions are empirical), their overall response indicated that they had a strong understanding of the key differences and connections between the two types of distributions and, thus, relational understanding for this concept. On the other hand, group 4's response was categorized as demonstrating minimal instrumental understanding because they attempted to explain the process of resampling by referring to empirical sampling distributions but their response was vague. For example, though they stated that a sample is taken from the parent sample, what is done with the sample is unclear. Thus, even though some understanding is demonstrated about the basics of sampling distributions, there is not enough detail to suggest that the group has achieved even instrumental understanding of sampling distributions.

Finally, the third stage of analysis involved looking across the story-based and follow-up tasks to determine if any common types or levels of understanding emerged from the data. The focus was on categorizing the information from the codes found in stages 1 and 2 by looking for themes and patterns related to different types of understanding. For example, one way that participants demonstrated instrumental understanding in all four story-based tasks was by producing statistical measures. Therefore, all of the codes that related to participants correctly producing statistical measures were grouped into a theme that I named algorithmic understanding. From this, six themes emerged of ways that participants demonstrated their understanding:

- *Algorithmic understanding* indicates that the participant could correctly follow a procedure.
- *Terminology understanding* indicates that the participant could correctly provide a definition of a statistical term, in their own words.
- *Contextual understanding* indicates that the participant could correctly interpret the statistical measure in the context of the story.
- *Choice understanding* indicates that the participant could choose between at least two different statistical measures and could correctly justify that choice.
- *Basis understanding* indicates that the participant could correctly explain the reasoning behind a statistical concept.
- *No understanding* indicates that the participant did not demonstrate any of the above types of understanding and that their answer was incorrect.

These six types provide a more detailed interpretation of the Skemp's (1976/1978) construct of understanding demonstrated by the participants.

To further analyze the participants' understanding in relation to Skemp's (1976/1978) definition, the data were reviewed to determine whether the participants were able to adapt their apparent relational understanding to new or different tasks or contexts. To investigate this, a comparison was done between participants' responses in the story-based task and their responses in the follow-up task. Only responses to portions of the story-based tasks that had analogous portions and different contexts in the follow-up task were examined. This examination focused on instances where a participant demonstrated relational understanding for a specific topic in both the story-based task and the follow-up task to identify adaptation of understanding and instances where the same understanding was not demonstrated for both to identify lack of adaptation. For example, one group of students (group 10) correctly explained the reason behind why the shape of the sampling distribution would be normal and was successful at applying their understanding of the central limit theorem from one context to the next, thus demonstrating adaptation of basis understanding. On the other hand, another group (group 9) demonstrated relational understanding of the central limit theorem in the story-based task but did not do so in the follow-up task, thus demonstrating no adaptation. Thus, group 10 showed further evidence of their relational understanding for this concept, while group 9 did not.

Shifts in participants' beliefs of the usefulness in statistics

This category, related to the first and second research question, involved coding the data to determine how participants' beliefs and examples about the usefulness of statistics shifted from the beginning of the term (prior to the intervention) to the end of the term (after the intervention). The shifts in participants' beliefs provided insight into the second research question, while the shifts in examples provided insight into the first research questions and, specifically, participants' understanding of the discipline of statistics. The main data sources were the participants' responses to the written response items, the transcripts of the interviews, and participants' responses to the reflection task.

Shifts in beliefs about the usefulness of statistics. To examine participants' shifts in beliefs about the usefulness of statistics, I analyzed participants' responses to the written response items using a two-stage process. In the first stage, I analyzed the data to determine the participants' beliefs about the usefulness of statistics at the beginning of the term and their beliefs at the end of the term. As beliefs in this study are defined as someone's viewpoint or opinion (Gal, 2002; Pajares, 1992), I coded statements that expressed the participants' viewpoints and opinions about the usefulness of statistics. Then I coded the beliefs about the usefulness of statistics as either positive (i.e., they believed that statistics would be useful in their everyday lives), neutral (i.e., they were uncertain if statistics would be useful in their everyday lives), or negative (i.e., they believed that statistics would not be useful in their everyday lives) beliefs. For example, the statement "Yes, I am majoring in marketing so it will be important for me to have good understanding of statistics to gather accurate consumer research" (participant 8) was coded as a positive belief regarding the usefulness of statistics in her career.

Stage 2 of the analysis focused on the responses of the five participants who submitted both the pre- and post-intervention written response items to determine if there was a shift in their beliefs as a result of the intervention. This involved comparing the pre-intervention beliefs with the post-intervention beliefs coded in the first stage to identify no change or changes that were positive (e.g., originally neutral to positive) or negative (e.g., originally positive to negative).

***Shifts in examples of the usefulness of statistics*.** To determine shifts in the participants' examples about how statistics can be used in their everyday lives, data from the written response items, students' responses to the reflection tasks (i.e., students' class artefacts), and transcripts of interviews were analyzed in three parts. In the first part, I focused the analysis on the examples of the usefulness of statistics provided by participants in the reflection task for each story-based task. In particular, I used my disciplinary knowledge to determine if the participants' examples of the usefulness of statistics were appropriate; that is, I determined whether the examples correctly applied the statistical topics of the story and whether the examples were realistic examples.

In the second part of the analysis, I compared the examples provided in the pre- and post-intervention written response items to determine changes in appropriateness (i.e., demonstrated an accurate understanding of how statistics could be used), specificity (i.e., provided a definite and specific use of statistics), and details (i.e., how statistics would be used in the example was provided). For example, one participant at the beginning of the term, inappropriately equated the usefulness of statistics to the general idea of being proficient at using software but, at the end of the term, appropriately recognized how they could use statistics in gambling. Further, the example became more specific at the end of the term as it moved from a generic example of using Excel in some nebulous way to the specific example of using statistics for March Madness gambling. But both examples lacked details as the participant did not indicate how statistics was specifically used in either example. Thus, the participant's examples of the usefulness of statistics were coded as more appropriate and specific at the end of the term compared to the beginning of term, but there was no change in details.

Finally, the third part of the analysis involved classifying the examples provided by the participants in all data sources as related to their work (e.g., business-related ideas of marketing campaigns by group 5 participants) or personal life (e.g., March Madness gambling by participant 14).

Relationship between features of the story-based tasks and meaningful learning

This category, connected to the third research question, involved coding the student class artefacts to determine what features of the story (if any) impacted participants' meaningful learning. The data sources were the participants' responses to the prompts in the four story-based tasks. I analyzed the data for how the specific aspects of the nature of stories impacted the dialogue written by the students for the story-based tasks. In particular, I coded the participants' responses to the story-based tasks for ways that the characters, plot, problem, context, imagery, emotions, and humor (Carter, 1993; Roberts & Stylianides, 2013; Zazkis & Liljedahl, 2009) impacted their dialogue and their learning of the statistical topics.

To illustrate, when examining the impact of the nature of the characters on the participants' written dialogue, a theme that emerged was viewing the concept from the perspective of an expert character talking to a novice. For example, participant 17 had the expert character consider the perspective of the novice by using a gambling example to explain why we look at "even better evidence against the null hypothesis" when calculating the *p*-value. Here, participant 17 considered the novice character by relating the abstract concept to a more concrete example of gambling. Other themes emerged related to the problem, context, emotions and humour aspects of stories. However, the elements of story related to plot and imagery did not appear to have an impact on participants' responses.

Once the themes were identified, further analysis was done to determine the impact of the theme on the participants' understanding of the topics. To do this, for participants' responses that demonstrated a story aspect, it was determined what type of understanding (i.e., algorithmic, terminology, basis, contextual, choice or none) was demonstrated through the use of the story element. For example, through the use of the gambling example, participant 17 demonstrated basis understanding of the concept as they correctly explained the reasoning behind the concept.

Participants' beliefs of the impact of the story-based tasks

This category, connected to all three research questions, involved coding the data to determine participants' beliefs of the impact of the story-based tasks both on their beliefs about the usefulness of statistics and on their learning of statistics. The data sources examined were the participants' responses to the post-intervention written response item and the transcripts of the interviews. Using an open-ended coding process, I read the relevant data sources and made note of any instance where a participant indicated their viewpoint or opinion about the impact of the stories on their beliefs about the usefulness of statistics or on their learning of statistics. Then I determined if there were common ways that participants expressed these beliefs. For example, in the post-intervention written response item, participant 1 stated "when applying statistics to these so called real life situations it showed me that statistics is used more than I had initially realized". Here the participant's response was interpreted as expressing the view that the stories provided more examples of how statistics could be used in their everyday lives. Common ways that participants expressed their beliefs about the impact of the story-based task on their learning included how writing dialogue, the story-context, and the implementation of the stories by the instructor impacted their learning.

23. Validation Process

Validity and reliability of the data collection and analysis processes were established in the following ways. This interview process was piloted in the pilot study, which resulted in revisions that structure questions to more accurately collect information about the intervention and its impact from participants. Member checks were also done for the interview data. The pre- and post-intervention written response items were evaluated by three experts in the field of statistics and education, which resulted in changes to strengthen the items.

Triangulation was conducted both within similar data sources and across data sources. For triangulation within similar data sources, once a code or theme was identified for a specific data source (e.g., a specific story-based tasks), I checked similar data (e.g., other story-based tasks) to determine if it emerged from all similar data sources (e.g., all four story-based tasks). To illustrate, the theme "viewing the concept from the perspective of an expert character talking to a novice" originally arose from the analysis of the participants' responses to *The Dragon Lady* story-based task. I then analyzed the participants' responses to the other story-based tasks to determine if the code was found in other stories. The consistency of the code across more than one similar data source was an indication of validity.

For triangulation across data sources, I compared themes that arose from one data source with different data sources. For example, the theme of "viewing the concept from the perspective of a novice learning from an expert" originally arose within the analysis of the responses to the story-based tasks. I then read the post-intervention written response item and the transcripts of the interviews to determine if any participant commented on the use of the novice character in this way. If there were such comments, this bolstered the validity of the theme.

Finally, to establish reliability of the results, I repeated the analysis process multiple times over a period of time to determine if each time I arrived at similar results. For example, to establish the reliability of the process of classifying participants' adaptation of understanding, I repeated the analysis twice. After I performed the first analysis, I waited approximately one week and then performed the analysis again. I then compared the results for consistency. If there were any responses that were coded differently, I revisited the definitions of the codes and compared the response to other similarly coded responses to determine which code was appropriate.

24. Ethical Considerations

Richards and Schwartz (2002) outlined four areas where qualitative researchers need to be aware of possible ethical issues: anxiety and distress, power dynamics, misrepresentation, and confidentiality. Direct participation in a study (e.g., interviews and written response items) has the potential of causing anxiety and distress in the participants even when these risks are minimized. This issue may be hard to avoid, but researchers can be sensitive to it by considering the timing of their data collection, the context of the participants, and providing information to participants about resources to seek help if they become distressed. In the researcher/participant relationship, there are always issues of power dynamics. Informed consent can help elevate some of this issue by clearly informing potential participants of the volunteer nature of participation, the opportunity for withdrawing, and how any conflicts will be managed. The third issue that Richards and Schwartz (2002) raised is about the misrepresentation of participants' comments. This can be addressed by using respondent validation (also known as member check) as part of the process. This involves the researcher providing the participant with the opportunity to review and revise their responses or the analysis of the results prior to publication. The last issue is of confidentiality of the participant's identity. To protect the data, researchers need to secure all research material using locks and passwords. As clues to the identity of the participant may be in some of the data, researchers need to remove or change these identifiers prior to publication.

To ensure that this study met the standards of ethical research, approval was sought and received from both the Conjoint Faculties Research Ethics Board at the University of Calgary and the appropriate ethics board for the university where the study was conducted.

To ensure that participants were aware of the nature of the study and what it entailed, details of the study, including the letter of consent, were posted on the course website three days prior to in-class recruitment. During the in-class recruitment, I provided each of the students with a copy of the consent letter approved by the ethics boards and orally went over its content with the class by following a script also approved by the ethics boards. In the absence of the course instructor, the students, who chose to, filled out the consent form and a signed copy was later sent to them. The consent form and my contact information were also posted on the course website.

To ensure ongoing consent, at the beginning of the four classes I attended to observe the instructor, I reminded the class who I was, that they could withdraw from the study at any time, and how they could contact me. Further, prior to the beginning of the interviews, I ensured ongoing informed consent by going over the purpose of the study with the participants, what I would ask them to do, and what will happen to the information they provided. In addition, I reminded them that they could withdraw consent without consequence or explanation at any time including during or after the interview was done. They were asked to sign another letter of consent indicating their ongoing consent. The instructor for the course was also asked to sign an approved consent letter for conducting the study in his class, the classroom observations and access to his class notes. Before each class observed, I ensured his ongoing consent prior to the class either by asking him in person or through email.

Multiple efforts were made to address potential areas of stress and anxiety for the participants. For example, for the classes observed, I addressed the class and told students that I was only there to observe the instructor and no information would be collected about them. Also, participants who agreed to be interviewed were given the choice of whether they would like to do the interview before or after the final exam.

In an effort to address the issue of misrepresentation of participants’ comments, they were provided the transcripts of their interview and asked to provide comments. Finally, to address confidentiality of the participants’ identity, pseudonyms were used for each participant and each group. Additionally, all efforts to remove any identifying features of their responses was done. The professional transcriber was also asked to sign a confidentiality form. Lastly, to protect the data, security precautions were taken in storing the data.

Chapter 4 - Findings

The findings of the study are being presented in terms of the themes and sub-themes that emerged from the data analysis. There are six themes: 1) participants' understanding of statistics concepts, 2) participants' beliefs and understanding about the usefulness of statistics, 3) differences in participants' representation of their understanding for story-based tasks, 4) participants' use of the characters (expert/novice) in the stories to demonstrate understanding, 5) impact of authentic context on participants' understanding of statistics, and 6) participants' beliefs about the story-based tasks. Each theme is presented with its sub-themes and supporting evidence.

As previously noted, regarding research questions 1 and 2, which are about the 'intervention impact' on participants' understanding and beliefs about statistics, respectively, the 'intervention impact' refers to what the students' understanding and beliefs look like when they learn statistics through the intervention. Thus, the findings of the study are not intended to establish cause and effect between the intervention and students' learning, but to indicate what happened when the students engaged in the intervention.

Throughout the results, the story-based tasks are referred to, so as a reminder, Table 4.1 provides a summary of the names of story-based tasks (in order covered in the course), the characters' names, the statistical problem investigated, and the statistical topic covered.

Table 4.1 - *Summary of story-based tasks*

	Problem presented in the story	**Key statistical concept(s)**	**Expert character(s)** / **Novice character**
Bob's Bikes	Determining if an inventory system is undervaluing items by on average more than $12	Sampling techniques and descriptive statistics	Jolene, Franca / Bart
Can Dolphins Communicate?	Whether the dolphin Aries can understand an oral communication from the dolphin Daphne	Informal inferential statistics (with a focus on informal hypothesis testing) and the binomial distribution	Emily / Sam
The Dragon Lady	Determining if electric scooters are meeting the contractual obligations	Sampling distributions of sample means, including the central limit theorem	Reema / Jed
Can They DIG It?	Determining if company should go forward with business expansion plan	Hypothesis testing and confidence intervals for mean and proportion	Kate and students chosen name of expert character / Leor

25. Participants' Understanding of Statistics Concepts

In this section, I present the first theme of the findings in terms of three subthemes: (1) the types of understanding that emerged from the analysis, (2) the understanding demonstrated by the participants for the selected statistical topics covered in the intervention, and (3) the understanding demonstrated by the participants of the discipline of statistics. These subthemes are related to research question 1.

Types of understanding that emerged from the analysis

This subtheme addresses the five major types of understanding, summarized in Table 4.2, that emerged in the analysis of the story-based and follow-up tasks (i.e., student class artefacts):

1) algorithmic, 2) terminology, 3) contextual, 4) choice, and 5) basis understanding. To provide context for these understandings, the participants' responses to the tasks (student class artefacts) were compared to the instructor provided resources (instructor class artefacts).

Table 4.2 - *Types of understanding that emerged from participants' responses*

Type of understanding	**Definition**
Algorithmic	Ability to correctly perform a procedure.
Terminology	Ability to correctly provide a definition of the concept in their own words.
Contextual	Ability to correctly interpret a concept in a context.
Choice	Ability to correctly choose appropriate statistical measure *and* explain why it is appropriate.
Basis	Ability to correctly explain the reasoning behind a concept in depth.

Algorithmic understanding involves correctly following a procedure. This could include correctly finding statistical measures (e.g., the mean, confidence intervals) by following a given/known procedure or following an algorithm to arrive at a decision or solution. In the course, students utilized a statistical software or a simulation to find all statistical measures for the story-based and follow-up tasks. Thus, when students found a statistical measure, it was by correctly following the steps of where to go in the program to find the measures. Additionally, students could follow a clearly outlined process to make a decision. As an example, comparing the *p*-value to the level of significance to correctly decide whether to reject or not reject the null hypothesis involves following a clearly outlined process to arrive at a decision. In other words, when participants did this, they were demonstrating understanding by following steps outlined by the instructor or by following an algorithm. Thus, I call this *algorithmic* understanding.

As an example of algorithmic understanding, multiple groups correctly produced the numerical and visual descriptive statistics of their data sets by using statistical software in the story-based task on descriptive statistics. For this type of understanding, no data was collected on how the participants used the algorithm (e.g., what buttons they pressed when using the software). Thus, understanding can only be determined by the result. That is, understanding was established based on whether they produced appropriate statistical measures from the statistical software or simulation, or made the appropriate decisions.

Terminology understanding involves correctly stating definitions of statistical terms. This understanding occurred in situations where the definition students provided was presented in a way that was different from the course resources (e.g., textbooks). That is, simply reproducing the definition from the textbook was not sufficient to demonstrate this type of understanding. As the participants are demonstrating understanding of definitions of terminology, I call this *terminology* understanding.

As an example of terminology understanding, group 6 provided this correct definition of outliers in *Bob's Bikes*: "An outlier is data that lies an abnormal distance from other values." This definition demonstrates terminology understanding because, not only is the definition correct, it is written in the participants' own words as it differs from the definition in the textbook ("An **outlier** is an observation of data that does not fit the rest of the data."; Holmes, Illowsky, & Dean, Section 2.2, para. 4). On the other hand, group 4 defined an outlier as "data points that are far from the average." This is not terminology understanding as an outlier is different from the rest of the data values and not just the average.

Contextual understanding relates to ability to synthesize the contextual and the statistical. It includes correctly interpreting a statistical measure in the context of the problem and appropriately using the context and statistical measures to arrive at a decision. Students could demonstrate this type of understanding if, for example, they correctly interpreted what a confidence interval means within the context of the problem or correctly used the interpretation of a confidence interval to decide on whether a new business model should be used or not. As this demonstration of understanding relies on understanding demonstrated within a context, I call this *contextual* understanding.

As an example of contextual understanding, participant 4 interpreted the *p*-value in the context of the story-based task *Can Dolphins Communicate?* as follows: “Well if Aries [the dolphin] is guessing, then yes it just has a 2% chance of getting 15 or more right out of the 20.” Here the participant correctly interprets the statistical measure (*p*-value) in the context of the problem by accurately stating the assumption in the conditional probability (the dolphin is guessing) and that the probability is of the evidence and even better evidence against the assumption (at least 15). On the other hand, participant 3 interpreted the *p*-value as follows: “So Aries would have a 0.02% chance of get it right 18/20 times if he was just guessing.” Participant 3 has correctly stated the condition, but has not correctly stated the even better evidence portion. Thus, this participant has demonstrated some contextual understanding of the *p*-value, but not a complete understanding.

Choice understanding involves choosing appropriate statistical measures or models for a situation and justifying their choices. Participants could demonstrate this understanding if, for example, they chose the best measure of centre for a specific set of data and justified this choice by referring to the presence or absence of outliers or to the problem context, or they chose which model best fits a situation (e.g., z or t means test) and correctly justified that choice. As this understanding relies on students making choices, I call this *choice* understanding.

As an example of choice understanding, group 14 explained why they chose their confidence level within the story-based task *Can They DIG It?* that covered the topic of confidence intervals:

> The 95% confidence interval was chosen so that we could be reasonably sure of the risk Leor might be going into (investing his time and money into a business that could be unprofitable), while taking into consideration the risk he could be taking by not pursuing this endeavour (moving on from this business idea). Investment of time and money is more of a risk because he has a lot to lose but moving on from the idea would make him unable to reap the rewards of his work. Because of this, a 95% confidence interval seems to be the reasonable choice in this situation.

Based on the instructor's course notes, he advised students to choose their confidence level by considering which error was more severe based on the context. Here the group correctly justified their confidence level by using the story-context to examine the competing risks (or consequences of errors) to determine that a more balanced approach of a confidence level of 95% was appropriate. On the other hand, group 16 provided this answer: "We chose a 95% confidence interval because there was only one low outlier and this shows that they won't be risking that much money." This answer does not demonstrate choice understanding as the level of confidence should be decided prior to finding the data to avoid bias and the outlier in question was not about money. Thus, though their choice may be appropriate, they have not correctly justified it.

Basis understanding involves correctly explaining a concept in terms of both how and why, that is, the underlying reasons (or basis) for the concept. Participants could demonstrate this type of understanding if, for example, they explained both how and why the sample size impacts sampling variability. I call this *basis* understanding as participants who demonstrated this understanding showed that they knew the underlying reasons (or basis) for the concept.

As an example of basis understanding, in the quote below, participant 17 correctly explained "why the 'at least 17/20' probability is more meaningful in helping us answer our research question" instead of "exactly 17" in the second story *Can Dolphins Communicate?* on informal inferential statistics. That is, they explained the reasoning behind the "or even better evidence against the null hypothesis" portion of the definition of the *p*-value within the story-context.

> *Emily*: The "extreme as" phrase on the applet is essentially testing the probability reaching a certain number. For example, if I were to try and find the probability of my winnings from a slots machine reaching $50 I would also consider winning $60 as reaching $50. If you won $80 and I asked you if you won $75 what would you say?
>
> *Sam*: I would say no, I won $80! But I kind of understand what you're saying. I won at least $75.
>
> *Emily*: In our case, this is what we want to look at as well. If Aries were to get it right 19 times out of 20, then we would count that as a success since he got it right at least 17 times.

Here the participant correctly explained that higher scores were even better evidence of communication between dolphins by relating the concept to the analogy of gambling. Then they returned to the story-context and explained how the analogy related to the statistical concept. Thus, their dialogue demonstrated basis understanding as they correctly explained the underlying reasons for the concept. On the other hand, participant 16 did not demonstrate basis understanding of this concept in their response: "We test the probability of getting at least 17 out of 20 because as the value increases the probability of meeting these restrictions will only decrease." Part of the difficulty with the answer is a lack of clarity; for example, what does "value" represent? Further, if the "as the value increases" means that as more values are added to the probability (i.e., 17, 18, 19, 20), then it is incorrect to say the probability decreases. In addition, this response describes what happens, rather than why it happens.

In summary of the first subtheme, five types of understanding emerged from the students' responses to the story-based and follow-up tasks. These types of understanding were seen across all four story-based tasks but, as will be discussed in the next section, were not all equally prevalent. As will be argued in Chapter 5, these types of understanding relate to Skemp's (1976/68) framework. Algorithmic and terminology understanding are related to instrumental understanding. While contextual, choice and basis understanding are related to relational understanding.

Understanding of selected statistical topics demonstrated by participants

In this section, I address the second subtheme of theme 1 of the findings regarding understanding demonstrated by the participants for the selected statistical topics covered in the intervention. Participants demonstrated understanding through their responses to the story-based and follow-up tasks (student class artefacts). I present the findings, first, for each selected topic based on the different types of understanding demonstrated by the participants, and second, the participants' adaptation of relational understanding to new contexts.

Types of understanding of statistical concepts demonstrated by participants. This section on the types of understanding (as defined in Table 4.2) of the statistical concepts demonstrated by participants is organized to highlight each of the four selected topics: descriptive statistics, informal inferential statistics, sampling distributions of sample means and confidence intervals and their related concepts (see Table 3.8 for details of the selected topics and their related concepts). Table 4.3 provides a summary of the types of understanding identified as demonstrated by the participants for each selected topic. Note that not all types of understanding are appropriate for all concepts. A blank cell in the table means that that type of understanding was not appropriate for the concept.

Selected topic 1: Descriptive statistics. This section describes the types of understanding demonstrated by the participants for each of the following concepts for descriptive statistics: visual descriptive statistics (histograms and box plots), measures of centre (mean, median and mode), measures of variation (range, standard deviation, interquartile range, coefficient of variation), outliers, and conclusions made based on descriptive statistics.

All participants demonstrated *algorithmic understanding* by using the statistical software package to generate correct descriptive statistics including visual descriptive statistics (histograms, box plots), measures of centre (mean, median, mode), measures of variation (standard deviation, range, interquartile range, coefficient of variation), and outliers both in the story-based task and the follow-up task.

All participants demonstrated *terminology understanding* of the measures of centre by accurately providing a definition of the mean, median and mode in their own words. They all also provided an accurate definition of at least one measure of variation, but no one provided accurate definitions of all measures of variation (i.e., range, standard deviation, interquartile range, and coefficient of variation). Most of them provided an accurate definition of an outlier in their own words while few provided a definition of either a histogram or a box plot. Instead of definitions, most of them provided accurate descriptions of what key parts of the visual representation meant. For example, they accurately stated that the heights of the histogram referred to frequency and that the ends of the box in the box plot were the first and third quartiles.

Table 4.3 - *Summary of type of understanding demonstrated by participants for each selected topic*

<table>
<tr><th>Selected Topics</th><th>Related Concepts</th><th>Algorithmic</th><th>Terminology</th><th>Contextual</th><th>Choice</th><th>Basis</th></tr>
<tr><td rowspan="4">Descriptive statistics</td><td>Visual descriptive statistics</td><td>All</td><td>Most</td><td>Most</td><td></td><td>None</td></tr>
<tr><td>Outliers</td><td>All</td><td>Most</td><td>Most</td><td></td><td>None</td></tr>
<tr><td>Measures of centre</td><td>All</td><td>All</td><td>Most</td><td>Most</td><td>None</td></tr>
<tr><td>Measures of variation</td><td>All</td><td>Most</td><td>Few</td><td>Many</td><td>None</td></tr>
<tr><td rowspan="4">Informal inferential statistics</td><td>Assumptions</td><td></td><td>Most</td><td>Most</td><td></td><td>Few</td></tr>
<tr><td>Empirical p-value</td><td>All</td><td rowspan="2">None</td><td>Few</td><td rowspan="2"></td><td rowspan="2">Some</td></tr>
<tr><td>p-value*</td><td>Many</td><td>Most</td></tr>
<tr><td>Unlikely events and levels of significance</td><td>Most</td><td>Many</td><td>Most</td><td>Most</td><td>Few</td></tr>
<tr><td rowspan="3">Sampling distributions of sample means</td><td>Construction of sampling distributions and difference from population distributions</td><td>All</td><td>Most</td><td>Most</td><td>Few</td><td>Most**</td></tr>
<tr><td>Sampling variability</td><td></td><td>Most</td><td>Some</td><td></td><td>Most</td></tr>
<tr><td>Central limit theorem</td><td>Most</td><td>Most</td><td>Some</td><td></td><td>Most</td></tr>
<tr><td rowspan="2">Confidence intervals</td><td>Confidence intervals</td><td>Most</td><td>Many</td><td>Most</td><td>Few</td><td>None</td></tr>
<tr><td>Confidence levels</td><td></td><td>None</td><td>None</td><td>Most</td><td>None</td></tr>
</table>

Note: ‘Most’ means over 75%, ‘many’ means 50-74%, ‘some’ means 25% to 49% and ‘few’ means less than 25%.

* Empirical *p*-values were found in the story-based task using a simulation. Theoretical *p*-values (or simply *p*-values) were found in the follow-up task using statistical software. They are separated here as the processes for demonstrating algorithmic understanding were different and because different contextual understanding was demonstrated between the story-based task and the follow-up task.

** Most participants demonstrated developing basis understanding of this concept.

Most participants demonstrated *contextual understanding* of measures of centre, outliers, and visual descriptive statistics by correctly interpreting them in the story-based task and the follow-up task. However, only few of them demonstrated contextual understanding of the measures of variation by correctly interpreting it in either task. In particular, most participants did not appear to understand what the measure of variation would indicate about the nature of the data. When considering the descriptive statistics together to draw a conclusion about the problem, most participants were successful at arriving at a conclusion to address the problem presented in the task. Yet, most participants arrived at their conclusions by examining the measures of centre and variation, but few participants considered visual descriptive statistics when making a conclusion.

Most participants demonstrated *choice understanding* by choosing the appropriate measure of centre for their data and correctly explained their reasoning behind their choice in both the story-based task and follow-up task. In both tasks, most participants choose the appropriate measure of variation for their data. In the story-based task, all participants provided correct but incomplete justifications of their choice. In particular, they justified their choice by comparing different measures of variation, but no participant considered all measures of variation. In the follow-up task, the participants were asked to choose between two measures of variation (interquartile range or coefficient of variation). Most participants provided correct and complete reasoning for their choice. No participant stated which appropriate visual descriptive statistic was better for a situation. Determining choice understanding for outliers is not be appropriate as only one method of finding outliers was presented to students in the course.

No participant demonstrated *basis understanding* of descriptive statistics. For example, no participant explained why the mean, median or mode were measuring the centre of the data.

In summary, for selected topic 1, most participants demonstrated algorithmic, terminology, contextual and choice understanding for most concepts for the topic of descriptive statistics, but no participant demonstrated basis understanding of any concept. Though participants demonstrated some type of understanding for most of the concepts for descriptive statistics, many participants had difficulty demonstrating contextual, choice or basis understanding of variation.

Selected topic 2: Informal inferential statistics. This section describes the types of understanding demonstrated by the participants for each of the concepts of informal inferential statistics: assumptions made to evaluate evidence, *p*-values, unlikely events, and the level of significance. Note that although students were asked to produce *p*-values, they were not given a formal definition of it, instead, they were introduced to *p*-values within the context of the story and were guided to finding the *p*-value by following the plot of the story.

All participants demonstrated *algorithmic understanding* for the empirical *p*-value by correctly generating it using a simulation in the story-based task. However, many of them did not demonstrate this understanding for the theoretical *p*-value in that they had difficulty correctly calculating it using statistical software in the follow-up task. In particular, they had difficulty using the assumption or determining what “or even better evidence against the null hypothesis” would mean when using the software. Most participants also demonstrated algorithmic understanding of an unlikely event by determining whether it was less than the level of significance by following the procedure of determining an unlikely event. For the concept of the assumption made when evaluating evidence, algorithmic understanding was not appropriate, as it did not require a procedure to follow.

Most participants demonstrated *terminology understanding* for an assumption by providing a correct definition of it in their own words. Many of them demonstrated this understanding of the level of significance as it relates to the boundary for an unlikely event, but only in terms of a beginning definition of it. Only a few provided a complete definition. That is, many participants stated that a *p*-value was unlikely if it was less than 1%, but few extended this definition to a general level of significance and even fewer stated what would happen if the *p*-value was greater than the level of significance. No participant provided a definition of the *p*-value, suggesting a lack of terminology understanding of it.

All participants demonstrated *contextual understanding* of the assumption by correctly stating it in the context of the story-based task but most demonstrated it in the follow-up task. In addition, most participants demonstrated this understanding for all aspects of the *p*-value by providing correct interpretations of them in the follow-up task, but only few demonstrated it for all aspects of the *p*-value in the context of the story-based task. In particular, most participants correctly identified the condition for the probability and the evidence being examined, but few correctly included the “or even better evidence” portion in their interpretation. All participants also demonstrated contextual understanding by correctly using the *p*-value and the concept of unlikely events to make an appropriate conclusion in the story-based tasks. However, in the follow-up task, some participants did not find a *p*-value, but of those who did, most correctly used it and the concept of unlikely events to make an appropriate conclusion using a level of significance.

Most participants demonstrated *choice understanding* by successfully choosing an appropriate level of significance and correctly justifying their choice in the follow-up task. However, choice understanding was not applicable in the story-based task for level of significance, as it was provided, and not applicable for both tasks for *p*-value since there was only one way to find it. This situation was also the case for the assumption.

Few participants demonstrated *basis understanding* of the assumption made when doing inferential statistics (i.e., the null hypothesis). Though most participants stated that they made this assumption as they needed to start from a position of skepticism, few explained why a position of skepticism was necessary. Thus, most participants did not demonstrate basis understanding, as they could not explain the reasoning behind the definition. However, many participants demonstrated this understanding by correctly explaining why the *p*-value included "or even better evidence against the null hypothesis" using the context, but no one explained the formulation of the *p*-value as a whole or why the *p*-value could be used to evaluate evidence. Also, only few of them correctly explained why the level of significance could be used as a threshold for an unlikely event.

In summary, for selected topic 2, most participants demonstrated algorithmic, terminology, contextual and choice understanding of most of the relevant concepts for the topic of informal inferential statistics. Basis understanding was only demonstrated by a few participants for the concept of assumption made when evaluating evidence and the level of significance. Most participants demonstrated basis understanding of some aspects of the *p*-value, but no participant demonstrated basis understanding for all aspects of the *p*-value.

Selected topic 3: Sampling distributions of sample means. This section describes the types of understanding demonstrated by the participants for each of the concepts of this topic: the construction of a sampling distribution, differences between sampling and population distribution, sampling variability, and properties of sampling distribution (centre, variation, shape; central limit theorem).

All participants demonstrated *algorithmic understanding* by correctly using a simulation to construct an empirical sampling distribution and to determine the empirical mean and standard deviation of the sampling distribution in both the story-based task and follow-up task. In the follow-up task, they also demonstrated this understanding by correctly finding the theoretical mean of the sampling distribution, but there were some participants who had difficulty calculating the theoretical standard deviation of the sampling distribution using the formula. Instead, they incorrectly stated that the standard deviation was the same for both the population and sampling distributions. For the concepts of difference between population and sampling distributions and sampling variability, algorithmic understanding is not applicable as neither requires a calculation nor a procedure to follow.

Most participants demonstrated *terminology understanding* of a sampling distribution, sampling variability, and the properties of the sampling distribution of sample means by providing correct definitions for them.

Most participants demonstrated *contextual understanding* by correctly using a probability determined from the sampling distribution to draw conclusions in both the story-based and follow-up tasks.

While most participants demonstrated *choice understanding* by correctly choosing the sampling distribution as the accurate model in the follow-up task, only few correctly justified why the sampling distribution was the appropriate model compared to the population distribution. As only sampling distributions of sample means were covered in the tasks, choice understanding was not applicable for sampling variability and the central limit theorem.

In the story-based task, the participants demonstrated *basis understanding*. Most of them could accurately explain how a sampling distribution of sample means is constructed and could explain the differences between a population and sampling distribution. Not only could most correctly state what sampling variability is, they could also correctly explain why it would occur. Further, most of them could correctly state the properties of sampling distributions and justify them.

In the follow-up task, while the participants did not demonstrate *basis understanding* when asked to apply their understanding of the differences between the population and sampling distribution, they did demonstrate it by successfully explaining sampling variability and the central limit theorem. Thus, they appeared to understand the difference between the population and sampling distribution when they were generalizing the concept, but had difficulty applying it to a specific situation, which suggests that most participants had developed partial basis understanding of sampling distributions.

In summary, for selected topic 3, most participants demonstrated algorithmic, terminology, contextual, and basis understanding for the related concepts, but did not for choice understanding regarding the use of a sampling distribution to model a problem.

Selected topic 4: Confidence intervals. This section describes the types of understanding demonstrated by the participants for each of the concepts of this topic: confidence intervals for one mean and one proportion, and confidence levels.

Most participants demonstrated *algorithmic understanding* by being able to use the statistical software package to generate correct confidence intervals for one mean and one proportion both in the story-based task and the follow-up task. As the confidence level is not calculated, it is not applicable for this understanding.

Many participants demonstrated *terminology understanding* for confidence interval by providing an accurate definition of it in their own words but no participant demonstrated it for confidence levels for which they did not provide an accurate definition.

Most participants demonstrated *contextual understanding* for confidence intervals by correctly interpreting it within a context and correctly using these interpretations to arrive at a conclusion to address the problem presented in the task. No participant correctly interpreted a confidence level within the context demonstrating lack of contextual understanding of it.

All participants demonstrated instrumental aspects of *choice understanding* by choosing the appropriate confidence interval for their data, which included choosing both the appropriate measure being estimated (e.g., mean vs. proportion) and the appropriate model to use (e.g., *z*-based or *t*-based normal distribution). Bu many participants had difficulty in justifying their choice. In their justification, few participants explained the difference between the measures being estimated. Those that did, provided an accurate explanation. Most participants attempted to explain why they choose a specific model. Of those that did attempt to explain, most could correctly explain why they choose a *z*-based means model, but few could correctly explain why they choose a *z*-based proportions model. Most participants chose an appropriate level of confidence for their data and justified their choice appropriately.

No participant was successful in demonstrating *basis understanding* of either a confidence interval or a confidence level. In particular, no participant could explain the relationship between a confidence level and a confidence interval.

In summary, for selected topic 4, for the concept of confidence intervals, most participants demonstrated algorithmic, terminology, and contextual understanding, but had difficulty demonstrating choice understanding. For the concept of confidence levels, most participants demonstrated choice understanding, but had difficulty demonstrating even terminology understanding.

Summary. For the four selected topics, the intervention was most successful at supporting participants to demonstrate algorithmic, terminology and contextual understanding for multiple concepts across multiple statistics topics, somewhat effective at helping participants demonstrate choice understanding, but was less effective at helping participants demonstrate basis understanding.

Adaptation of understanding to new and different contexts. The above section highlighted different ways that participants demonstrated their understanding of the four selected topics. This section examines whether participants could adapt their choice, contextual or basis understanding of the statistical concepts to new and different contexts as another way of determining their dept of understanding. The results are based only on the participants' responses to the first three story-based and follow-up tasks (i.e., student class artefacts). As the context was the same for both the story-based and follow-up task in the fourth task that covered confidence intervals, the fourth task was not included in the analysis.

Most of the participants demonstrated adaptation of understanding to a new task, that is, they showed the same type of contextual, choice or basis understanding of the topics covered in both the story-based task and the follow-up task. This adaptation occurred for contextual understanding of the assumption when evaluating evidence (i.e., the null hypothesis), choice understanding of the measure of centre, and basis understanding of the properties of the central limit theorem. However, participants had difficulty adapting their basis understanding of the differences between a sampling distribution and the population distribution. There were mixed results for the adaptation of contextual understanding of the *p*-value (i.e., its conditional nature and the idea of even better evidence). That is, some participants did show adaptation while some did not.

In summary, there is evidence that suggests that if a participant demonstrated choice, contextual or basis understanding in the story-based task, it was likely that they would continue to demonstrate the same type of understanding in a follow-up task. But this adaptation of understanding was not the same for all statistical concepts.

Understanding of the discipline of statistics demonstrated by participants

In this section, I address the third subtheme of theme 1 of the findings regarding understanding demonstrated by the participants for the discipline of statistics. While the previous section presented the understanding demonstrated by participants for the four selected statistics topics, this section addresses their understanding of the discipline of statistics with a focus on the usefulness of statistics. Participants demonstrated their understanding of the usefulness of statistics in two ways: (1) providing appropriate examples of the applications of the concepts covered in the story-based tasks and (2) providing more appropriate and specific examples at the end of the term compared to the beginning. The data for these findings were the student class artefacts (specifically, participants' responses to the reflection tasks) and the pre- and post-intervention written response items. In particular, in the reflection tasks, the participants were asked to provide an example of how the statistical concept in the story could be applied either in their personal or work life.

The first way in which the participants demonstrated understanding of the usefulness of the statistical concepts was through the appropriateness of the examples of applications of the concepts they created for each of the four reflection tasks and the pre-and post-intervention written response items.

For the reflection tasks that covered the topics of sampling techniques and descriptive statistics, and formal inferential statistics, participants provided appropriate examples of the usefulness of these topics. For example, group 5 provided the following example to show how statistics could be used in a future career that combined the stories main statistical concepts of sampling techniques and descriptive statistics.

> If you ever decide to open up a personal business, statistics can come in very useful for monitoring inventories or targeting your market by sampling how popular or demanded your product or service is. For example, say you open up a business and you're not receiving as many customers as you thought. You can conduct a quick random survey and find out where your consumers are within the city and where the demand is.

This group correctly applied the idea of sampling techniques as a way to gain demographic information for a business. Thus, they recognized how collecting data and analyzing could help solve a real-world business problem.

On the other hand, for the reflection tasks that covered the topics of informal inferential statistics and the binomial distribution, and sampling distributions of sample means, all participants had difficulty providing appropriate examples for the concepts covered in these stories. For example, for the reflection task related to informal inferential statistics, many examples mostly focused on simplistic applications of the general concepts of probability (e.g., Participant 13: "When I'm getting ready to leave for work, I give myself extra time on weekdays because I work downtown and the probability of running into traffic is much higher on the weekdays than on weekends.") or attempts at using the binomial distribution to model situations:

> An example could be a car manufacturing company. They could use this to test the brakes of all new cars. It would fit within the criteria of binomial and the 5 characteristics of it. It would allow them to know success and failure rate and determine the variable between the 2 independent events – brakes would either work or not. (Participant 10)

Though the situation described might fit the criteria of a binomial distribution, the participant appeared to misunderstand how a binomial distribution works in that meeting the binomial criteria does not help in determining the success rate.

From the analysis, it appeared that the type of story affected the ability of participants to see the usefulness of the statistical concepts beyond the one shown in the story. In particular, participants provided appropriate examples that correctly applied the statistical concepts covered in the comprehensive stories (i.e., stories used at the end of the unit on a topic). On the other hand, participants struggled to find appropriate examples or examples that fully incorporated the statistical concepts covered in the introductory stories (i.e., stories that introduced the statistical concepts and were at the beginning of the unit on the topic). Thus, there is evidence to suggest that participants found it easier to provide examples of the usefulness of topics at the end of the unit rather than at the beginning, even when presented with a story that introduced the topic.

As an additional part of the analysis, the examples provided in the reflection tasks and pre- and post-intervention written response items were classified as related either to their work or personal lives. In general, the majority of examples were classified as related to the participants' work life. For the examples provided for specific topics in the reflection tasks, the exception to this trend was for the topic of formal inferential statistics. Instead, the majority of examples related to the personal life of participants. For the written response items, participants were asked separately for examples that related to their personal life and to their work. At the beginning of the term and prior to the intervention, all participants provided an example of how they could use statistics in their future careers but only some participants provided an example for their personal lives. At the end of the term, most participants could provide an example both for their personal lives and work. This suggests that participants could provide examples of the usefulness of statistics in their everyday lives, but most saw the usefulness of statistics as related to their work.

The second way in which the participants demonstrated understanding of the usefulness of the statistical concepts was by providing more appropriate and specific examples at the end of the term compared to the beginning, which was determined by examining the examples provided in their responses to the pre-and post-intervention written response items.

In the pre-written response item, most participants gave examples that were appropriate applications of statistics, but were vague ideas of how statistics could be used and did not provide details of how exactly statistics would be used. However, in the post-intervention written response items the participants provided examples of the usefulness that were appropriate, were more specific, and provided some details of how statistics would be used in the application.

Summary of findings for the first theme

Through the analysis of participants' responses to the story-based and follow-up tasks, five types of understanding emerged: algorithmic, terminology, contextual, choice, and basis understanding. Participants were most successful at demonstrating algorithmic, terminology and contextual understanding. Though many could demonstrate choice understanding, they had difficulty with justifying their choice for measures of variation and which model to use for formal inferential statistics. Many participants also had difficulty demonstrating basis understanding for most topics. The exception is with the topic of sampling distributions of sample means. For this topic, most participants demonstrated basis understanding.

Findings regarding how participants demonstrated understanding of the discipline of statistics through their understanding of the usefulness of statistics included demonstrating understanding of how the selected topics of descriptive statistics and formal inferential statistics are useful in their everyday lives. The participants also demonstrated a better understanding of the usefulness of statistics at the end of the term compared to the beginning of the term. Finally, though they provided examples of the usefulness of statistics for their everyday lives, most examples focused on work life. Thus, the intervention appears to support their development of a better understanding of the usefulness of statistics and, thus, the discipline of statistics.

26. Participants' Beliefs and Understanding of the Usefulness of Statistics

In this section, I address the second theme of the findings that deals with the participants' beliefs and understanding of the usefulness of statistics, which is related to the second research question. For this theme, I present the results of the analysis of the participants' responses to the pre- and post-intervention written response items.

Based on responses to the pre-written response items, the majority of the participants already believed that statistics was useful in their everyday lives, few participants presented neutral beliefs about the usefulness of statistics, and no participant presented negative beliefs about the usefulness of statistics. Though participants were asked about their beliefs about the usefulness of statistics in their personal and work lives, all participants focused on the usefulness of statistics in their work and, specifically, future careers.

Based on the responses to the post-intervention written response items, the majority of participants expressed positive beliefs about the usefulness of statistics in their everyday lives and only one participant presented a neutral belief about statistics. Similar to the beginning of the term, all participants focused on how statistics would be useful in their future career but few mentioned their personal lives.

Of the participants who responded to both pre- and post-intervention written response items, all of them who had positive beliefs about the usefulness of statistics at the beginning of the term continued to have positive beliefs at the end of the term. Further, those who had neutral beliefs at the beginning of the term, shifted to positive beliefs at the end of the term.

In summary, there is evidence to suggest that the intervention may have a positive impact on students' beliefs about the usefulness of statistics. But this result is presented cautiously as it is based on a small sample size of only five participants.

27. Differences in Participants' Representation of their Understanding for Story-based Tasks

In this section, I address the third theme of the findings, related to the first research question, regarding differences in how the participants represented their understanding for the selected statistical topics in relation to the story-based tasks.

Even though participants were given the same prompts in the story-based tasks, there were significant differences in the way they responded to them. These differences are over and above the variation that can be explained by the correctness of the answers. That is, even amongst participants whose responses to the story-based tasks demonstrated similar understanding, there were differences in how they represented their understanding. Thus, in this section, I present the differences in participants' written *dialogues* (i.e., their responses to the story-based tasks; student class artefacts) as a way of further identifying the nature of the impact of the intervention on their learning and understanding of the statistics topics. These differences, illustrated next with two of the story-based tasks, were observed in three ways: 1) among the participants, 2) in relation to the textbook definitions, and 3) in relation to the explanations provided in the stories.

Differences among participants' explanations of the statistics concepts

For the first way that participants had differences in their representations, the focus is on their written dialogues for *prompts* from two different story-based tasks (Prompt A and Prompt B) to illustrate the differences among their explanations related to their understanding of the statistics concepts. The two story-based tasks were chosen as they cover the two different types of stories in the intervention, were done at different times in the term, and required different types of explanations.

Prompt A is from the first story *Bob's Bikes*, which is comprehensive (i.e., brings all of the concepts together at the end of the unit). This prompt occurred after the story has led students to gather a sample to address a problem of a badly performing computer inventory system. It asked the students to discuss the presence or absence of *outliers* for the statistical data on the difference between the actual price for goods and the inventoried price.

Prompt B is from the third story *The Dragon Lady*, which is introductory (i.e., introduces students to the topic). This prompt occurred after the story introduced the students to a problem of quality control of the peak speed of scooters that the company was initially trying to address using a population distribution but, as the expert character in the story points out, should be modelled by a sampling distribution of sample means. It asked students to explain why the *standard deviation* for the peak speed of the scooters of the parent sample is larger than the standard deviation for the mean peak speed of a batch of 30 scooters for the sampling distribution.

In Prompt A, the students needed to explain their understanding of *what* an outlier is and what it means within the story-context, while Prompt B required them to explain *why* the standard deviations are different. Due to the different natures of a "what" versus "why" question, the dialogue for both prompts are compared separately.

Prompt A. Table 4.4 provides the dialogue (participant's response to story-based task) for all of the participating groups for Prompt A (actual wording of prompt on first row of table).

Table 4.4 - *Participants' written dialogue demonstrating differences among participants for Prompt A*

Prompt A (Bob's Bikes): Have at least two of the characters determine if there are outliers in the data set. Explain how you did this. If there are outliers, explain what that means in the context of the story. If there are no outliers, explain what that means in the context of the story. One of the characters should be Bart who will either ask a question or make an incorrect suggestion. Then either Jolene or Franca will answer the question or correct his misunderstanding.	
Groups	**Participants' dialogue written for characters in *Bob's Bikes* story for prompt A**
1	*Jolene:* So by the looks of the data given in the Box plot, we don't have any outliers to deal with! *Bart:* Sooo, that's a good thing, right? I know they are kinda bad or at least important in some way, right? *Jolene:* Ya Bart that is correct, it is a good thing! They are meant to show us points in the data that have diverged from the rest of the data pattern, and in this case we have none! *Bart:* Oh that's sweet, less work then!
2	*Franca:* You're right Bart, they are important to look out for. The reason we look for them is to prevent there being errors in the data when trying to come to a sound conclusion as they can pull the data in one way or another that could falsify the information. *Bart:* How do you determine when there are outliers? *Franca:* An outlier is data that lies an abnormal distance from other values. To determine if there are outliers in this data set, we can use MegaStat on Excel.

	Jolene: Already on it. According to MegaStat we have 4 outliers which are: $4.50, $5.50, and two $6.50
3	*Jolene:* There aren't any outliers. I used the formula to determine whether there were any. *Bart:* What formula? *Franca:* Outliers can be found by calculating if any numbers fall outside of a certain range. We use the 25th percentile minus 3 times the interquartile range to find the boundary value. Any number smaller than that range is an outlier. We use the 75th percentile plus 3 times the interquartile range to find the upper boundary, and anything larger than that is an outlier. *Bart:* What's the IQR [interquartile range]? I don't know what you're talking about… *Jolene:* The IQR [interquartile range] is the range between the highest and lowest value in the middle 50% of the values, that area in the rectangle part of the box plot. I'll show you later how to do this easily using MegaStat. Until then, you'll just have to trust me.
4	*Bart:* So from the information that we have, are there any outliers? *Franca:* There are no extreme outliers in the data that we collected, so there are no extreme values. The data is skewed to the left, which means that the when the lower values of the data are more spread out, we say that data is skewed to the left. We do not have any data points that are far from the average. *Bart:* Makes sense, I think.
5	*Jolene:* Honestly Bart, I don't know what you're talking about 99% of the time. *Franca:* Hey Jolene, can you check under the descriptive statistics data in Excel if we have any outliers? *Jolene:* It looks like we don't have any! *Franca:* Oh you're right! I should have just looked at the box plot and it would have been obvious. *Bart:* Why would it be obvious?

	Jolene: Because if there were outliers, there would be little lines extending from the longer lines on the side of the box. *Bart:* Well why do we even need those? Are they important? *Franca:* We don't need them Bart, but if we had them then we would have to make note of that. An outlier would have been data that wasn't closely related to the rest of the data we collected. For example, the maximum difference between the audited and the recorded amount is $25.50. An outlier would have been $35 or -$15.00 since the lowest difference was -$5.00. If we had an extreme outlier, it would have indicated that something was either recorded wrong, or something very unusual is happening with the inventory system. *Bart:* I guess that makes sense.
6	*Jolene:* Bart, maybe you were dreaming of yourself- you remind me of an outlier- you don't fit in. *Bart:* Jolene, aren't you hilarious! Now instead of making fun of me, can one of you please explain to me what outliers are and if we have any in our data? *Franca:* Jokes aside, based on the box plot alone, it is evident that there are no outliers present within our data as they would be visible on the extremes. *Bart:* I thought that there would always be outliers in box plots? *Jolene:* No, like Franca said, outliers only exist when there are extremes or when something doesn't fit in with the rest of the data. This just means that Bob's inventory system wasn't undervaluing the products with huge differences, that the amount the products were being undervalued by was consistent. That's why we have no outliers, everything fits in and nothing extremely unusual is happening with the inventory. Sure it is undervaluing the inventory, but not by an insane amount, so that is why there are none present in this data. *Bart:* Ok, I think I might get it now.

The dialogues presented in Table 4.4 demonstrate differences in how the participants wrote dialogue for their understanding of outliers. The differences in participants' dialogue to Prompt A can be observed by how the participants defined outliers, described how they found the outliers, interpreted the outliers in the story-context, and used the personalities of the characters.

Differences in how the participants defined outliers. Five of the six groups provided a definition of an outlier. In the definitions, the groups focused on how an outlier was unusual, but they used different terminology when describing unusual. Four of the five groups correctly defined an outlier as a data value that “has diverged from the rest of the data pattern” (group 1), “lies an abnormal distance from the other values” (group 2), “wasn’t closely related to the rest of the data we collected” (group 5), and “outliers only exist when there are extremes or when something doesn’t fit in with the rest of the data” (group 6). Lastly, group 4 described an outlier as “data points that are far from the average”. Their definition incorrectly relates unusualness to the average rather than the whole data set. Of the five groups that provided a definition of an outlier, all of them described an outlier as an unusual value, but used different terminology in how they did so. Each group presented a unique understanding by using their own language for the idea of unusual and most groups chose language that was appropriate.

Differences in how the participants described how they found the outliers. Five of the six groups described how they found the outliers (or lack of outliers). These explanations included looking at the descriptive statistics output from MegaStat or Excel (groups 2 and 5) or using a formula for the outer fences to determine “if any numbers fall outside of a certain range” (group 3). But the most common way, with half of the dialogues, was to look at the box plot. But even within this simple idea of looking at the box plot to identify outliers, there are differences with how the process is described. Group 1 simply stated that they looked at the box plot. Group 5 explained the answer was “obvious” from the box plot because if there were outliers, “there would be little lines extending from the longer lines on the side of the box”. Based on the remainder of the answer, it appears that they are referring to the fences being included in the box plot output as the “little lines”. The last group (6) stated that the outliers would be visible as "extremes" on the box plot. Thus, the groups presented a unique understanding by using different ways to find the outliers and expressing how they found the outliers differently. In addition, most of the groups' choices in the language they used to describe their process were appropriate. Further, the groups chose appropriate aspects of the process to focus on, but some groups' descriptions were minimal.

Differences in how the participants interpreted the outliers in the story-context. Three of the six groups provided an interpretation of what the outliers meant in the story-context. Group 1 had a minimal interpretation but stated that the lack of outliers is "a good thing" because the presence of outliers is a "bad thing". It is not indicated what that would mean for the inventory system. On the other hand, groups 5 and 6 provided more detail into what an outlier would mean for the inventory system. Group 5 stated that if there was an outlier "it would have indicated that something was either recorded wrong, or something very unusual is happening with the inventory system". While group 6 stated that the lack of outliers

> means that Bob's inventory system wasn't undervaluing the products with huge differences, that the amount the products were being undervalued by was consistent. That's why we have no outliers, everything fits in and nothing extremely unusual is happening with the inventory. Sure it is undervaluing the inventory, but not by an insane amount, so that is why there are none present in this data.

In the last two interpretations, both groups focused on the inventory system, but group 5 focused on whether there is an error in the data entry or if there is something wrong with the inventory system. Group 6 did not consider whether an outlier would be an error, but instead focused on the story-context of the inventory system undervaluing the stock and what a lack of outlier means about this undervaluing. The differences in the responses here are seen in the extent of the interpretation and what the interpretation focuses on. Thus, these groups presented unique understandings by focusing on different aspects of the story-context in their interpretation of the outliers (or their absence) and choices made by the groups of what to focus on were appropriate.

Differences in how the participants used the personalities of the characters. All groups utilized the personality of the characters in some way in their dialogue and how they used the personalities differed between the groups. In this story-based task, Bart's personality was presented as lazy, incompetent, and indifferent. On the other hand, Jolene and Franca were both presented as competent, go-getters and frustrated with Bart. All groups utilized the incompetent nature of Bart's character by having him ask questions. Yet the questions asked differed between the groups. For example, group 1 had Bart ask questions about the importance of outliers. While group 3 had Bart ask about the formula used to find outliers and group 6 had Bart ask about his misconception that all box plots would have outliers. Thus, even though all the groups had Bart ask questions that indicated his lack of understanding of statistics, they did so in different ways. Through the use of Bart's questions, the groups addressed different aspects of the concept. For example, by having Bart ask about the importance of outliers, group 1 expanded on their initial interpretation of the outliers.

Two of the six groups utilized aspects of the opposing nature of Jolene and Bart's personality to introduce humour and emotions in their dialogue. Again, even though the group were using similar aspects of the characters, they did so indifferent ways. Group 6 used Jolene's sharp wit to inject humour into their dialogue by having Jolene joke that Bart is an outlier. Group 5 instead used Jolene's frustration to insert emotion into their dialogue by having Jolene express her anger at Bart's incompetence. In group 6's case, the use of Bart's personality allowed them to show a deeper understanding of the idea of outliers. While in group 5's case, the use of Bart's personality did not add to their demonstration of understanding. Thus, in the injection of humour and emotions into the dialogue sometimes resulted the demonstration of a deeper understanding of the concepts.

Thus, by using the personalities of the characters in writing dialogue, the participants could direct the dialogue to specific aspects of the concept, and could inject humour and emotions into their dialogue. It should be noted that the story-based task for Prompt A had the most distinct personalities of all the stories. That is, there were strong differences between the expert and novice characters that was not seen in other stories. Therefore, though the use of personalities to demonstrate unique understanding was seen in each of the story-based tasks, it was predominantly seen in this story-based task.

Prompt B. Table 4.5 provides the dialogue (participant's response to story-based task) for all of the participant groups for Prompt B (actual wording of prompt on first row of table).

Table 4.5 - *Participants' written dialogue demonstrating differences among participants for Prompt B*

Prompt B (The Dragon Lady): Have Jed and Reema discuss why the standard deviations [of the parent sample vs. the empirical sampling distribution of sample means] are so different.	
Groups	**Dialogue written for characters in the story-based task *The Dragon Lady* for prompt B**
3	*Jed*: Our sample of individual scores does not take into account the average, or I guess the mean, and no matter how many different samples you take from the parent sample, their means should result in the bell curve we saw just now. So the standard deviation of the original sample must be larger since the values aren't averaged out, meaning they are likely larger or smaller than what we get from taking the mean of a sample.
4	*Jed*: Why is one smaller than the other? Don't they have the same data? *Reema*: Both of them have the same data, but the population sample looks at all the individual scooters and that's why their extreme values have high variation. The

	sampling distribution does not look at the individual scooters but measures the means of the batches of the scooters. Therefore there won't be any of the extreme values.
6	*Jed*: Would this relate to the law of large numbers? The parent sample is smaller and the sampling distribution is much larger, therefore it would be more representative of the population? *Reema*: No, this isn't necessarily the reasoning behind the difference we are seeing in the standard deviations. It is because the parent sample will have extremes in it, and these extremes have an impact on the standard deviation as it is measuring the average distance to the mean. The standard deviation of the parent sample is 3.403. The standard deviation of the sampling distribution, is much smaller at 0.589 because it is comprised of the means which smoothes out the extremes we see in the parent sample.
7	*Jed*: This is mainly because we took the means of a larger sample, rather than the means of the individual scooters. This eliminates extremes, such as outliers. The histogram on the right [sampling distribution] showcases how taking the means of the larger sample eliminates the extremes as they regulate to how a theoretical sample would look.
8	*Reema*: The standard deviations are so different because the sampling distribution takes a look at the means of the samples taken from the parent samples. This means that the numbers collected from the parent sample will be closer together. This will make the standard deviations different because the majority of the data falls between 48 and 52 with few outliers. *Jed*: When looking at individual scooters there are some scooters that are way below 48 and others that are much higher than 52 which will change the shape of the histogram and make it not look normal.
9	*Jed*: So there is a smaller standard deviation because you are looking at the middle numbers instead of them individually. By looking at the middle numbers they end up being closer together which causes the effect of having a smaller standard deviation.
10	*Reema*: To understand this concept, we should go over what each graph is measuring. The Parent Sample is measuring individual scooters, with a large variation in speed, while the sampling distribution measures means of a group of 30 taken from our original sample. The speeds recorded in the Sampling distribution will be much closer

	therefore because averages, with far less extreme numbers are used in the applet. The standard deviation is narrower or smaller for the sampling distribution because this is using the mean of groups of numbers taken from the entire sample. There will be far less representation of extreme numbers. While the parent sample is a collection of individual values, and will therefore by skewed by extreme outliers.
12	With a smaller amount of data the average speed of the batches of 30 scooters are able to change a great deal, but with a much higher amount of data the average speeds of 10000 batches of 30 scooters, the data will have more accurate data to average out and in this case the standard deviation is much closer to the mean.

The dialogues presented in Table 4.5 demonstrate differences in how the participants explained why the standard deviation of the sampling distribution is smaller than the population distribution. In the majority of the dialogues, the participants correctly explained that the reason that the standard deviations are different had to do with the data for the parent sample being for the peak speed of individual scooters while the data for the sampling distribution being the mean peak speed for a batch of 30 scooters. Even though the participants used similar reasoning, they represented their reasoning in different ways.

Groups 8, 9 and 10 described the means in the sampling distribution as being "closer" compared to the individual scooters in the parent sample. Group 9 explained the importance of this by simply stating that this "causes the effect of having a smaller standard deviation." Group 10 explained that the closer means results in "the standard deviation [being] narrower or smaller for the sampling distribution because this is using the mean of groups of numbers taken from the entire sample. There will be far less representation of extreme numbers." Group 8 focused on the range of the types of data:

> *Reema*: This will make the standard deviations different because the majority of the data falls between 48 and 52 with few outliers.

Jed: When looking at individual scooters there are some scooters that are way below 48 and others that are much higher than 52 which will change the shape of the histogram and make it not look normal.

Even though these three groups initially utilized the similar idea of the means being closer, their reasoning for how this relates to a smaller variation for the sampling distribution was quite different.

Group 3 took a different, but similar, approach to the above groups. Instead of focusing on the means being closer, this group examined why the data in the parent sample would be more spread out. They explained that the data from the parent sample is “likely larger or smaller” than the means in the sampling distribution as “they aren’t averaged out”.

Another approach that the groups used to explain their reasoning for why the standard deviations were different relied on the means in the sampling distribution having less “extreme” values as the parent sample. Groups 4, 6, and 7 used this explanation. Both groups 4 and 7 stated that for the means in the sampling distribution “there won’t be any of the extreme values” and this “eliminates extremes, such as outliers”, respectively. Though group 4 stated that this is why “the extreme values have high variation”, neither group made a clear link to how the extreme values relate to variation. On the other hand, group 6 stated that “the parent sample will have extremes in it” while the means in the sampling distribution will “smooth out the extremes we see in the parent sample”. Unlike groups 4 and 7, group 6 attempted to make a direct link to the standard deviation by invoking the definition of the standard deviation: “It is because the parent sample will have extremes in it, and these extremes have an impact on the standard deviation as it is measuring the average distance to the mean.” Thus, groups 4 and 7 provided appropriate but incomplete explanations, while group 6 provided an appropriate and more complete explanation.

From these examples, we can see that participants presented unique understandings by choosing three different approaches to explain why the standard deviation of the sampling distribution is smaller (i.e., means being closer, individual scores being spread out, and lack of extreme values in the means). Even when the groups used similar correct reasoning, there were differences in the extent of explanation provided by the participants and how they choose to justify their reasoning. That is, groups made appropriate choices around the language that that used and which aspects of the concept to focus on in their explanations. Though most groups focused on appropriate aspects of the concept, some did not provide complete explanations.

***Summary*.** The dialogues above show how the participants wrote dialogue that demonstrated a unique understanding of outliers (Prompt A) and unique reasoning for why the standard deviation of the sampling distribution is smaller than the parent sample in different ways from each other (Prompt B). These differences were observed both in the type of language used in their definitions and explanations, what aspects of the definition or explanation that they chose focus on, and how the personalities of the characters were used. For most groups, the choice of language and choice of what aspect of the explanation or definition to focus on was appropriate, but in some instances not all aspects of the explanation or definition was included.

Differences from textbook definition of concepts

For the second way that participants had differences in their representations, the focus is on how their responses to the story-based tasks differed from the textbook using their written dialogue presented in the above tables. The textbook for the course provided multiple definitions for outliers but provided no explanation for why the standard deviations of the sampling distribution of sample means would be smaller than the standard deviation for the parent sample. Due to this, this section only focuses on how the participants' definition of an outlier found in the dialogues for Prompt A are different from the textbook definition. To illustrate this, the dialogues from Table 4.4 are compared to the definition of outliers provided in the textbook. The textbook provided the following definition of outliers:

> An **outlier** is an observation of data that does not fit the rest of the data. It is sometimes called an **extreme value**. When you graph an outlier, it will appear not to fit the pattern of the graph. Some outliers are due to mistakes (for example, writing 50 instead of 500) while others may indicate that something unusual is happening. (Holmes et al., 2016, Section 2.2, para. 4)

In a later section, outliers are further defined as “a data point that is significantly different from the other data points. These special data points may be errors or some kind of abnormality or they may be a key to understanding the data” (Holmes et al., 2016, Section 2.3, para. 16).

The textbook definition of the outliers focused on data values that are “significantly different” or that do not fit the “pattern” of the data set. Though the participants’ definitions aligned with this, they tended to use different terminology when defining a data value that is an outlier. For example, instead of using the terminology of “significantly different”, group 2 wrote “abnormal distance”, group 5 used “wasn’t closely related”, and group 6 wrote “don’t fit in”. Though group 6’s choice of “don’t fit in” appears to be similar to the textbook definition, they have made the definition their own by making a joke that the character Bart reminds Jolene of an outlier as “you don’t fit in”. Group 1’s dialogue most closely resembled the textbook definition of outlier: “points in the data that have diverged from the rest of the data pattern.” Even though it resembles the first definition in the textbook, it does show some differences by using the term “diverged” rather than “fits in”. Group 3’s definition uses the definition of “far” instead of “significantly different”, but incorrectly states that an outlier is “far from the average” rather than the data points in general. Group 4 did not provide a definition of an outlier, but instead focused on the formula for finding outliers.

Groups 2 and 5 related their definition of outliers to the idea of errors that is part of the textbook definition. Yet, group 2 focused on why errors may impact the conclusion: “The reason we look for them is to prevent there being errors in the data when trying to come to a sound conclusion as they can pull the data in one way or another that could falsify the information.” Group 5 focused on what the idea of an error would mean within the context of this problem: “If we had an extreme outlier, it would have indicated that something was either recorded wrong, or something very unusual is happening with the inventory system.” Thus, even when two groups used similar aspects of the textbook definition of outliers, they interpreted what it meant in different ways.

The above examples highlight the unique choice of language used by the groups and how that choice differs from the definition provided in the textbook. It also illustrates how different groups choose different aspects of the definition to focus on. This suggests that participants presented unique understanding of outliers by using language that made sense to them and focusing on aspects of the definition that they felt were most relevant.

Differences from story-based tasks explanations of concepts

For the third way that participants had differences in their representations, the focus is on how their responses to the story-based tasks differed from the explanations provided within the story by using their responses in the tables. For Prompt A, the story-based task *Bob’s Bikes* is a comprehensive story, which means that the portions of the story provided by the instructor only provided context and did not provide any explanation of the concepts. On the other hand, Prompt B comes from the introductory story *The Dragon Lady*, which means that the portions of the story provided by the instructor also included explanations of the concepts covered in the story. Due to this, this section focuses on how the participants' dialogues written for Prompt B were different from the explanations provided in the story-based task. To illustrate, the participants' written dialogue from Table 4.5 are compared to the story-explanation provided right before Prompt B. Students filled in the blanks for the portions in orange. Here is the story-explanation:

> "Nice work, Jed. Now, have a look at the standard deviation on the 'parent sample'", continued Reema.
>
> "Yeah, it's about **insert standard deviation**."
>
> "Right. But now look at the standard deviation on the sampling distribution. Notice anything?"
>
> "Yeah. It's **Pick either "larger", "smaller" or "equal to"**. That's weird. I thought we used the same data to produce both histograms."
>
> "The same data, yes, but not the same measurement, Jed. The sampling distribution of sample means is a histogram of mean scores (each one from a sample of 30, remember?), whereas the histogram of the 'parent sample' (Figure 1) is a histogram of 320 individual scores. Plus, the sampling distribution contains data that has come from bootstrapping (re-sampling over and over again). There are 10,000 means on that distribution, remember?"
>
> "Sorry, you'll need to explain the different standard deviations more thoroughly. My head is spinning again." Jed slouched.
>
> "Nope."
>
> "What?"
>
> Reema winked. "I'd rather you try. I want to see if you've learned anything."

The story-explanation states that the statistical data for the parent sample and for the sampling distribution are different but does not explain how the difference relates to why the standard deviations are different. As this story-explanation focuses on the two types of statistical data, it is worthwhile to examine whether the participants utilized similar terminology to the story-explanation. When comparing the participants' written dialogue to the story-explanation, the story-explanation makes a minimal connection between the statistical data and the story-context. The parent sample data is only described as "individual scores" while the sampling distribution data is described as "mean scores". What they are scores of is not stated in this portion of the story. The participants' written dialogues differ from the story-explanation by providing more of the story-context. This was done in some cases by simply using "scooters" instead of "scores" (groups 7 and 8). Group 10 provided more details by indicating that the data represented speeds of scooters and group 4 indicated that the means were for "batches of the scooters". Group 12 provided the most comprehensive use of the context by stating that the data for the sampling distribution represented "the average speed of the batches of 30 scooters". Based on this, we can see that the groups have expanded on the story-explanation and have provided more context-dependent dialogue. Thus, they are providing explanations that are unique from the story-explanations.

Two groups (4 and 10) had written dialogue that bore similarities to the conversation between Reema and Jed provided in the story-explanation. Group 4 most closely resembles the dialogue by having Jed state "Why is one smaller than the other? Don't they have the same data?", which echoes the story quite closely. As the closeness in the dialogue is seen mostly in Reema's portion, Table 4.6 places her dialogue for the story-explanation and the two groups beside each other for easier comparison.

Table 4.6 - *Comparison of instructor-provided story dialogue vs. Groups 4 and 10's responses with colour added to highlight similarities for Prompt B*

Instructor-provided story	**Group 4**	**Group 10**
The same data, yes, but not the same measurement, Jed. The sampling distribution of sample means is a histogram of mean scores (each one from a sample of 30, remember?), whereas the histogram of the 'parent sample' (Figure 1) is a histogram of 320 individual scores. Plus, the sampling distribution contains data that has come from bootstrapping (re-sampling over and over again). There are 10,000 means on that distribution, remember?	Both of them have the same data, but the population sample looks at all the individual scooters and that's why their extreme values have high variation. The sampling distribution does not look at the individual scooters but measures the means of the batches of the scooters. Therefore there won't be any of the extreme values.	To understand this concept, we should go over what each graph is measuring. The Parent Sample is measuring individual scooters, with a large variation in speed, while the sampling distribution measures means of a group of 30 taken from our original sample. The speeds recorded in the Sampling distribution will be much closer therefore because averages, with far less extreme numbers are used in the applet. The standard deviation is narrower or smaller for the sampling distribution because this is using the mean of groups of numbers taken from the entire sample. There will be far less representation of extreme numbers. While the parent sample is a collection of individual values, and will therefore by skewed by extreme outliers.

Even though the dialogues have similarities, the participants' responses were written to focus on the salient points relevant to the prompt. For example, neither of the dialogues mentioned the portion of the story-explanation related to 10,000 means, which would not be relevant to explaining this concept. Further, even though it is apparent that both submissions used the portion of the story written by the instructor to guide their answer, the differences between groups 4 and 10 are also apparent. For example, group 10 focused more on the context and provided a more detailed explanation of why the differences were important. Thus, even when the dialogues have similarities to the portions of the story provided by the instructor, there is still variation and expansion within the participants' dialogue.

In summary, the participants presented unique understanding by using different language from the story-explanation, making their dialogue more context-dependent, and by using the story-explanation as a beginning point, which is then expanded upon.

Summary of findings of the third theme

This theme addressed differences in the participants' representation of the concepts. The written dialogue for Prompts A and B illustrated how the participants' responses were different from each other and different from instructor provided material (i.e., textbook and story-explanations). Within the samples provided, there were differences among the participants in their choices of what language they used and what they chose to focus on. The differences were also seen when we compared the participants' dialogues with the instructor-provided resources of the textbook and story-explanation.

Though only two prompts were presented, the written dialogues as responses to the story-based tasks are representative of what was seen throughout the analysis and provide evidence that participants were presenting unique understanding by representing it in their own words.

This was done by choosing unique language to define terms and to explain reasoning that differed both from peers and instructor-provided resources. Across the story-based tasks, most participants used appropriate language when they did this. This was also done by choosing which aspects of the concept to focus on in an explanation. Across the story-based tasks, most participants choose appropriate aspects of the concept to focus on, but there were some instances where important aspects of the explanations were missing.

Unique understanding was seen across the story-based tasks and for multiple concepts. In general, for the topic of descriptive statistics, all participants presented a unique understanding of visual descriptive statistics, outliers, measures of centre, measures of variation, and drawing conclusions. For the topic of informal inferential statistics, most participants presented a unique understanding of the assumption, *p*-values, and the level of significance. For the topic of sampling distributions of sample means, most participants presented a unique understanding of how the sampling distribution is constructed, the differences between it and the population distributions, sampling variability, and the central limit theorem. For the topic of confidence intervals, many participants presented a unique understanding of confidence intervals.

28. Participants' Use of the Characters (Expert/Novice) in the Stories to Demonstrate Understanding

In this section, I address the fourth theme of the findings, related to the first research question, regarding the participants' use of the expert/novice characters in the stories to demonstrate their understanding of the statistics concepts. Each of the story-based tasks had two types of characters: novice and expert. The students were asked to write dialogue for both types of characters and certain prompts within the story-based tasks required the participants to include an exchange between the two types of characters. From the analysis of the participants' responses to the story-based tasks, there was evidence that through writing dialogue, participants considered the perspective of each character in demonstrating their understanding. Thus, the findings consisted of two ways in which they viewed the concepts: 1) viewing the concept from the perspective of an expert talking to a novice, and 2) viewing concept from the perspective of a novice learning from an expert. To illustrate these two ways, participants' written dialogue as responses to multiple story-based tasks (i.e., student class artefacts) is presented. The participants' written dialogues were examined for how the consideration of the different perspectives was done and how it demonstrated participants' understanding of the relevant statistical concept.

Viewing concept from the perspective of an expert talking to a novice

In this way of viewing the concept, the participants' focus was on the expert character explaining a statistical concept to the novice character. Throughout the story-based tasks, participants primarily wrote dialogue in response to the prompts in the story-based tasks where the expert character explained some aspect of the statistical concept to the novice character. When writing dialogue for the expert character, the participants had the opportunity to consider how to explain the concept to someone new to the topic. The participants considered the perspective of the expert character talking to a novice character by writing dialogue where the expert character would take into consideration the novice character by adjusting the explanation to suit the level of understanding of the novice. This adjustment occurred in three ways in which they had the expert character: 1) use examples in their explanation, 2) re-state previous knowledge or 3) use informal language in their explanations. To illustrate these, participants' responses from various stories are presented.

Used examples in their explanation. One way in which the participants had the expert character consider the perspective of the novice character was by using examples in their explanations of the statistical concepts and terms. To illustrate, consider the dialogue participant 17 wrote for the story-based task that covered informal inferential statistics for a prompt that asked students to "why the 'at least 17/20' probability is more meaningful in helping us answer our research question" instead of "exactly 17":

> *Emily*: The "extreme as" phrase on the applet is essentially testing the probability reaching a certain number. For example, if I were to try and find the probability of my winnings from a slots machine reaching $50 I would also consider winning $60 as reaching $50. If you won $80 and I asked you if you won $75 what would you say?
>
> *Sam*: I would say no, I won $80! But I kind of understand what you're saying. I won at least $75.
>
> *Emily*: In our case, this is what we want to look at as well. If Aries were to get it right 19 times out of 20, then we would count that as a success since he got it right at least 17 times.

In this dialogue, the participant had the expert character Emily use the analogy of gambling to explain the concept. This suggests that the participant recognized that someone new to the concept might find the concept too abstract and, thus, attempted to connect the concept to a more concrete situation that would potentially fit with the life experiences of the novice character.

As another example, in Table 4.4 above, group 6 did something similar when they had the expert character explain what an outlier is with a joke: "Bart, maybe you were dreaming of yourself- you remind me of an outlier- you don't fit in." Thus, they connected the abstract idea of an outlier to the concrete example of the character Bart's personality.

In both of these written dialogues, the participants took an abstract concept and made it more concrete through an example. Further, the examples they used were appropriate for the concept. By considering the difficulties that the novice character may have in understanding the concept, they created examples that could make their own understanding stronger and more memorable. Thus, the participants wrote dialogue where the expert character considered the perspective of the novice character in their explanation by using an example to make their explanation more concrete. When participants used examples in this way, they choose appropriate examples that properly illustrated the concept.

Re-stated previous knowledge. The second way in which the participants had the expert character consider the perspective of the novice character was by having the expert character reiterate previous knowledge. For the same prompt about "at least 17" in the story that covered informal inferential statistics, participant 11 had the expert character explain why they looked at "at least 17" as follows:

> *Emily*: Well this is very significant as it tells us how strong our evidence is. Getting at least 17 right means that there is a possibility of getting more than 17. If he is **NOT GUESSING**, then it shows how close we are to getting a perfect score. Looking at the 0.012 chance, we can assume that Aries is not guessing, at that there is potential of receiving a perfect score. It **SHOWS** us how close we are to perfect score.

Here participant 11 had the expert character Emily reiterate the conditional nature of the *p*-value rather than assuming the novice character would remember all of the details of the experiment.

This was also seen in the story-based task that covered descriptive statistics and sampling techniques when the expert characters reminded the novice character, Bart, of the definitions of terms. For example, group 1 wrote dialogue that had the expert character Franca not only explain which is the best measure centre for the data, but she also provided the definitions of each of the three measures of centre.

> *Franca*: You might be onto something Bart! The mean is the most relevant center of measure in this case. It is calculated through the sum all the data divided by the amount of data. In other words, the average. We would choose this over the mode or median. The mode is the number which is repeated most often in the data set. In the case of our data it would be an average difference of $2.00 which would completely throw off our findings. As for the median we wouldn't use that either as it is the middle value. If we took that value it wouldn't be taking the other data into consideration like the mean would. The middle number wouldn't be relevant.

Here the group did not assume that Bart would be familiar with these terms and instead took the time to remind him of the definitions.

In both of these examples, the participants may have recognized that someone new to a concept might not remember all of the details of the experiment or the definitions of statistical terms. Thus, they had the expert consider the perspective of the novice by repeating previous information that they deemed were relevant to the novice understanding the concept at hand. When participants had the expert character re-state previous knowledge, the previous knowledge was usually accurate and relevant.

Used informal language. The third way in which the participants had the expert character consider the perspective of the novice character was by using informal language in their explanations.

This result was already alluded to above in the section on participants' unique understanding of statistical concepts. But to highlight it here, consider as an example how participants described an outlier in the dialogue presented in Table 4.4 above. When describing that an outlier is a data value that is significantly different from the other data values, the participants used informal language. An outlier "has diverged from the rest of the data pattern" (group 1), "lies an abnormal distance from the other values" (group 2), "wasn't closely related to the rest of the data we collected" (group 5), and "outliers only exist when there are extremes or when something doesn't fit in with the rest of the data" (group 6). In this dialogue, participants utilized language that could be categorized as "everyday" rather than the more formal language used in the textbook. When participants had the expert character use informal language, it was usually appropriate.

Summary. The findings summarized here indicate how the participants viewed the concept from the perspective of an expert talking to a novice. The findings, based on examples of responses to the story-based tasks, show that when participants wrote explanations from the viewpoint of the expert character that there were various ways in which the participants considered the novice character. In the examples where the dialogue considered the novice character, the participants choose to use examples to make a connection between the abstract and the concrete, re-state previous knowledge, and translate definitions and explanations into less formal language. Throughout the story-based tasks, by far the most common way that participants had the expert character consider the perspective of the novice character was in using less formal language in their explanations. All participants did this multiple times throughout the story-based tasks. As was stated in the section on differences in participants' representation of their understanding for story-based tasks, most participants choose appropriate language to describe the concepts in their own words. Since this language was often informal, the participant demonstrated understanding by using appropriate informal language.

Having the expert character re-state previous knowledge or use examples to make connections between abstract concepts and concrete situations in their explanations was done less frequently. Instead, there were a few participants who did this repeatedly while other participants did not do it at all. When participants did do this, their examples were appropriate and the re-iteration of previous knowledge was usually correct and appropriate. In particular, the use of examples usually bolstered participants’ explanations by making them clearer and allowing the participants to demonstrate their understanding more fully compared to participants who did not use examples. The re-iteration of previous knowledge often allowed participants to demonstrate terminology understanding (a type of instrumental understanding).

Viewing concept from the perspective of a novice learning from an expert

For the second way of viewing the concept, the participants’ focus was on the novice character learning from the expert character. In many of the dialogues written by participants, the novice character was a passive receiver of information from the expert character. That is, the expert explained the concept with minimal interaction with the novice. This section examines written dialogue as responses to the story-based tasks where the two types of characters interact in a meaningful way and the novice character became an active participant in their own learning. These interactions were observed in two ways: 1) the novice character asked meaningful questions and 2) the novice character stated their understanding of the concept.

Novice character asked meaningful questions. The first way in which the participants had the novice characters actively interact with the expert characters was by making them ask questions. When the participants wrote dialogue as responses to the story-based tasks where the two types of characters interacted, they often had the novice character ask questions of the expert character. Almost all of the participants had the novice character ask a question multiple times throughout the story. This is not surprising, as there were multiple prompts throughout the story-based tasks that asked students to do this. Yet almost all participants had the novice character do this at least once without being prompted.

Often these questions were used to start the dialogue between the characters and, thus, were simply re-statements of the prompts. But there were instances where the novice character asked questions that could meaningfully contribute to their learning. These questions generally asked the expert character to interpret a statistical measure in the story-context (e.g., "*Bart*: So I am looking at the histocrap and all I see is a bunch of random bars and number gibberish. What the HELL does this all mean?" from *Bob's Bikes*, group 1), or for clarification of the explanation provided by the expert (e.g., "*Sam*: Yes but why couldn't we just do exactly 17 and know that it is possible he gets more than that?" from *Can Dolphins Communicate?*, participant 5).

Through having the novice character ask questions, the participants could direct the expert character to explain certain aspects of the concept. For example, in the story-based task on descriptive statistics and sampling techniques, after group 4 wrote dialogue that had the expert character explain the difference between a histogram and a box plot, they had the novice character ask "so why do we need a histogram?" This question focused the expert character on providing an explanation of why a specific visual representation would add information to their analysis. As another example, in the story-based task on informal inferential statistics, for the prompt that asks the students to explain why the histograms for a smaller sample size differ from one of a larger sample size, participant 19 has the novice character ask "But how did we get as many as fifteen [heads] and as little as 5?" and "If we look at the first five histograms how does it make sense that none of them resemble a bell curve?" Here the novice character directs the expert character to discuss the variation in the histograms with a smaller sample and to explain why the smaller samples did not have a shape similar to the histogram for the larger sample. In both instances, through the novice character's questions, the participants used the perspective of the novice character to personalize their understanding by focusing the explanations of the expert characters on specific aspects of the statistical concept.

Overall, the questions themselves do not necessarily demonstrate understanding on the part of the participant as most of the questions simply directed the dialogue to certain aspects of the concept. Having said that, in general, most of the questions did direct the dialogue to relevant aspects of the concept, but the questions did not always lead the expert to address all relevant aspects of the concept. That is, even with the novice character asking questions, there were many instances where the dialogue did not completely explain the concept.

Novice character stated their own understanding of the concept. The second way in which the participants had the novice characters actively interact with the expert characters was by making them state their understanding of the concept being considered. Though the expert character pre-dominantly provided the explanations of the statistical concepts, there were instances where the participants, in their responses to the story-based tasks, had the novice characters express their understanding of the concept. This was observed in two ways: 1) the novice character made a summary statement of the expert characters explanation and 2) the novice character provided an incorrect understanding of the concept. To illustrate, dialogue from various stories is presented.

Novice character made a summary statement of the expert characters explanation. There were two ways in which the participants had the novice character summarize the explanation of expert character: summary of story-explanation and summary of their own explanation.

The first way occurred when a prompt within the story-based task specifically asked for the novice character to provide a summary of a story-explanation. An example of this is the dialogue presented in Table 4.5 where the participants were prompted to write the novice's understanding of the story-explanation. Prompts like this occurred in the two introductory story-based tasks (as these were the only tasks that contained story-explanations) and covered the statistical concepts of level of significance, construction of sampling distributions, differences between population and sampling distributions, and sampling variability. For a summary of the understanding demonstrated by participants for these concepts, see Table 4.2 above.

The second way the participants had the novice character make a summary statement was when the participants wrote an explanation of the statistical concept first through the expert character, then had the novice character state their understanding of the statistical concept. For example, in the story-based task on sampling distributions, group 9 wrote a response that had the novice character express their understanding of the expert character's explanation of the properties of the central limit theorem.

> *Reema*: If the parent population was normally shaped, then automatically the sample distribution would also be normally shaped. Also, the mean will always stay the same as the parent population; the standard deviation will always change. This is because when we take a sample from the parent sample, it's just a smaller representation, so it would make sense that the mean would stay the same.
>
> *Jed*: Okay so basically you're saying that the standard deviation changes because when we increase our sample size from the parent sample, the sample distribution becomes narrower and narrower. This makes the standard deviation become smaller and smaller while the mean becomes more true to the entire population!

In this dialogue, the expert character used statistical language (e.g., “normally shaped”, “smaller representation”) to explain the properties of the sampling distribution as the sample size increases. The novice character then translated that explanation into terms that made sense for their character. For example, they used terms the less formal terms of “narrower” and “smaller” to describe the process. By having Jed state his understanding, the group demonstrated a deeper understanding of the concept by not only stating the properties of the central limit theorem, but also providing a justification for the properties and doing so using accessible language.

As another example, the next written dialogue not only demonstrated a novice character summarizing their understanding, but it also showed the expert character considering the lack of knowledge of the novice character in their explanation and the novice character asking question. That is, the dialogue written by the participant as a response to a story-based task demonstrates both considering the concept from both the perspectives of the expert and novice character. The dialogue was written by participant 14 and provided an explanation for why the probability of “at least 17” is more appropriate than “exactly 17”. It is from the story-based task on informal inferential statistics.

> *Emily*: The as extreme “as” phrase basically means “at least” as compared to “exactly”. Which in turn basically gives us more freedom when approaching the research question because getting 17 or above would help prove our research question. right would mean the same thing. When you say exactly 17/20, that excludes the other numbers above 17 to represent and answer the research question.
>
> *Sam*: But isn’t that the same thing? “At least”, “exactly”, ugh it’s making my head spin.
>
> *Emily*: Frankly, no it isn’t. By saying the word “exactly” we are putting restrictions on the research question whereas by saying “at least” we are giving more freedom to the

research question with more room to prove the research question right or wrong. Let me explain it in more depth. So, we are trying to prove whether dolphins can communicate abstract information… right?

Sam: Well, yes.

Emily: So, if Aries was to get 17/20 trials right, that would mean he communicates abstract information. But what if Aries was to get 18/20 right? Wouldn't that mean that he can communicate abstract information?

Sam: Oh, I'm starting to understand it more now! So, if Aries gets 17/20 or higher right, it would still prove the research question. If we were to say exactly, Aries would have to get no more or no less than 17/20 to prove the research question.

Emily: Yes! You've got it Sam.

In the dialogue between Sam and Emily, participant 14 started off by having Emily (the expert) provide an explanation, but they did not stop there. Instead the participant recognized that this explanation may not be clear to a novice. Thus, they had Sam ask a question that addressed where a novice character might get confused. In the response to the question, the participant had Emily start to break down her reasoning by first reminding Sam of the goal of the experiment (i.e., re-state previous knowledge). Next Emily stated her reasoning in more simple terms ("what if Aries was to get 18/20 right?) and asked Sam to think about it. Then Sam expressed his understanding of the concept. Thus, this participant considered both the viewpoint of the expert and the novice throughout their explanation. By doing this they demonstrated a deep understanding of the concept, but also wrote an in-depth explanation that perhaps addressed where the participant themselves may have had difficulty in understanding.

This type of written dialogue only appeared for the introductory story-based tasks of *Can Dolphins Communicate?* (topic of informal inferential statistics) and *The Dragon Lady* (topic of sampling distributions of sample means). Further, there were no prompts that elicited this type of dialogue. There were few participants who wrote this type of dialogue, but when they did almost all instances demonstrated strong and deep relational understanding of the concept.

Novice character stated a misconception of the concept. Participants wrote dialogue as responses to the story-based tasks where the novice character believed that they understood the concept, but in fact they expressed a misconception of it. After the novice stated their misconception, participants would have the expert attempt to address this issue. These misconceptions were interpreted as being intentional as the novice character attempted to address them.

To illustrate, in the story-based task on sampling distributions, when the characters Jed and Reema were prompted to discuss why the standard deviation for the parent sample is different from the sampling distribution, group 6 wrote dialogue as follows:

> *Jed*: Would this relate to the law of large numbers? The parent sample is smaller and the sampling distribution is much larger, therefore it would be more representative of the population?
>
> *Reema*: No, this isn't necessarily the reasoning behind the difference we are seeing in the standard deviations. It is because the parent sample will have extremes in it, and these extremes have an impact on the standard deviation as it is measuring the average distance to the mean. The standard deviation of the parent sample is 3.403. The standard deviation of the sampling distribution, is much smaller at 0.589 because it is comprised of the means which smoothes out the extremes we see in the parent sample.

The novice character incorrectly suggested that the reason has to do with the law of large numbers because, as the law states, for random samples, as the sample size increases the statistic will likely approach the parameter. Though the law of large numbers is the basis for the correct answer, the novice character incorrectly applied the law to this situation. The response from the expert character correctly addressed this misunderstanding. For the participants to write this dialogue, they would need to understand that the law of large numbers can apply in multiple situations and how it applied (and did not apply) to this situation, which suggests a strong understanding of the concept. Further, the group personalized their answer by choosing informal language (e.g., “smoothes”) and choosing not to use the story-context in their explanation. Thus, by their formation of the misconception, the participants demonstrated a strong understanding of the concept of law of large numbers in a way that may not have been possible if the misconception had not been written.

As another example, in the story-based task on formal inferential statistics, group 14 had the novice character Leor state a misconception about interpreting confidence intervals that the expert character then corrected:

> *Leor*: OK. Wow there are a lot of percentages here! So what you’re saying is that the confidence interval indicates that we have 95% certainty but we only need 60% certainty?? That’s great news!!
>
> [*expert character*]: Yikes! Just hang on a minute. Let’s go back to the whole *interval* thing. Our interval is 73.9%-96.1%. This means that we are 95% certain that the mean from the *population* lies somewhere in these numbers. When we calculated the confidence interval we *chose* 95% and hoped that the calculation would come back with an *in-ter-val* above 60%.

Leor: OH WAIT I GET IT!! Let me guess

[*expert character*]: go on….

Leor: Since the entire interval is above 60%, we can accept that this endeavor will be cost effective for Ayre and Oxford!

In this dialogue, Leor incorrectly believed that the level of confidence of 95% was what he needed to look at to determine whether the condition of whether the business expansion plan would be cost effective (i.e., at least 60%) was met. In the previous example, the participants utilized their deep relational understanding of the law of large numbers to come up with a plausible misuse of it. This is not the case here. Leor's misunderstanding is simply confusing percentages. Yet the expert character's response not only provided the definition of the confidence interval, but also clarified what the confidence level was and that the actual result of the confidence interval was not known beforehand. Thus, the group chose to address Leor's error not simply by highlighting the correct interpretation, but also providing additional details into what a confidence interval actually is. Further, unlike the first example provided, the novice character applies the expert character's explanation by revising their interpretation of what the confidence interval means in the story-context. The combined response from both characters demonstrated relational understanding through stating the definition of a confidence interval and interpreting it within the context of the story. Further, the participants personalized their dialogue by using informal language (e.g., "yikes"), choosing not to use the story-context to explain what a confidence interval is, and having the characters meaningfully engage with each other.

Though not all participants wrote dialogue that had the novice character state a misconception, there were examples of this type of dialogue in all story-based tasks. Though there were prompts that asked students to write a question or state a misconception for the novice, when provided the choice, participants pre-dominantly chose to write a question rather than state a misconception. Yet in most instances when the participants wrote dialogue that had the novice character state a misconception, the participant demonstrated relational understanding of the concept by having the expert character correctly addressing the issues that arose in the novice character's understanding.

Summary. The findings summarized here indicate how the participants viewed the concept from the perspective of a novice learning from an expert. In the participants' written dialogue as responses to the story-based tasks, there were instances where the novice character was written to no longer be a passive receiver of information, but instead was active in their learning. By engaging the novice character in the explanations of the statistical concepts, the participants had the opportunity to consider where a novice character may have difficulties in their understanding and to address them. This allowed for the participants to potentially deepen their understanding of the concepts. These types of dialogue were not ubiquitous in the participants' work, but when they did occur the participants usually demonstrated understanding of the statistical concept.

29. Impact of Authentic Context-Rich Problem on Participants' Understanding of Statistics

In this section, I address the fifth theme of the findings, related to the first research question, regarding the impact of authentic context-rich problems on the participants' understanding of statistics. Again, impact does not mean a cause-effect relationship, but rather what students were able to understand from engaging in these problems. Part of the reason for using story-based tasks was to create context-rich problems that allowed the participants to apply their understanding of the statistical concepts in a meaningful way and to learn about the statistical concepts in a motivating, authentic context. This theme examines what role the context and problem presented in the story (i.e., story-context) had on participants' understanding. The role of the story-context was observed in the participants' varying degrees of use of the story-context in two ways consisting of: 1) the application of statistical concepts and 2) the explanation of statistical concepts. The data for this theme were the participants' responses to the four story-based tasks (i.e., student class artefacts), which was presented as written dialogue.

Application of statistical concepts

The first way in which the context-rich problem impacted the participants' understanding of statistics involved the application of the statistical concepts. In all of the story-based tasks, the students were asked to apply their understanding of relevant statistical concepts within the story-context. In particular, all of the stories included a problem that needed to be resolved through statistical analysis. Within the analysis of the dialogue, the participants used the context in their statistical analysis to various degrees. To illustrate, participants' written dialogue from the first story-based task that covered the topics of sampling techniques and descriptive statistics, is presented in Table 4.7. In this story-based task, students needed to determine if Bob's inventory system needed to be repaired, which would happen if the average amount the inventory is being undervalued by in the system is more than $12. The dialogue presented here is from the end of the story-based task where participants have produced various descriptive statistics and are writing dialogue to explain to Bob, the owner of the bike shop, whether the statistical analysis conclusively shows whether the inventory system needs to be repaired, whether it can be postponed, or whether the analysis is inconclusive.

Table 4.7 – *Samples of participants' written dialogue as responses to* Bob's Bikes *demonstrating use of story-context application of concepts*

Group	Dialogue written for characters in the story-based task ***Bob's Bikes*** **(topic: descriptive statistics)**
3	*Jolene*: Unfortunately, this data puts us in a bit of a grey area. The median is $15. If we had used the median as a measure of centre, the decision would be obvious that you should fix the system. Since we decided the mean was a better measure of centre for our data, that puts the average at $11.15, which means you would be slightly better off to keep the system the way it is and just pay the extra tax. That being said, Bob, we have decided to advise you to replace the system to be safe. If you leave it the way it is, you might have a small savings but since none of us know what is wrong with the system, it is possible that the price generating software is completely erratic and it might keep changing prices.
5	*Jolene:* The thing is we worked really hard to find a solution Bob, we just don't know whether you should replace the system or not right now. *Franca:* Jolene is right, we did some hard work but it seems to be very hard to determine the right solution. You see we created a Histogram to make it a littl- … *Bob*: A Histo - what? *Franca*: A Histogram. It's just a chart that visually graphs out all the data to help better understand it, see? *Shows Bob Histogram* When you look at this graph, it's not so clear to determine whether or not the average is above or below $12.00. *Jolene:* … Then, to find the measure of the centre, we couldn't use the mode or median because there was no significant mode when it came to price differences. We also did not want to rely on the median as the bulk of our information was skewed to either side and not in the middle due to the differences between large and small items. *Franca*: We then finally calculated the standard deviation, which is the average distance of all values from the mean, helping us determine how far off some values were from the average. It turned out that the values were ranging from 1.51 to 20.77 that was, on average again, neither above nor below $12.00.

6	*Franca*: Bob, it is very difficult to tell what is going on with the inventory system, we aren't getting the average number of $12 which makes it difficult to tell whether or not you should repair the system. However, by seeing that most products are being undervalued, there is definitely a problem. The biggest issue is making the best decision. *Jolene*: Seeing how the inventory is being consistently undervalued, gives us our answer, doesn't it? The product being undervalued makes the biggest difference, so I think that the system definitely needs to be repaired. It'll be better in the long term anyways! Plus, looking at the histogram and box plot, we can clearly see that there is a problem as well which supports the undervaluing/overvaluing of the inventory. Bob, we can see based on the histogram, the largest percent of our sample of your inventory is being undervalued by $18-$21. The second largest portion is being undervalued by $21-$24.

The dialogue provided by group 5 is an example of where the story-context was used, but in a minimal way. Here, it was used to decide whether the inventory system should be used or not (i.e., is the average clearly above or below $12) but otherwise the story-context is ignored. This is seen in particular in their final concluding sentence: "It turned out that the values were ranging from 1.51 to 20.77 that was, on average again, neither above nor below $12.00". Though they use the boundary of the average of $12 in their conclusion, other important aspects of the story-context are missing. For example, there is no mention of what the "values" represent. Further, the story-context is not used in the choice of descriptive statistics used to justify the conclusion. For example, the group indicated that they considered whether to use the mean or median based on the presence or absence of outliers, but they did not consider the actual story-context in their choice. This is an example of dialogue where the group correctly applied their understanding of the statistical concepts but did not utilize the story-context to strengthen or deepen their analysis in any substantial way.

In contrast, group 6 integrated the story-context into their conclusion. Unlike group 5, they indicated what the statistical data represents (i.e., how much the inventory-system was undervaluing the stock at the bike shop) and utilized the story-context in their choice of which statistical measures to use to support their conclusion. For example, group 6 provided a specific analysis of the histogram by using it to determine the most frequent ranges for how much the items were being undervalued. That is, they used the story-context to determine what the "biggest issue" was and used that to highlight the statistical measure that best represented that issue. Like group 5, their response is appropriate based on their analysis. But unlike group 5, they considered the story-context fully by utilizing multiple aspects of it in their conclusion and by using it in their choice of which statistical measures to investigate. Group 6 demonstrated relational understanding of the application of the statistical concepts in the story by recognizing that a statistical analysis is more than simply comparing the descriptive statistics to a condition (i.e., $12) but also includes considering what the descriptive statistics tells one about the problem being solved.

Group 3's response demonstrates something different from the other two groups. In this dialogue, the group allowed the story-context to override their statistical analysis. They begin by referring to the uncertainty of their conclusion by comparing the mean and median to show that the two measures provide different conclusions. But then, rather than considering other descriptive statistics, they instead focused on the problem of having an "erratic" inventory system and how the initial cost savings might not be worth it. Thus, the statistical analysis was ignored and the conclusion was made based on their opinion that it is safer to replace the system. That is, group 3 allowed their personal beliefs to override the statistical analysis. Thus, their response demonstrated a minimal understanding of how to apply the statistical concepts.

These three dialogues are representative of how participants utilized the story-context to varying degrees across all story-based tasks. In general, participants used the story-context to deepen their analysis, used the story-context appropriately but not to deepen their analysis, or let the story-context override the statistical analysis. When considering all four story-based tasks, the most prominent way that participants utilized the story-context was to use it appropriately but not to deepen their analysis. Similar to what was seen in Wroughton and colleagues' (2013) study, there were a few instances where participants allowed the story-context to override the statistical analysis.

Explanation of statistical concepts

The second way in which the context-rich problem impacted the participants' understanding of statistics involved the explanation of the statistical concepts. In the two story-based tasks that were used to introduce a topic, the students were asked to explain their understanding of statistical concepts within the story-context. Within the analysis of the written dialogues as responses to the story-based tasks, the participants used the context in their explanations to various degrees. To illustrate, dialogue from *The Dragon Lady*, which covers the topic of sampling distributions of sample means, is presented. In Table 4.8, dialogue is presented where participants explained what a sampling distribution of sample means is and how it differs from a population distribution. The story-context here involves quality control of the peak speeds of a batch of 30 scooters.

Table 4.8 – *Samples of participants' written dialogue as responses to* The Dragon Lady *that demonstrate how the story-context was used in the explanation of statistical concepts*

Group	**Dialogue written for characters in the story-based task *The Dragon Lady* (topic: sampling distribution of sample means)**
3	*Reema*: Right. *Jed*: A sampling distribution takes a large number of samples from the sample data we have collected, so essentially it's as though the sample itself is the population. But we use the mean from those samples, whereas for the original sample we only used the measured values.
6	*Jed*: So a sampling distribution is used when the entire population is unknown, so we will take our sample of 320 and randomly select a sample of 30 from this larger sample and measure the mean. We will put these 30 back into the large sample and we will then randomly take another sample of 30 from the 320, measure the mean, and continue on until we have enough means to create a sufficient enough sampling distribution from all of our sample means. You said this process is called bootstrapping. *Reema*: Right on, it is important to note that there is a known difference between the data in the sample and the data on the sampling distribution. The sample of the 320 scooters collected were randomly sampled every 15 minutes, and then tested for peak speed which was then recorded and plotted on the curve. This is the actual sample containing raw data. The data collected from the process of bootstrapping, is the same data, except when we look at the sampling distribution, this is strictly made up from the sample means derived from bootstrapping. So it is the means of the 30 empirically sampled scooters.
9	*Reema*: Jed, are you understanding why exactly we are working with the bootstrapping strategy? *Jed*: Yeah I think I do. So basically, since our contracts states the average speed of a batch of scooters needs to be 50KMPH, if we use this bootstrapping technique, we can figure out if the scooters are fulfilling the requirements of the contract. We are not measuring for individual scooters anymore, we need to measure the mean in a batch sample from the entire sample of 320.

	Reema: Exactly, and the difference between empirical and theoretical is just if we accurately have the population data or not. The bootstrapping strategy is based solely on the data and not on the population which is how we are going to be able to get more accurate results. *Jed*: Which is perfect since using the bootstrapping strategy for sampling our data saves us not only time but money as well!

Both groups 6 and 9 used the story-context significantly in their explanations by referring to the specific nature of the data (i.e., peak speeds of scooters) and the contractual agreement. Group 3, on the other hand, did not use the context at all in their explanation.

To demonstrate the impact of the use of the story-context, I will begin by comparing group 3's context independent response to group 6's context dependent response. Both groups showed understanding of what a sampling distribution is and how it differs from a population distribution. Both groups highlighted that the sampling distribution of sample means is found by taking multiple samples from a parent population and finding the mean of each sample, and that this differs from a population distribution as the data in the population are the individual values while data for the sampling distribution are means. Even though their explanations show similar understanding of sampling distributions, by using the story-context, group 6 provided more specifics about the sampling distribution and a clearer answer. For example, when group 3 described the process of re-sampling, they simply stated that a "large number of samples" are taken from the parent population. From their explanation, it is not clear whether the process is done with or without replacement. On the other hand, group 6 utilized the story-context to demonstrate the process by referring to how the sampling distribution would be created for this specific set of data. In addition, group 3 used the generic term of "measured values" to describe the population distribution while group 6 used the story-context to specifically explain the difference between the two sets of data. Thus, both groups demonstrated understanding of what a sampling distribution is, but group 6 demonstrated a deeper understanding by using the story-context to make their explanations clearer and more explicit.

The use of the story-context within the explanation does not necessarily result in a demonstration of understanding. Group 9 used the story-context in their attempt to describe a sampling distribution but did not make it clear that the sampling distribution is created by taking multiple samples and finding the sample mean of each sample. Instead, their answer appears to suggest that only one mean is looked at for the sampling distribution.

Even though group 9 did not demonstrate understanding of what a sampling distribution is, their dialogue does explain why the sampling distribution is relevant to the story-context. By referring to the contractual obligations of the company, group 9 motivated why a sampling distribution would be appropriate to use to model this situation. As group 3's explanation does not use the story-context, they demonstrated that they can explain what a sampling distribution is but have not demonstrated that they understand when it is appropriate to use.

In short, for these dialogues, the story-context was used by group 6 to provide more details that made their explanation clearer. The story-context was also used by group 9 to motivate why the sampling distribution is appropriate to use within the specific story-context. Group 3 did not use the story-context which resulted in a response that, though it was correct, was vague in the details of what a sampling distribution is and did not explain why a sampling distribution is relevant to this story-context.

Though only samples of dialogue were presented for how the story-context was used by the participants in their explanations of concepts, the dialogues are representative of what was seen throughout the analysis. In general, if two explanations demonstrated understanding of a statistical concept, the explanation that utilized the story-context usually resulted in an explanation that was more detailed and clearer compared to the explanation without the story-context. Yet the use of the story-context in the explanation does not necessarily result in the participants demonstrating understanding.

30. Participants' Beliefs about the Story-Based Tasks

In this section, I address the sixth and final theme of the findings regarding the participants' thinking about the story-based tasks, which is related to research question 3. From the analysis of the transcripts of the interviews with participants and the participants' responses to the post-intervention written response items, participants expressed their beliefs about the story-based tasks in supporting their learning in two ways: 1) usefulness of the story-based tasks, and 2) the implementation of the story-based tasks.

Participants' beliefs about the usefulness of the story-based tasks

The participants' beliefs of the usefulness of the story-based tasks emerged from the data in two ways: 1) beliefs about writing dialogue and 2) beliefs about the story-context.

Participants' beliefs about writing dialogue. The participants considered the writing of dialogues to be useful to their meaningful learning. This emerged in the post-intervention written response items and the interviews where some participants expressed how the writing of dialogue impacted their learning of the statistical concepts. Two of the participants expressed how the writing of dialogue resulted in a deeper understanding of the statistical concepts. In the post-intervention written response item, participant 1 explained that writing dialogue "meant that you needed to understand the concepts more in order to be able to explain in your own words". In the interview, when participant 12 was asked to explain how writing dialogue between characters helped with explaining statistical concepts, they stated:

> It definitely challenged me to expand more on my understanding and not just put like, the textbook definition down. … If I knew, like on a test, if I knew I was going to be asked to explain something I would probably just look at the textbook and then like, not really memorize, but half memorize, half understand. And with that, yeah, I think it made me have a really more of an understanding.

Both of the participants highlighted that by writing dialogue they needed to develop a deeper understanding of the statistical concepts. That is, simply restating the textbook definition was insufficient. Instead, they needed to develop a deeper understanding so they could write the dialogue. In addition, participant 12 spoke about how writing dialogue for the expert character mimicked explaining the concepts to a peer:

> But I think because you are not memorizing it, like you are forced to actually have an understanding of it, and that is sort of like a teaching – the teaching method – of understanding something well enough to teach it to someone, rather than understanding, or like, knowing it so you can tell someone what it says in the textbook or whatever. (interview)

For this participant, writing dialogue for the expert character engaged them in a form of peer learning. For their perspective, this allowed them to deepen their understanding by forcing them to think about how they would explain the concept to someone else.

Yet not all of the participants found the writing of dialogue to be helpful in developing their understanding. In the interview, participant 2 found that the use of dialogue felt “forced”. In particular, they commented on how dialogue in conversations could follow certain parameters, but the dialogue for the story did not allow for this:

> For example, the way that I am talking to you right now is not the way I would have talked in [*Can They DIG It?*] when I am explaining things. I can't be like, "Oh yeah, I just wanted to … I can't remember what the name of it is, but let me just go back and check," like you can't do that in a written dialogue, you would assume. In real life you would be like, "Well this is how you do it," and the person would be like, "Why did you do it like that?" and you would explain it. … So that is kind of why it feels forced because you are taking a real-life aspect, but at the same time it is not because it still has to be theoretical on paper.

When asked if they felt that the instructor would be uncomfortable with more informal dialogue, they stated,

> Well it is not that [the instructor] would have been uncomfortable with that, it is just that wouldn't have room in an educational sort of environment, because you are trying to get some terms, and those terms will be asked to you on a quiz.

What the participant suggests here is that the language of informal dialogue would not be acceptable on an exam. Thus, they felt that they needed to write more formal dialogue, which then felt forced. Due to this, they did not believe that writing dialogue benefited their learning.

In summary, some participants found that the challenge of writing their understanding as dialogue supported their learning as it pushed them beyond textbook understanding, forced them to write their understanding in their own words, and had them consider how they would explain a concept to a novice. Yet other participants found a disconnect between what was expected in the story-based task (informal dialogue) with what would be expected on an exam and this disconnect impeded their ability to develop understanding of the concepts.

Participants' beliefs about the story-context. The participants also considered the story-context to be useful to their learning. The story-based tasks provided the opportunity for students to learn about the statistical concepts through real-world applications. In the interviews and the post-intervention written response items, the participants expressed their thinking about how the story-context impacted their understanding of statistics in two ways: 1) learning the statistical concepts and 2) their beliefs about the usefulness of statistics.

Learning the statistical concepts. Some of the participants found that the story-based tasks created a situation where they could learn about the statistical concepts through the story-context. This was highlighted by participants in different ways. For example, in the post-intervention written response item, participant 5 found the most helpful story-based task to be *Can They DIG It?* because "it really secured my understanding for hypothesis testing and confidence intervals by giving me a hands-on way to practice". Thus, they focused on how the story-based tasks helped them learn by applying the statistical concepts in a real-world context. Participant 12 echoed this when they described the story-based tasks as "learning by doing" (interview).

Other participants focused on how the story-context took the abstract statistical concepts and made them more concrete. For example, participant 1 commented that "that the stories applied statistics to potentially real situations" (post-intervention written response item) and participant 5 found that "the stories helped give real life meaning to the statistical concepts taught in the course so that it wasn't all just numbers and formulas" (post-intervention written response item). For these participants, the benefits of the story-context were seeing how these seemingly abstract concepts had real-world applications. The context provided in the story-based tasks helped them to deepen their understanding of statistics by having them apply the concepts in a meaningful context and to see how the concepts were more than just "number and formulas" but had actual real-world applications.

Usefulness of statistics. Some of the participants also found that the story-based tasks created a situation where they could learn about the usefulness of statistics meaningfully. They noted that the stories impacted their understanding of how statistics was useful by highlighting that there were more uses than previously believed and providing specific instances of how statistics can be used.

Though most participants knew that statistics was useful prior to taking the course, that did not mean that they had a full understanding of how statistics could be used. Participant 1 stated that "when applying statistics to these so called real life situations it showed me that statistics is used more than I had initially realized" (post-intervention written response item). Participant 12 echoed this sentiment:

> Ever since I can remember, I've thought that stats is a useful thing to learn for critical thinking but I hadn't really thought about it so explicitly for business purposes so I think the stories just gave a better context for a business environment. (post-intervention written response item)

For these participants, the story-based tasks provided them with examples that helped solidify and expand on their beliefs about the usefulness of statistics.

The story-based tasks also helped students not only realize that statistics was more useful than they originally thought but it also helped them understand the specifics of how statistics is used in real-world applications. For these participants, prior to the class, they knew that statistics was useful, but they didn't know the specifics of how they were useful. For example, in the interview, participant 2 easily discussed various ways that statistics could be used in their everyday life. When asked if they had this understanding prior to the course, they responded,

> No, I didn't have those ideas before class. It does help clear up the uses of it, rather it is more accurate, because you can be like, "Oh yeah, use stats for demographics." Good! "What do you know?" I don't know. Or I do Google Analytics and, "What is it?" I don't know! It is just stats, don't worry about it, and then that is not very clear. But then you can be like, "Oh no, I use linear regression to plot the number of users that click through this versus usability," or, "I did this …" like I do the probability of someone purchasing that if they are from this neighbourhood, and that is more clear.

Participant 3 stated something similar when responding to the question about what impact the stories had on their beliefs about the usefulness of statistics in the post-intervention written response: "I knew statistics was useful in some fields, however it was interesting to see how exactly it could be used in a field that was more similar to mine". This highlights that these participants were aware of applications of statistical applications prior to the course and intervention, but they did not have the skills or knowledge to understand how the applications worked. Yet from learning through the story-based tasks, they developed a better understanding of the specifics of how statistics is used in these real-world scenarios.

Participants' beliefs about the implementation of the story-based tasks

The second way in which the participants considered the impact of the story-based tasks on their learning was in relation to the implementation of the tasks. In the interviews and post-intervention written response items, the participants provided their thinking about how the implementation of the story-based tasks during class time helped or hindered their understanding of statistical concepts. The participants' beliefs were expressed in two ways: 1) small group discussions and 2) level of instruction.

Small group discussions. The participants considered the small-group discussions involving the story-based tasks to be helpful in supporting their learning. In the interviews, both participants mentioned the benefit they found in the small group discussions that occurred during the classroom time provided to work on the story-based tasks. When asked about the usefulness of talking about the concepts with classmates while working on the story-based tasks, participant 2 responded:

> It does work in almost every situation, if you are able to explain it to somebody it means you understood the concept. If you are able to understand it from somebody it means you are learning the concept so there is more reinforcement rather than the teacher saying, "Look, it is this. It is this, it is this, you press this and you are done, and that is it."

They highlighted how explaining the concepts to others helped in their learning, but also how hearing other students' explanations was also useful. Thus, they focused on the peer learning that occurred during the small group discussions in class. Participant 12 also commented on the benefits of talking with classmates. Their group partner withdrew from the course, which meant that they had to complete the last story-based task on their own: "I was really disappointed at the end when I didn't have anyone to ask about. Yeah I think group work is good in this situation." This suggests that in the implementation of the story-based tasks allowing the students time to talk about the story-based task with their peers had perceived benefits for the students.

Level of instruction. The participants also considered the level of instruction in implementing the story-based task had an impact on their learning. They expressed their beliefs on the level of instruction provided by the instructor in regards to how much support they received when completing the story-based tasks. For example, in the interview, participant 1 stated that the story they found to be the most useful in their learning was the comprehensive story *Can They DIG It?* because prior to the story-based task "there was a lot of time spent on ensuring the class understood hypothesis testing and confidence intervals. When applying to the story it was easier to apply because an understanding was there" (post-intervention written response item). In the interview, participant 12 initially stated that the part that they found most useful about the stories was "having to figure it out yourself … sort of like, learning by discovery". When asked if that meant that the introductory stories were more useful in their learning than the comprehensive stories, the participant stated:

> Actually, no, I would say I probably learned … when we had the lectures beforehand I actually didn't study them, we actually had … for me I had an idea of what was happening and I didn't study it to prepare for the assignment. I think just having the little briefing beforehand helped a little bit, but that was still, like, it still allows for the learning by doing, or whatever, just with a little bit of background information.

Both of these participants highlighted that the stories that were more useful to their learning were those that had some instruction on the concept prior to doing the story-based task, which then allowed them to apply their understanding in the task.

When participant 1 was asked what story they learnt from the least, it was the introductory story *The Dragon Lady* because "it's hard to teach yourself the concepts and write the story not knowing if you're entirely correct" (post-intervention written response item). This highlights that the participant found it difficult to learn the concepts without some support from the instructor because they didn't know if they were developing a correct understanding. Participant 16, who had a negative view of the usefulness of stories, stated in their post-intervention written response that the stories "over-complicated concepts that we did not learn in class. Unfortunately the professor spent most of the time in class talking about the stories instead of lecturing" (post-intervention written response item). This suggests that some of the participants found that learning strictly from the stories had a negative impact on their learning as they were uncertain about their understanding and felt that they needed more support from their instructor in their learning. However, not all participants had these beliefs. Participant 5 felt that the story that was most useful in their learning was the introductory story *Can Dolphins Communicate?* because it "was the first story in which we had to teach our selves the content as we went along and didn't already know what was being used in the story" (post-intervention written response item). This suggests that some students found the challenge of learning the content through the story useful in developing their understanding of new statistical concepts.

Based on this evidence, it appears that students would like to be introduced at least somewhat to the concepts prior to learning about them in the story-based tasks and would like more direct support from the instructor as they learn the statistical concepts from the stories.

Summary

This summary addresses the sixth and final theme of the findings regarding the participants' thinking about the story-based tasks. Participants expressed various beliefs about the usefulness and implementation of the story-based tasks. Some of the participants believed that by writing dialogue within the story-based tasks they were challenged to develop a deeper understanding of the statistical concepts, while one participant found that writing informal dialogue did not connect with the more formal expectations of an exam. Participants found the authentic contexts presented in the stories to be useful as it grounded the statistical concepts in an application and allowed them to explore the statistical concepts in a "hands on" manner.

Participants believed that the story-based tasks impacted their understanding of the full extent of the usefulness of statistics and provided specific instances of how statistics can be used to solve real-world problems.

The participants also expressed views on how the story-based tasks were implemented in the classroom. From the interviews, the two participants highlighted that having the opportunity to discuss the story-based tasks in small groups was useful in their learning. Some participants voiced varying degrees of concerns about how much the instructor lectured prior to the story-based tasks. This ranged from suggesting that a little bit of instruction prior to the story-based tasks would be useful to others suggesting that significantly more lecturing was needed.

31. Chapter Summary

This summary addresses the findings of all six themes presented in this chapter that emerged from the data analysis. Through the story-based tasks, participants demonstrated their understanding of various statistical concepts in the four topic areas of descriptive statistics, informal inferential statistics, sampling distributions of sample means, and confidence intervals. In particular, most participants demonstrated algorithmic, terminology and choice understanding of most concepts for all four topics. Participants also demonstrated contextual understanding of most of the concepts but had particular difficulty with variation within a sample, *p*-values and confidence levels. With the exception of the topic of sampling distributions of sample means, most participants had difficulty demonstrating basis understanding of most concepts in the other three topic areas. Further, most participants who demonstrated contextual, basis or choice understanding of a statistical concept could adapt their understanding to a new context and task.

Participants demonstrated understanding of the discipline of statistics by demonstrating understanding of the usefulness of statistics. Most participants could provide examples of the usefulness of descriptive statistics and formal inferential statistics. Further, participants' examples of the usefulness of statistics became more appropriate and less vague from the beginning to the end of the course.

Most participants started the course with positive beliefs about the usefulness of statistics in their everyday lives. For those participants who expressed their beliefs both at the beginning and at the end of the term, all of them expressed positive beliefs about the usefulness of statistics at the end of the term.

Even though participants were given the same prompts, they wrote unique dialogue for their understanding of statistical concepts that differed both from the other participants and from the instructor-provided resources. These differences included using unique language and focusing on different aspects of the concept.

When writing dialogue as responses to the story-based tasks, the participants had the opportunity to consider the perspective of the expert and novice characters. When writing dialogue for the expert character, participants demonstrated how to explain a concept to a novice by using examples, informal language, and re-stating previous knowledge. When writing dialogues for the novice character, the participants considered the perspective of someone new to a concept by having the novice character ask questions and state their understanding of the concept.

The authentic context-rich problems in the stories also played a role in participants' written dialogue. In particular, participants used the story-context to various degrees when applying the statistical concepts and when providing explanations of the statistical concepts. When participants demonstrated understanding and utilized the story-context, their dialogue usually demonstrated a strong and deep understanding of the concept.

The participants expressed their beliefs of the usefulness of the story-based tasks both on their own learning and on their beliefs about the usefulness of statistics. The participants believed that the story-based tasks helped them develop their understanding of statistical concepts by forcing them to write their understanding in their own words and by applying the statistical concepts. They also found the opportunity to discuss the statistical concepts in small groups to be helpful. The story-based tasks did not help in some of the participants' learning because they found the dialogue to be "forced" and it did not align with expectations on exams. Some students also voiced concerns that the instructor did not provide enough support for their learning and instead relied too heavily on the stories to do the teaching. The story-based tasks reinforced participants' understanding of the usefulness of statistics by providing a deeper understanding of how statistics can be used and providing specific examples for the application of statistics.

Chapter 5 - Discussion

The chapter is organized into five sections. The first three sections highlight the various outcomes for the three research questions. The fourth and fifth sections connect the story-based tasks to the theoretical frameworks related to this study, which consist of the reform movement in statistics education and the theory of learning of constructivism. In discussing the findings of this study, it is important to note that 'impact' is considered to be what students' understanding and beliefs looked like when they engaged in learning statistics through the intervention. Throughout the chapter, connections to the theoretical perspectives and research literature are discussed as appropriate to situate the findings to the field.

32. Impact of Intervention on Participants' Understanding of Statistics

This section discusses findings of the ***first research question*** regarding the impact of the intervention on (1) participants' understanding of selected statistical topics and (2) participants' understanding of the discipline of statistics.

Participants' understanding of selected statistical topics

The study focused on the participants' understanding of four selected statistical topics: descriptive statistics, informal inferential statistics, sampling distributions of sample means, and confidence intervals. Findings indicate that the intervention impacted the students' understanding of these topics in terms of their related concepts in two ways: (1) development of various types of understanding and (2) personalization of knowledge.

Development of various types of understanding. Participants demonstrated five types of understanding of the statistical topics, collectively, based on their work with the story-based and follow-up tasks: algorithmic, terminology, contextual, choice, and basis. *Algorithmic* understanding occurred when participants correctly followed a step-by-step process to arrive at an answer. *Terminology* understanding was demonstrated when they correctly wrote a statistical definition in their own words. *Contextual* understanding occurred when they correctly interpreted statistical measures in a context or used the statistical measures to understand the problem being investigated. *Choice* understanding involved choosing an appropriate measure or model to investigate a situation and appropriately justifying that choice. *Basis* understanding occurred when participants could correctly explain the reasoning behind a concept.

Though all five types of understanding emerged from the participants' work, some of them were more prevalent than others. Most participants demonstrated algorithmic, terminology and contextual understanding for most concepts in the four selected topics. Choice understanding was demonstrated by most participants for the topics of descriptive statistics and informal inferential statistics. Basis understanding was only demonstrated by most participants for the topic of sampling distributions of sample means. Additionally, most participants who demonstrated choice, contextual or basis understanding in one task or context continued to demonstrate that type of understanding in a new task or context. The findings can be connected to: 1) Skemp's (1976/1978) theory and 2) the literature regarding the relevance of them to the field, discussed in the following two subsections.

Connection to Skemp. In relation to Skemp's (1976/1978) theory of instrumental and relational understanding, the participants demonstrated both of these understanding through the preceding five types of understanding. Algorithmic and terminology represent instrumental understanding (e.g., knowing how to do an algorithm and what a term means) while choice, contextual, and basis represent relational understanding (e.g., knowing which statistical measure or model to use and *why* it is appropriate to use; knowing *why* the mean is a measure of central tendency). For algorithmic understanding, there was no indication that participants understood why the steps of the algorithm were appropriate. For terminology understanding, while the participants were able to describe the definition in their own words, this does not necessarily indicate that they understood why it was appropriate. For example, just because they can write the definition of the mean in their own words does not imply that they understand why the mean is a measure of central tendency. In contrast, for choice, contextual, and basis understanding, participants demonstrated knowledge of 'why'. For example, choice included justifying why a concept/measure/model was appropriate to use, contextual included demonstrating what the result means and why it is important or relevant to understanding the problem (i.e., why and how a statistical measure can provide insight into the context and problem being investigated), and basis included explaining the reasoning behind a concept.

Table 5.1 summarizes the findings in terms of Skemp's theory, that is, for instrumental understanding, *how* (algorithmic) and *what* (terminology) and for relational understanding, *meaning/use* (contextual), *which* (choice), and *why/reason* (basis). For example, 'how' indicates knowing steps, 'what' indicates knowing definition, 'meaning/use' indicates ability to interpret and use concept, 'which' indicates ability to choose and justify, and 'why/reason' indicates knowing reason underling the concept.

Table 5.1 – *Summary of instrumental and relational understanding demonstrated by participants*

Topics	**Concepts**	**Instrumental**		**Relational**		
		How	**What**	**Meaning/use**	**Which**	**Why/Reason**
Descriptive statistics	Visual descriptive statistics	All	Most	Most		None
	Outliers	All	Most	Most		None
	Measures of centre	All	All	Most	Most	None

	Measures of variation	All	Most	Few	Many	None
Informal inferential statistics	Assumptions		Most	Most		Few
	Empirical *p*-value	All	None	Few		Some
	p-value	Many		Most		
	Unlikely events and levels of significance	Most	Many	Most	Most	Few
Sampling distributions of sample means	Construction of sampling distributions and difference from population distributions	All	Most	Most	Few	Most
	Sampling variability		Most	Some		Most
	Central limit theorem	Most	Most	Some		Most
Confidence intervals	Confidence intervals	Most	Many	Most	Few	None
	Confidence levels		None	None	Most	None

Note: ‘Most’ means over 75%, ‘many’ means 50-74%, ‘some’ means 25% to 49% and ‘few’ means less than 25%.

The table indicates that the intervention was more successful in supporting instrumental than relational understanding of the concepts. Within instrumental, it was more successful in supporting how to follow prescribed steps than defining the concept. Within relational, it was least successful in supporting knowledge of underlying why/meaning of the concept. The difficulty of participants developing this type of understanding is similar to the challenges faced by the reform movement in calculus. As Skemp (1976/1978) pointed out, instrumental understanding is easier for students to develop than relational understanding, which requires students to go beyond surface features and think deeply about the concepts. While the intervention was intended to provide opportunities for students to develop relational understanding, if students' prior learning experiences were oriented towards instrumental understanding they are more likely to engage in the story-based tasks in an instrumental way. Thus, providing students with experiences to understand what it means, and how, to engage in learning for relational understanding may be a necessary prerequisite to engaging in the story-based activities. In spite of this, the intervention did validate and extend what we know about students' knowledge or understanding of specific statistics topics, as discussed next.

Connection to other studies. This subsection on the findings regarding the participants' development of various types of understanding of statistics discusses connections of the findings to other studies. Some of these findings of this study are also similar to those of related research in the field of statistics education discussed in the literature review, thus providing further evidence of them. For example, for the topic of *descriptive statistics*, participants had difficulty with relational understanding of variation, which is similar to findings by Gougis et al. (2016) and Mathews and Clark (2007). For the topic of *sampling distributions* of sample means participants had difficulty distinguishing between variations within a sample versus between samples similar to findings of Reaburn (2010) and Stevens and Paloscay (2012). They also had difficulty choosing when it was appropriate to use the sampling distribution versus the population distribution similar to findings by delMas et al. (2007) and Jacob and Doerr (2014).

In addition, they were able to explain the concepts of the sampling distribution, they had difficulty applying them similar to findings by Jacob and Doerr (2014). For the topic of *confidence interval*, the participants were not able to explain confidence levels similar to findings by Jacob and Doerr (2014) and Reaburn (2010). Thus, the intervention did not help them to overcome these issues that were identified in other studies, but helped them for other concepts.

There were other concepts for which the intervention produced different outcomes compared to other studies as in the following examples for each of the statistical topics investigated. For *descriptive statistics*, the participants were able to provide accurate definitions of measures of centre and at least one type of variation (e.g., standard deviation), which is different from findings by Lavy and Mashiach-Eizenberg (2009) and Mathews and Clark (2007), which suggested that students struggle with even understanding the definition of standard deviation. Further, participants successfully interpreted a histogram in the context of the story, which is different from the findings of Kaplan et al. (2014) who found that students had difficulty reading histograms. For *informal inferential statistics*, the participants demonstrated contextual understanding of the *p*-value as a conditional probability, which is different from findings by Sotos et al. (2009) that indicated participants had difficulties understanding the properties of the *p*-values. Also, for informal inferential statistics, participants understood the "even better evidence against the assumption" portion of the definition of the *p*-value, which adds to the literature as previous research has not examined this. For *sampling distribution* of sampling means, the participants understood that the sampling distribution is a distribution of statistics, which is different from findings of previous research that suggested that students have difficulty understanding even the basics of sampling distributions (Chance et al., 2004, Reaburn, 2010; Stevens & Paloscay, 2012). For *confidence interval*, the participants correctly interpreted multiple confidence intervals within a context, which is different from findings by Canal and Gutiérrez (2010), delMas et al. (2007), Fiddler (2006), Reaburn (2010), and Stevens and Palocsay, (2012), which indicated that participants had difficulty interpreting confidence intervals.

The differences between the findings of this study and those of other studies is that while participants demonstrated learning of these concepts by engaging in the intervention, other students found that they were unable to do so. Taking this all together, this suggests that learning statistics through stories has the potential to have a positive impact on students' understanding of concepts in areas that previous research suggests is difficult for students to learn.

Personalization of knowledge. This section discusses the second way in which the intervention supported the students' understanding of selected statistical topics through personalization of the knowledge they developed. The findings indicate that the participants demonstrated four ways in which the intervention enabled them to personalize the knowledge they constructed of the statistical topics: 1) choice of language, 2) choice of which aspects of the concept to focus on, 3) choice of extent of story-context to use, and 4) choice of how much they had the story-characters engage with each other. Through these choices, participants personalized their explanations to make sense of the concepts for themselves. *Choice of language* occurred when participants provided definitions of key terms and explained concepts using informal language or language that differed from their peers and the instructor-provided resources (e.g., textbook). *Choice of which aspects of the concept to focus on* occurred when participants explained the reasoning behind concepts by focusing on different aspects of the concept, or had the novice character in the story ask questions or state misconceptions to direct the expert to aspects of the concept. *Choice of extent of story-context to use* occurred when participants chose how much or how little of the story-context to use in their interpretations of the statistical measures or in their explanations of the statistical concepts. Finally, the participants personalized their knowledge by *choosing how much or how little they had the characters in the story engage with each other* in the participant written dialogue.

Though all types of personalization emerged from the participants' work, not all types necessarily led to the demonstration of understanding. For example, most participants choose appropriate language to personalize their knowledge of terms and concepts, which allowed them to demonstrate terminology understanding (i.e., instrumental understanding). Further, when participants choose to have meaningful interactions between the two types of characters in the stories, this usually resulted in the demonstration of relational understanding by demonstrating basis, contextual or choice understanding. Yet, when choosing which aspect of the concept to focus on, many participants did not choose all relevant aspects of the focus in their explanations, which resulted in them not fully demonstrating relational understanding.

Based on my review of the literature, there were no studies found within statistics education that investigated students' personalization of knowledge. This suggests that this study adds to the field by providing evidence that learning through stories can result in the personalization of knowledge and that the personalization of knowledge can support learning of statistics.

Summary. The findings of the impact of the intervention on students' understanding suggest that the intervention can support the development of both instrumental and relational understanding of the four statistical topics investigated. Additionally, the findings suggest that the story-based tasks can support personalization of knowledge of the four statistical topics investigated, which can then support students' learning and understanding of the concepts. These findings are further supported by feedback from the participants who indicated that the story-based tasks supported a deeper understanding of the concepts through the process of writing dialogue and applying their understanding in concrete situations. Though most feedback from participants on learning statistics through stories was positive, some participants did indicate they would have preferred more direct support from the instructor when learning about the concepts through the stories.

Participants' learning of the discipline of statistics

This section discusses the second part of research question 1, which is the impact of the intervention on students' understanding of the discipline of statistics (as opposed to their understanding of statistics concepts previously discussed), with particular focus on their understanding of the usefulness of statistics. The findings indicate that the participants demonstrated understanding of the usefulness of statistics through their examples of the applications of statistics.

Understanding of the applications of topics occurred when participants provided appropriate examples of how the concepts and topics covered in the story-based tasks could be used in their everyday lives. Participants were significantly more successful understanding the applications of the topics of sampling techniques and descriptive statistics, and formal inferential statistics compared to the topics of informal inferential statistics and sampling distributions. Further, though participants could provide examples related either to their personal life or to their work, most participants provided examples related to their work. All participants who provided examples both at the beginning and at the end of the term demonstrated a positive shift in their examples. That is, they all provided more appropriate and specific examples of the usefulness of statistics at the end of the term compared to the beginning.

Few studies found in the literature focused on specific examples provided by students on the usefulness of statistics. Songsore and White's (2018) study is the most relevant in terms of examples of usefulness of statistics provided by participants. In both studies, participants understood applications related to sampling techniques and descriptive statistics, and formal inferential statistics. But the findings differ in terms of the types of examples participants provided. In Songsore and White's study, they found that most examples related to practical life skills and less on future career, which is the opposite of the results in this study. A possible reason for this difference could relate to the nature of the courses in which the studies were conducted. Songsore and White's study was conducted in a general first-year statistics course for social sciences, health sciences, science and business (p. 123) and, thus, would likely demonstrate examples across multiple disciplines. While in this study, as the course was specifically for business students, all examples of the usefulness of statistics provided in the intervention had connections to business. Thus, the use of specific business examples in this study may have impacted the types of examples chosen by participants.

The findings, then, suggest that the intervention can support the development of understanding of the discipline of statistics by supporting the development of relational understanding of core concepts and understanding of the usefulness of statistics. Further, the findings suggest that the intervention is effective at promoting the usefulness of statistics within work contexts, but less so in personal contexts. These findings are further supported by feedback from the participants who indicated that the intervention provided more specific and concrete examples of the usefulness of statistics for business purposes.

33. Impact of Intervention on Students' Belief about the Usefulness of Statistics

This section discusses the ***second research question*** regarding the impact of the intervention on participants' beliefs (thinking, viewpoints or opinion; Pajares, 1992) about the usefulness of statistics in their everyday lives (both personal life and work). The findings indicate that, in general, participants expressed positive beliefs about the usefulness of statistics related to their everyday lives both before and after the intervention. The few participants who expressed neutral views prior to the intervention expressed positive views after the intervention. Though positive beliefs about the usefulness of statistics to everyday life emerged, most participants appeared to interpret everyday life as referring to work. These findings suggest that the story-based tasks helped participants to maintain or positively improve their beliefs about usefulness of statistics.

These findings are in contrast to most studies on the impact of a statistics course on the usefulness of statistics, which found that students' beliefs about the usefulness of statistics can become more negative after taking a course in statistics (e.g., Kerby & Wroughton, 2017; Murtonen et al., 2008; Schau & Emmioglu, 2012). However, the findings do align with studies that examined the impact of context-rich problems on students' beliefs about the usefulness of statistics, which found that the students ended the term with favourable views about the usefulness of statistics (e.g., Dierdorp et al., 2011; Hiedemann and Jones, 2010). Thus, this study provides further evidence of the positive impact of context-rich problems (i.e., story-based tasks) on students' beliefs about the usefulness of statistics.

Another way in which participants demonstrated their beliefs was by their examples of situations, both before and after the intervention, that showed how they thought statistics would be useful to their everyday lives. These examples indicated positive change in participants' beliefs in that the examples were often non-specific and vague at the beginning while at the end of the intervention they became more specific and detailed.

The shifts in the participants' beliefs suggest that their initial beliefs were evidential (Green, 1971), that is, more susceptible to change through evidence, and that the intervention provided the participants with evidence that supported the change. For example, as they created more specific and appropriate examples at the end of the term of situations involving the usefulness of statistics, they became more aware of the usefulness that had a positive effect on their beliefs.

In summary, the findings suggest that learning statistics through stories has the potential to reinforce already positive, evidentially held beliefs about statistics. It is unclear how stories would impact someone with negative beliefs about the usefulness of statistics as no participant began the course with this belief. These findings are further supported by feedback from the participants who indicated that the story-based tasks supported their beliefs as they helped them develop a better understanding of the applications of statistics.

34. Features of Story-based Tasks that Supported Meaningful Learning

This section discusses the ***third research question*** in terms of the features of the story-based tasks that supported students' meaningful learning of selected statistical topics and the discipline of statistics, and reinforced students' positive beliefs about the usefulness of statistics. From the findings, three key features of the story-based tasks supported students' learning and beliefs: (1) the prompts, (2) the context, and (3) the characters.

The prompts

One feature of the story-based tasks that supported meaningful learning for the participants was the prompts. Each of the story-based task included prompts that were embedded in the story to guide students to think about particular aspects of the statistical topics and the discipline of statistics. For example, for the story that explored the topic of sampling distributions of sample means, the prompts included asking students to explain different aspects of the sampling distribution, which can be related to their development of basis (relational) understanding of the topic. This suggests that the prompts are important to the effectiveness of these tasks.

The findings indicate that the prompts in the story-based tasks impacted the students' learning of the selected statistical topics and the discipline of statistics by promoting types of relational understanding, such as contextual and basis understanding. Possible ways that the prompts supported the development of relational understanding include: (1) the nature of the prompts and (2) the placement of the prompts.

Nature of the prompts. The nature of the prompts refers to what the prompts invited the students to do, which played an important role in the students' learning. The findings suggest that the nature of the prompts supported students' learning of the statistical concepts in two ways: (1) type of understanding elicited in prompts and (2) type of dialogue elicited from the novice character.

Type of understanding elicited in prompts. The findings suggest that differences in the nature of the prompts could lead to different types of understanding demonstrated by participants and, in particular, the demonstration of basis understanding (a type of relational understanding, which involves explaining the reasoning behind a concept) or a lack of it. For example, in the case of the story-based tasks involving the topics of descriptive statistics, informal inferential statistics and confidence intervals, many prompts tended to elicit contextual understanding (another type of relational understanding, which involves interpreting statistical measures in a context). While for the story-based task involving sampling distributions of sample means, many prompts tended to elicit basis understanding. To illustrate, consider the prompt from the story-based task that explored descriptive statistics:

> Have at least two of the characters explain what the histograms and box plot mean in the context of the story. One of the characters should be Bart who will either ask a question or make an incorrect suggestion. Then either Jolene or Franca will answer the question or correct his misunderstanding.

The prompt does not ask the students what a histogram or box plot is or to explain the different information that the two different visual descriptive statistics would provide. Instead the focus is only on the synthesizing the statistical and the contextual, that is, on developing contextual understanding. Though there were prompts that asked students to generate statistical measures and to choose an appropriate measure and justify their choice, there were no prompts in this story-based task that would elicit basis understanding from the participants for this topic.

As another example, consider the prompt from the story-based task that explored informal inferential statistics:

> Have both characters discuss what you think this probability means in the context of the initial research question. Again, because Sam is new to this material (as you are), his primary role is to ask questions or to state his own misconceptions. Be as thorough and concrete as possible.

Again, the prompt does not ask what the probability represents in general or why it is the appropriate probability to examine. Instead it, again, focuses on interpretation. Though there were prompts that asked participants to explain aspects of informal inferential statistics, most focused on interpretation of the results within the context. Thus, many of the prompts in these story-based tasks would elicit contextual understanding.

In contrast, the prompts for the story-based task that explored sampling distributions of sample means asked students to explain the reasoning behind a concept, which resulted in a different type of relational understanding (i.e., basis understanding). For example, one prompt in the story-based task is

> Have Jed and Reema discuss what a sampling distribution is. In particular, comment on how the data in the sample and the data on the sampling distribution are different and how the process of collecting a sample is different from the process of creating a sampling distribution.

For this prompt, the focus is not on interpreting a sampling distribution but is instead on explaining the differences between them. Thus, the students were invited to consider why the two distributions were different. Though there were prompts that asked for interpretation of results and the generation of statistical measures, there were more prompts in this story-based task that asked students to delve into the reasoning behind the concepts compared to the other three story-based tasks. Though the other story-based tasks had the occasional prompt that could elicit basis understanding, this was only the story-based task that consistently and repeatedly asked students to consider the reasoning behind the concepts and it was the only story-based task where most participants demonstrated basis understanding. Further, all story-based tasks asked students to synthesize the statistical and contextual, and contextual understanding was the most common type of relational understanding that emerged from participants' work. Thus, there is evidence that suggests that a feature of the story-based task on the topic of sampling distributions of sample means that made it effective at supporting basis understanding is the nature of the prompts that were included in that task. Further, there is evidence to suggest that a feature of all the story-based tasks that made them effective at supporting contextual understanding was also the nature of the prompts.

Therefore, differences in the natures of the prompts relating to the type of understanding elicited appears to have an impact on the type of understanding demonstrated by the participants. This finding is similar to the results found by Smith (2014), which suggested that a feature of prompts that led to students to engaging in informal inferential reasoning and statistical thinking were that the prompts specifically asked students to do so. This is also consistent with the work of Elia et al.'s (2010) study, which suggested that for stories to generate learning, appropriate prompting is required. Further, this result is not dissimilar to what Kaplan and colleagues (2018) found when they examined whether changing the wording of the prompt impacted students' answers. But Kaplan et al. focused on how the prompts could "cue" the participants to provide a more complete description of a distribution (p. 97), while this finding focuses on how the prompts could "cue" the participants to focus on types of relational understanding of a concept. This suggests that this study adds to the field by providing further support of the importance of prompts in statistics education in supporting students' learning. Additionally, as the participants in Smith's and Elia et al.'s studies were elementary children, this study adds to the field by providing evidence of the importance of prompts in stories at the post-secondary level.

Type of dialogue elicited from the novice character. The findings also suggest that the differences in the nature of the prompts could lead to students demonstrating different levels of understanding based on how they personalized their dialogue through the novice character. For example, when the prompts asked the students to write a question for the novice character, the responses tended to result in personalization that did not elicit relational understanding. While the prompts that asked for the novice to state a misconception resulted in personalization that elicited relational understanding of the concept.

In contrast, when participants had the novice character state a misconception, they demonstrated relational understanding by considering not only how the concept applied to the situation but also how it did not apply. Therefore, there is evidence to suggest that the prompts that invited students to personalize their knowledge through writing misconceptions were more effective in supporting students' development of relational understanding than those that invited students to write questions for the novice characters.

Placement of prompts. In addition to the nature of the prompts, the placement of the prompts also contributed to students' learning of statistical topics by supporting the personalization of knowledge. For example, after the key concepts were explained in the instructor-provided portion of the story (i.e., the story-explanations), there were prompts that asked the participants to write their understanding of the story-explanation from the perspective of the novice. For these prompts, the findings showed that the participants personalized their knowledge by writing dialogue that differed from the story-explanation by making choices to use different language than the story-explanation, to focus on different aspects of the concept from the story-explanation, and choosing how much or how little to use the story-context. This suggests that the placement of prompts may be important in the effectiveness of the tasks to promote personalization of knowledge, which can support student learning.

To conclude regarding prompts, based on this study, for prompts to be effective in supporting students' learning of selected statistical topics and the discipline of statistics, they should have the following features: be written to support students' development of relational understanding of statistical topics and the discipline of statistics; be written to balance the types of understanding elicited to ensure different types of relational understanding are supported; be written to support students' personalization of knowledge; and be placed at key points in the story to support the development of relational understanding of key concepts.

The context

A second feature of the story-based tasks that supported meaningful learning for the participants was the context of the tasks. In addition to the prompts, the context of the stories also played a role in the students' meaningful learning. In particular, the context of the stories supported the development of relational understanding of the selected statistical topics and the solidification of students' positive beliefs about the usefulness of statistics. Each story had a unique context. For example, the story exploring descriptive statistics was about evaluating an inventory system, while the story exploring formal inferential statistics was investigating the feasibility of expanding a business. These contexts were relevant to the students' business program as they demonstrated ways that statistics could be used to explore and solve real-world business problems. The contexts also had relevance regarding real-world situations that brought the statistical concepts to life. For example, the abstract concept of sampling distributions was explored by seeing how it could be used to model the process of quality control.

These tasks, therefore, provided a type of context-rich problem to support students' learning. The importance of such problems is well established in the field of statistics education and supports the meaningfulness of stories to students' learning. The benefits of exploring concepts through context-rich problems has been supported by multiple research studies (e.g., Dierdorp et al., 2011; Nowacki, 2011). One of the defining aspects of statistics is the importance of context (Moore & Cobb, 2000). As such, many statistics educators argue that statistics should be taught using context-rich problems (e.g., Everson et al., 2016; Konold & Higgins, 2003).

The findings indicate that there is an important relationship between the context in the story-based tasks and the students' (1) learning of selected statistical topics, (2) learning of the discipline of statistics, and (3) beliefs about the usefulness of statistics.

Relationship between context and students' learning of selected statistical topics. The findings suggest a relationship between the nature of the contexts and students' learning, in particular, development of relational understanding of the statistical topics. For example, the context provided real-world meaning of the abstract statistical concepts that likely supported their development of basis understanding of the concepts (a type of relational understanding, which involves explaining the reasoning behind a concept) of the concepts. To illustrate, in their explanations of the reasoning behind the concepts, participants melded together the story-context and the abstract concept. For example, when explaining the abstract concept of how a sampling distribution is constructed, some participants utilized the story-context in their explanations to connect the process of constructing a sampling distribution to the problem of quality control for solar-powered scooters being investigated in the story. In doing so, the findings suggest that when participants did this, their explanations were more detailed and clearer compared to those who did not use the context in the explanation. That is, when participants balanced the abstract concepts with the concrete real-world context, the reasoning they presented behind the abstract concept was stronger and more complex than if they only focused solely on the abstract. Thus, the context has the potential to support students' development of basis understanding by providing real-world meaning to the abstract concepts.

Not only did the findings indicate that the nature of the context can support basis understanding, they also suggest that it can support contextual understanding, another type of relational understanding, which involves interpreting statistical measures in a context. For example, the context provided the participants with the opportunity to apply their understanding of the concepts in a real-world situation, which allowed them to demonstrate contextual understanding. To illustrate, the stories asked the students to apply their understanding of statistical concepts to explore and solve a problem. For example, throughout the story that explored descriptive statistics, participants used the context to interpret the statistical measures in the context of the problem, and used their interpretations to conclude whether the inventory system should be repaired. From the findings, contextual understanding was the most prevalent type of relational understanding demonstrated by the participants. Thus, through applying their understanding within a context, they had the opportunity to develop contextual understanding of the selected statistical topics. These findings are further supported by feedback from the participants who indicated that the intervention provided them the opportunity to apply the abstract statistical concepts in concrete situations.

Relationship between context and students' learning of the discipline of statistics. The findings also suggest a relationship between the nature of the context and students' learning of the usefulness of the statistical topics, which is an aspect of the discipline of statistics. For example, in the stories that had straightforward and obvious business contexts, participants could provide examples of how these topics would be useful to their everyday lives. While for stories that had tenuous connections to business or where the business context was complex, participants had difficulty providing appropriate examples of the usefulness of the topics. To illustrate, for the topic of descriptive statistics, the context of the story-based task was evaluating an inventory system for a company. The context provided a clear and easy to understand business problem: whether it was more cost-effective to fix the inventory system immediately or to wait. Further, the idea of doing inventory for a business is likely a context that most students either have direct experience with through their own jobs or at least can understand why it is useful to do properly. In the reflection task for this story, when participants were asked to provide examples of how this topic would be useful to them, they provided examples that were appropriate to the topic and involved the topic as a whole. Thus, through an easy to understand and relevant context, participants were able to see the usefulness of the concepts beyond the story, which can contribute to reinforcing their positive beliefs about the usefulness of statistics.

In contrast, for the topic of informal inferential statistics, the context of the story-based task was determining whether dolphins could communicate abstractly through experimentation. The context had a tenuous connection to business by motivating the reason for the experiments as part of a new marketing campaign for an aquarium. In the reflection task for this story, participants had difficulty providing appropriate examples to the topic. Instead, they focused on specifics of the topic (e.g., basic probabilities), but no participant provided an example of using informal inferential statistics to solve a problem either in their lives or in their future careers.

Thus, when the context had a tenuous connection to business, participants had difficulty understanding the usefulness of statistics beyond the story, which could make the context less effective at supporting students’ understanding of the discipline of statistics. This suggests that the relevance of the context may play a role in students’ ability to understand the applications of statistics, which could then make the context less effective at supporting learning about the discipline of statistics.

Relationship between context and students’ beliefs about the usefulness of statistics. The findings also suggest a relationship between the nature of the context and students’ beliefs about the usefulness of statistics. For example, the contexts and problems presented in the story-based tasks were examples of authentic real-world problems that required the use of statistics to address. As has been argued above, the findings suggest that participants' beliefs about the usefulness of statistics were evidentially held and, thus, can be swayed or solidified by evidence. Thus, the story-context has the potential to provide further evidence for students by providing specific examples of the usefulness of statistics. This is supported by the participants' examples of the usefulness of statistics becoming more specific and appropriate after the intervention compared to before. Further, participants indicated that the story-based tasks impacted their beliefs by highlighting that there were more uses of statistics than previously believed to their future careers and providing specific instances of how statistics can be used in their future careers. This suggests that a feature of the story-based tasks that support positive beliefs about the usefulness of statistics is the authentic nature of the problem and context.

To conclude regarding the context, this study suggests that for the story to be effective in supporting students' learning and development of relational understanding, development of understanding of the discipline of statistics, and development of positive beliefs about the usefulness of statistics, it should include an authentic real-world context in which to explore the abstract statistical concepts and provide the opportunity to apply the statistical concepts in a meaningful way. For the context to be effective in supporting understanding the discipline of statistics, the relevance of the context to the students' program of study needs to be clear and easy to understand. Thus, it may be important for the instructor to ensure that the class had established a common understanding of the story-context by relating the story-based tasks to the students own life experiences. For example, to start the discussion of story-based tasks, the instructor can ask students about experiences related to the stories such as doing inventory or seeing a dolphin show. It may also be important to make connections between the concepts in the stories and a broader context. For example, it would be beneficial to provide additional real-world examples of informal inferential statistics to evaluate evidence to help students see the usefulness of more abstract concepts.

These findings add to the growing literature on the benefits of learning statistics through real-world contexts. Similar to Pfannkuch's (2011) study, participants in this study utilized the context to provide real-life meaning to the abstract concepts. Dierdorp et al.'s (2011) study also found that students could connect their knowledge of the statistical concepts to understand a context-rich problem and that the use of authentic contexts had a positive impact on students' beliefs about the usefulness of statistics. This suggests that the study adds to the field by providing further evidence of the importance of authentic and rich contexts in statistics education in supporting students' learning and beliefs about the usefulness of statistics.

Further, from the literature review, Nolan and Temple Lang (2015) found that when the contexts were not accessible to the students it could have a negative impact on their learning. Thus, this study contributes to the field on context-rich problems by providing evidence of the importance of the accessibility of the context is not just for learning of statistical topics but is also important for supporting understanding of the discipline of statistics.

The characters

A third feature of the story-based tasks that supported meaningful learning for the participants was the characters of the stories. Through the different types of characters (i.e., expert and novice), the story offered students the opportunity to view the concepts from different perspectives that likely played a role in the learning and understanding of the selected statistical topics. In particular, all stories required the students to take on the characters of various expert and novice business professionals, which invited them to think about the concept from two different perspectives. For example, for the story-based task that explored sampling distributions of sample means, the participants were required to take on the role of a consultant who is well-versed in statistics (the expert) and a business owner who is unfamiliar with statistics (the novice). Through these characters, the participants explained concepts such as sampling variability and the central limit theorem from the perspective of an expert explaining the concepts to a novice and from the perspective of a novice new to the concept. From the perspective of the expert talking to the novice, in general, the participants tended to use informal language, provide examples, and inter-weave the story-context into their explanations of the statistical concepts. From the perspective of the novice, in general, the participants asked questions about the concept and stated their understanding of the concept (which included both demonstrating understanding and purposefully stating misconceptions). By adopting the perspectives of these two types of characters throughout the four story-based tasks, the participants personalized their knowledge and demonstrated relational understanding of the statistical concepts. This suggests that incorporating such characters in the story-based tasks allowed the participants to explain their understanding of the concepts in a way that makes sense to someone else, which results in them considering the concepts relationally.

The findings also suggest that the personality of the characters in the stories can impact the levels of understanding demonstrated by the students. For example, when the novice character was presented as lazy and uninterested in statistics and the expert characters were presented as frustrated with this, the participants interpreted the role of the expert as providing minimal information to the novice with no consideration of whether he/she understood the statistical situation. This resulted in the participants demonstrating instrumental or minimal relational understanding of the statistical concepts.

In contrast, when the expert characters were presented as caring and they wanted the novice character to learn the concepts and the novice characters were interested and engaged in the learning, the participants interpreted the role of the expert to be a teacher and the role of the novice to be a motivated student. For example, in one case, through the expert role, the participant explained the concept to the novice, then checked the novice's understanding and re-explained parts of the concept the novice did not understand. In another case, the participant used the expert role to explain the concept to the novice and, in the novice role, the novice re-stated their understanding in their own words. In these cases, they demonstrated strong relational understanding by demonstrating basis, contextual or choice understanding of the concepts.

This consideration of engaging students in multiple perspectives of viewing the statistical concept forms a central part of the nature of the story-based tasks that is unique to this study. Though there are other studies that use stories in statistics education (e.g., D'Andrea and Waters, 2002; Sherwood, 2018), within the literature review, there were none that required the students to consider the concepts from multiple perspectives. As this study has demonstrated possible benefits of multiple perspectives regarding students' development of relational understanding of statistical concepts, this study contributes to the field a different and meaningful way to engage students in learning and understanding statistical concepts through the use of story-based tasks.

To conclude regarding the impact of the story characters on students' understanding of selected statistical topics, the nature of the story characters that allowed for more meaningful consideration of the concepts from different perspectives was important for their development of relational understanding of the concepts. This occurred when there were at least two types of characters (a novice and an expert) and the personalities assigned to the characters of the stories are written to promote cooperation between the two types of characters.

35. The Story-Based Tasks and the Reform Movement in Statistics Education

In the previous sections, the findings were summarized as related to the three research questions. In this section, how the story-based tasks align with the learning goals of the reform movement in statistics education are considered. The reform movement in statistics education calls for instructors to focus on the learning goals of statistical knowledge, reasoning, and thinking (Ben-Zvi & Garfield, 2004). Thus, this section examines whether the story-based tasks were effective at achieving the goals of statistical knowledge, reasoning and thinking.

Statistical knowledge is defined within this study as the basic skills and knowledge needed to engage with statistics, which includes the ability to produce descriptive measures (e.g., making graphs, finding measures of centre and variation), and have the knowledge of the terminology, concepts and symbols used in statistics to effectively communicate statistically (Ben-Zvi & Garfield, 2004; Gal & Garfield, 1997).

The findings indicate that the intervention supported participants' development of both aspects of statistical knowledge, collectively, based on their work with the intervention (in particular, with the story-based and follow-up tasks). Participants demonstrated the skills needed to perform statistics when participants found various statistical measures, that is, when they demonstrated algorithmic understanding. Participants demonstrated knowledge of the terminology needed to perform statistics occurred when participants wrote definitions of terms in their own words, that is, when they demonstrated terminology understanding. As was described above, participants demonstrated both algorithmic and terminology understanding of the selected topics. This suggests that the intervention was effective at supporting the learning goal of statistical knowledge.

Statistical reasoning, in this study, is defined as the ability to interpret the results within the context; to understand how and why statistical processes work; to justify and explain the reasons behind choices made in the statistical investigations; and to understand how the choices made relate to the conclusions drawn (Ben-Zvi & Garfield, 2004; delMas, 2004).

The findings indicate that the intervention supported participants' development of the first three of the four aspects of statistical reasoning, collectively, based on their work with the intervention (in particular, with the story-based and follow-up tasks). The ability to interpret results occurred when participants interpreted statistical measures within the context of the story, that is, when they demonstrated contextual understanding. The ability to justify and explain the reasons behind choices made in statistical investigations occurred when participants chose between various statistical measures and justified their decision, that is, when they demonstrated choice understanding. The ability to understand how and why statistical processes work occurred when participants explained the reasoning behind statistical concepts, that is, when they demonstrated basis understanding. Yet, the ability to understand how the choices made relate to the conclusions drawn was not demonstrated by the participants in this study. As was described above, participants demonstrated contextual and choice understanding for the selected topics, while basis understanding was only demonstrated for specific concepts. Thus, the findings suggest that the intervention was effective at supporting most aspects of the learning goal of statistical reasoning.

Statistical thinking in this study, is defined as the thought processes used by statisticians when engaging in statistical investigations which include understanding the importance of properly collected data; that different representations of the data foster understanding (transnumeration); that data are numbers in a context; that variation is everywhere and in everything; that chance and probability play a role in statistical analysis; the purpose, logic and process of statistical investigations; that data are better than anecdotes; and the limitations of statistics (Cobb, G. W., 1992, 2007a; Wild & Pfannkuch, 1999).

The findings indicate that the intervention supported participants' development of the first five of the eight aspects of statistical thinking, collectively, based on their work with the intervention. The need for properly collected data occurred when participants correctly explained why a sampling technique would result in a good sample and the limitations of conclusions from poorly collected data. Transnumeration occurred when participants correctly used various statistical measures to provide meaning to the raw data. Understanding that data are numbers in a context occurred when participants demonstrated contextual understanding. The omnipresence of variation occurred when participants correctly recognized the impact of both variation within a sample and sampling variability (also called variation between samples) in their statistical analysis. The role of chance in statistical conclusions was demonstrated when participants recognized that conclusions were 'likely' rather than 'proven'.

Though five aspects of statistical thinking emerged from the participants' work, some aspects were more prevalent than others. Most participants demonstrated statistical thinking by demonstrating that they understood the need for properly collected data, transnumeration, that data is numbers in a context, and importance of sampling variability in statistical analysis. Yet, few participants demonstrated understanding of the importance of variation within a sample when doing statistical analysis and few participants recognized the role of chance in conclusions. Further, three aspects of statistical thinking related to the overarching ideas of statistical investigations (the purpose, logic and process of statistical investigations; that data are better than anecdotes; and the limitations of statistics) did not emerge from the participants' work.

Thus, the findings suggest that the intervention was effective at supporting many aspects of the learning goal of statistical thinking.

To conclude, the findings suggest that the intervention can support the development of aspects of statistical knowledge, reasoning, and thinking. But there is no evidence to suggest that intervention supports the development of aspects of statistical thinking related to the big picture ideas of statistical investigations. That is, they were least effective at promoting students' understanding of the purpose, process, logic and limitations of statistical investigations (Wild & Pfannkuch, 1999). Thus, the story-based tasks align with many aspects of the reform movement in statistics education in relation to the learning goals, but does not align with all of them.

Previous research suggests that statistics educators continue to have difficulties successfully converting the reform movement's recommendations into practice (Tishkovskaya & Lancaster, 2012) and many innovations in statistics education are based on intuition rather than on evidence (Ramirez et al., 2012). Thus, this study adds to the field by providing an example of an intervention that has successfully implemented the reform movement's recommendations into practice. Further, the effectiveness of the intervention at promoting the learning goals of the reform movement are based on evidence rather than intuition.

36. The Story-Based Tasks and the Emergent Perspective of Constructivism

The findings suggest that the story-based tasks helped students engage in aspects of the learning process as outlined by the emergent perspective of constructivism (Cobb & Yackel, 1996, p. 186). These include actively constructing knowledge using prior knowledge, developing relational understanding, and negotiating meaning through social interactions.

A central tenet to constructivism is that new knowledge is actively constructed using prior knowledge (rather than passively received) from experiences (Merriam et al., 2007). The findings suggest that the story-based tasks can lead to students actively constructing their understanding of the statistical topics and that they do so using prior knowledge. For example, when participants wrote dialogue that were different from their peers and from the instructor-provided resources, this suggests that they were actively engaging in the construction of their *own* understanding of the concepts rather than passively receiving the information. Further, when participants personalized their knowledge by choosing language that was unique to them, this suggests participants were connecting the new experiences in the classroom to their prior knowledge. The choice of language suggests that participants were taking the new terms and concepts, and were connecting them to terminology they were already familiar with. For example, when defining outliers, students used the words “diverged”, “abnormal distance”, “extreme”, “far” to describe what an unusual value is. This suggests that they were connecting the idea of an “unusual value” to terminology they were more familiar with. Thus, the findings suggest that the story-based tasks were effective in supporting students to actively construct their knowledge using prior knowledge to understand the new concepts and topics covered in the story-based tasks.

The learning theory of constructivism arose in part out of concerns that learning strategies based on behaviourism resulted in students who had only an instrumental understanding of course concepts and not a deeper, relational understanding (Confrey & Kazak, 2006). The findings suggest that the story-based task support the development of relational understanding of the concepts. For example, participants demonstrated the ability to understand the concepts within a context (i.e., contextual understanding), to make choices between different statistical measures and to justify those choices (i.e., choice understanding), and to explain the reasoning behind the concepts (i.e., basis understanding). Thus, as has been previously noted, participants collectively demonstrated relational understanding of most of the statistical topics and concepts covered, based on their work with the intervention. This suggests that the story-based tasks are effective at supporting students' development of a relational understanding of statistical topics.

Within the emergent perspective, social interaction is important for learning as it allows students to negotiate meaning through mutual adaptation (Cobb & Yackel, 1996). Due to ethical concerns, observations of the students while they wrote the dialogue in class was not done. Therefore, there was no evidence collected regarding the effectiveness of the intervention at stimulating social interactions that allow for the negotiation of meaning. But the story-based tasks themselves invite the students to write dialogue between characters and, thus, can act as simulated social interaction. The findings suggest that writing dialogue between characters for the story-based tasks can lead students to negotiate meaning of the statistical concepts. For example, when participants personalized their knowledge by choosing to have the characters engage with each other extensively, the written dialogue simulated the negotiation of meaning. In particular, participants had the novice character ask questions to clarify understanding or to address different aspects of the topic, state misconceptions about the concepts that the expert character addressed, and state their own understanding of the concept after the expert character had explained the concept. Thus, the two types of characters engaged in "a process of mutual adaptation wherein individuals negotiate meanings by continually modifying their interpretations" (Cobb, P., 1994, p. 14). The findings suggest, then, that the story-based tasks have the potential to provide students with the opportunity to negotiate meaning through writing dialogue.

To conclude, the story-based tasks have the potential to be effective at supporting students' active construction of knowledge of selected statistical topics. In addition, the story-based tasks have the potential to be effective at supporting students' development of relational understanding of the topics. Finally, the story-based tasks have the potential to be likely effective at providing students with the opportunity to engage in negotiation of meaning through writing dialogue between characters. Thus, taking these findings together, this suggests that the story-based tasks align with the emergent perspective of constructivism.

Chapter 6 - Conclusion

This study set out to examine the impact of an intervention on post-secondary students' understanding of statistical topics, their understanding of the discipline of statistics, their beliefs about the usefulness of statistics, and what features of the stories supported meaningful learning. The findings indicate that the intervention, which consisted of story-based, reflection and follow-up tasks, can be effective at supporting students' understanding of statistical topics and the discipline of statistics, and can solidify already positive beliefs about the usefulness of statistics. In particular, the intervention was most successful at supporting participants'

- development of instrumental understanding of all four selected topics (i.e., descriptive statistics, informal inferential statistics, sampling distributions of sample means and confidence intervals);
- development of aspects of relational understanding (such as contextual understanding) of most aspects of the selected topics;
- development of understanding of the usefulness of statistics;
- personalization of knowledge as part of the process of developing understanding;
- reinforcing already positive beliefs about the usefulness of statistics;
- achievement of the learning goals of the reform movement in statistics education of statistical knowledge, reasoning and thinking;
- active construction of their understanding using prior knowledge, and;
- negotiation of meaning through writing dialogue for characters.

Features of the story that supported meaningful student learning and positive beliefs about the usefulness of statistics include the prompts, the context and the characters. The prompts contributed to the findings by inviting students to consider certain aspects of the statistical topics. The context contributed to the findings by providing a meaningful and concrete situation to explore the statistical concepts, examples of authentic real-world applications of statistics, and providing students with further evidence to support their positive beliefs about statistics. The characters in the story contributed to the findings by presenting students with a situation where they needed to consider multiple perspectives of the concept, which contributed to the development of their understanding of the concepts and supported the personalization of their knowledge.

The reform movement in statistics education suggests that courses in statistics should promote statistical knowledge, reasoning and thinking. From the findings, the intervention was effective at supporting statistical knowledge and aspects of statistical reasoning and thinking. The story-based tasks were least effective at promoting aspects of statistical thinking that relate to understanding the big picture ideas of statistical investigations.

Finally, through the intervention, participants had the opportunity to actively construct their understanding, develop relational understanding of concepts, and engage in the negotiation of meaning. Thus, the intervention aligns with the emergent perspective of constructivism, which is the primary learning theory related to this study.

The remainder of this chapter outlines the implications of the study for the reform movement in statistics education, the nature of understanding, the nature of stories, teacher practice, and teacher education. Then the limitations of the study and suggestions for future research are presented.

37. Implications for the Reform Movement in Statistics Education

One of the reasons for this study was to add to the research on how to effectively support students' understanding of statistical concepts and promote positive beliefs about the usefulness of statistics using suggestions of best practices from the reform movement. Thus, it is important to consider whether the findings support these suggestions of best practice provided by the reform movement in statistics education.

The intervention in this study was designed using the best practices suggested in the GAISE report (Guidelines for the Assessment and Instruction in Statistics Education; Everson et al., 2016). As this study suggests that the intervention was successful at supporting most of the learning goals of the reform movement, there is evidence that the best practices for teaching suggested by the reform movement can provide guidance on how to reform a statistics course to achieve the learning aims of statistical knowledge, reasoning and thinking. To illustrate, the suggested best practices of focusing on conceptual understanding influenced the inclusion of prompts in the story-based tasks that invited students to explain the “why” behind statistical concepts. As was discussed in Chapter 5, there is evidence that the nature of the prompts supported students’ meaningful learning of statistics. Further, the suggestion of designing assessments that support learning influenced the creation of the threefold nature of the intervention (i.e., the story-based, reflection and follow-up tasks). This was done in an effort to support student learning by providing students with the opportunity to consider the statistical concepts more than once and in different contexts. As was outlined in Chapter 4, there is evidence that the participants could adapt their understanding from one context (the story-based tasks) to a new context (the follow-up tasks). Thus, there is evidence in this study that appears to validate the best practices suggested by the reform movement in statistics education.

Having said that, I would also argue that this study provides evidence that the best practices have room for improvement. In particular, there is one key aspect of the intervention that did not come from the best practices and, yet, was important for student learning: the writing of dialogue between two types of characters. This feature came from addressing the deficiencies that arose from pilot of the intervention. As was discussed in Chapter 5, the nature of the characters invited the students to consider the statistical concepts from multiple perspectives and to personalize their knowledge, which appears to have supported meaningful learning of the concepts. That is, writing about statistics in a meaningful way may contribute to learning. In this regard, the reform movement in statistics education could learn from the calculus reform movement by adding to their best practices the importance of writing *about* statistics. Though GAISE recommends that students write reports and minute papers as different types of assessment (Everson et al., 2016), they do not appear to highlight the importance of writing about statistics by explaining the concepts in detail or by considering how to explain the interpretation or concept to someone new to statistics. This study suggests that this practice may be beneficial in supporting learning.

In short, this study appears to validate the current suggested best practices in the reform movement. But it also suggests that these best practices can be improved to include the importance of writing about statistics.

38. Implications for the Nature of Understanding

Skemp's (1976/1978) definition of understanding broke understanding into two types: instrumental and relational. This study expands on this interpretation of understanding by providing additional sub-categories. In particular, instrumental understanding can be understand to involve *algorithmic* understanding (follow the steps to arrive at an answer) and *terminology* understanding (state definition in own words, without understanding the reasoning behind the definition). Further, relational understanding can be understood to involve *contextual* understanding (applying understanding in a context), *choice* understanding (choosing between different methods and justifying that choice), and *basis* understanding (explaining the reasoning behind a concept).

Though this study focused on statistics, these additional interpretations can be applied more broadly to mathematics education in general. For example, in calculus, students can demonstrate algorithmic understanding of derivatives by following steps to find a derivative. Further, they can demonstrate choice understanding by choosing between different integration techniques and justifying why the choice would be appropriate.

Thus, researchers may find this expanded interpretation of Skemp's (1976/1978) instrumental and relational understanding framework useful in their research of students' understanding of mathematics.

39. Implications for the Nature of Stories

As was noted earlier, the stories in this study do not fit nicely into any of the previous definitions of stories in mathematics education. For example, Zazkis and Liljedahl (2009) provided six ways in which stories can be used in mathematics education, but in each of these versions, the teacher presents a complete story to the students (story as teaching tool; Roberts & Stylianides, 2013). Though some of these stories may require a response from the students, it is after the story is told by the instructor and not during the story (e.g., Albano & Pierri, 2016). When students were asked to engage in story-telling, they were asked to create their own story from scratch (e.g., Cho et al., 2015; Sherwood, 2018). In contrast, the stories in this study combine these two perspectives and require the teacher and student to contribute to its construction. Thus, the stories in this study extend our understanding of stories for learning by presenting the idea of *interactive stories*.

Using the stories in this study as a blueprint, interactive stories are incomplete short stories written by the instructor (or someone who is not a student) and require the students to complete them as a basis for their learning. The interactive stories were used in this study to engage students in learning statistics concepts, but they have the potential to be used for other mathematics concepts or in other disciplines. In general, these are stories that are left intentionally incomplete at key points, with these points chosen to engage students in making sense of important (mathematical) concepts presented in the story, applying the concepts to address the problem presented in the story, or any other place where the writer of the story believes that students may benefit from actively interacting with the story. In short, interactive stories are designed to invite the students to insert themselves (their thinking) into the story and, thus, become co-authors with the teacher in co-constructing the story as a way to support their learning. Thus, an implication of this study is the extension of the understanding of story in education, in general, and mathematics education, in particular, to include interactive stories. More details on what aspect of the interactive stories may benefit student learning in mathematics education are presented in the next section.

Implications for Teacher Practice

As was noted by Ramirez and colleagues (2012), many innovations in statistics education are based on intuition rather than research. Thus, it is important to discuss the results of this study on an innovation in the context of teaching practice. Based on the results, there is evidence that the use of story-based tasks aligns with the learning goals of the reform movement in statistics education and can result in both instrumental and relational understanding of key statistical topics. In particular, there is evidence that story-based tasks can support students' development of statistical knowledge, most aspects of statistical reasoning and some aspects of statistical thinking. Further, they can support the development of instrumental understanding and aspects of relational understanding such as contextual and choice understanding. Finally, the story-based tasks can support students' already positive beliefs about the usefulness of statistics

Yet the story-based tasks did not support all aspects of learning. In particular, though students' developed basis understanding for certain topics (e.g., sampling distributions), there was less evidence that they developed basis understanding across all key statistical topics. Finally, there is not enough evidence to suggest that the story-based tasks were effective at supporting students' learning of the big picture of statistical investigations.

Thus, statistics educators who are interested in students developing relational understanding of the role of contexts in statistical investigations and aspects of statistical reasoning may find story-based tasks useful in supporting this learning. While statistics educator who are interested in students developing a big picture understanding of statistical investigations may not find story-based tasks as useful.

For statistics educators who are considering using stories in their courses, there is evidence to suggest that the stories in this study have addressed the two major problems found in the pilot study. In the pilot of the story-based tasks, students lost the context of the story in their responses and had difficulty connecting what they learned in the stories to the course content. To address the first problem, the stories in this study were changed to have more accessible contexts to the students and to have the prompts embedded in the story. Based on the results of this study, it appears that these changes were successful in addressing the problem of loss of context. For example, in this study, students utilized the context to reinforce their explanations of statistical concepts. To address the second problem, there were less stories in this study (four compared to seven) and two types of stories were used that provided students with the opportunity to consolidate their previous learning (comprehensive stories) or to motivate the learning of new concepts (introductory stories). In this study, students who demonstrated relational understanding in a story were likely to adapt that understanding to a different context and task. This suggests that the stories in this study provided students with the opportunity to connect the content in the stories to the course content. In summary, there is evidence to suggest that the story-based tasks in this study addressed the two major problems that arose from the pilot of the stories. Thus, educators who are considering using stories to teach statistics may find the framework for the stories in this study more useful than the one used in the pilot of the stories.

Further, for statistics educators who would like to develop their own story-based tasks, this study provides evidence of certain features of these tasks that were most effective at supporting learning. In particular, the nature of the prompts used in the story-based tasks appear to play an important role in developing understanding. For example, the nature of the prompts can lead to the development of certain types of understanding. Thus, educators would be advised to write multiple types of prompts that support multiple types of understanding. Further, the nature of the context also played a role in learning. In particular, educators are advised to choose authentic contexts that have a clear connection to the students' lives and future careers to ensure both the development of understanding and positive beliefs about the usefulness of statistics. Finally, the characters in the story also play a role in supporting learning. For example, having both an expert and novice character appears to support students in considering multiple perspectives of the statistical concepts. Thus, when writing story-based tasks, considering how the characters would interact and their level of expertise can be important in supporting learning.

40. Implications for Teacher Education

Though not described in detail in this study, the process of writing the stories by myself and a colleague resulted in our deeper understanding of the content covered in the stories and better pedagogical understanding of how to present the content. To illustrate, through writing the first story in which the plot follows a statistical investigation, I had to carefully consider the steps in a statistical investigation, which helped me unpack my implicit process. Further, the consideration of how these steps should unfold, helped me scaffold the process for the students through the plot of the story. As another example, for the two stories that introduced concepts, while writing the stories my colleague and I actively considered portions of the concept that students had difficulty with in the past and wrote these difficulties into the story in an attempt to address them.

Based on this, the stories in this study have implications for teacher education. In particular, the writing of story-based tasks on a specific topic has the potential to engage pre-service teachers in developing a better understanding of the topic. Further, encouraging pre-service teachers to actively consider potential areas of difficulty in understanding the topic, they can begin to develop strategies on how to effectively address these difficulties. Lastly, by considering where it is appropriate for the students to 'take over' the story (i.e., prompt students to write dialogue), pre-service teachers can begin to determine when students can benefit from the opportunity to consider concepts and applications on their own. In short, my experience with the study suggests that writing stories has the potential to help pre-service teachers develop an understanding on how to effectively present a topic. Thus, the writing of story-based tasks may be an effective tool of supporting pre-service teachers' development of both content and pedagogical knowledge.

41. Limitations of Study

Though this study suggests that the intervention has the potential to enhance students' understanding of statistics and their beliefs about the usefulness of statistics, there are at least five limitations for this study. The first limitation is that the data for this study was collected only for one class, in one term. This makes it difficult to generalize the results beyond the participants in the study. However, since case study methodology calls for naturalistic generalizations (Stake, 1995), it is possible for readers to learn and gain insight from the case.

The second limitation relates to data collection. In particular, due to ethical considerations, no data was collected on the instructor's interactions with the students. There was only data collected on what the instructor said during whole group discussions on the first day each of the four story-based tasks were introduced. Thus, the study does not provide sufficient insight on how the instructor's action impacted students' learning through the intervention. This is one area that future research should consider to further extend the findings of this study.

The third limitation is that, since most of the data analyzed for this study was group work, the understanding of individual participants for the selected topics and the discipline of statistics is unclear. Further, due to this limitation, there were instances where data from individual participants and from groups were compared. However, since this was an exploratory study regarding the level of understanding the story-based tasks could support, the findings were less about individuals and more about the group of participants as a unit, with the individual and group work/assignments providing examples of the levels of understandings within the unit. Future studies could be designed to address individual and group work independently.

The fourth limitation is that all of the story-based tasks focused on a business context. Therefore, it is unclear if the intervention would be effective in other disciplines at promoting beliefs about the usefulness of statistics. Thus, this is another meaningful area for future studies to investigate.

The fifth and final limitation of the study is the extent to which the story-based tasks played a role in supporting students to develop their understanding. Within the study, there was no data collected on how the students determined what to write in their dialogue. For example, there were no direct observations of the students engaging with the stories in the classroom. Further, there was no data collected on the extent of outside resources (e.g., textbook, peers, internet) used to create the dialogue, which makes the extent of the impact of the story of students' understanding unclear. In addition, there was no data collected on participants' prior knowledge of the course material. Therefore, it is not clear if the participants who demonstrated understanding had prior knowledge of these concepts. In short, though participants demonstrated understanding of the concepts through the story-based tasks, it would be inappropriate to suggest that the story-based tasks were the sole cause. The implication is that a research design involving an experimental and control group or a pre- and post-test design would provide more reliable evidence of the impact of the intervention and establish a cause-effect relationship. However, as previously explained, impact in this study was not about cause-effect but the nature of the understanding students demonstrated from engaging in the intervention.

42. Future Research

From this study, there is evidence that students can develop their understanding through story-based tasks, but there are aspects of their understanding and learning that were not explored. This suggests that further research into student learning through such tasks is needed.

The story-based tasks elicited understanding from the students and their knowledge was personalized. Yet from this study, what aspects of the story enabled students to do this can only be theorized. Thus, further research might explore how writing dialogue for an expert and novice character within a story contributes to students' understanding of statistical concepts.

Within this study, no data was collected on whether participants understood the big ideas of statistical investigations. Thus, it is unclear if the stories as models of statistical investigations aided students in the learning these big ideas. Further research could explore this deficit.

Participants demonstrated strong relational understanding when they had the two characters engage in a meaningful dialogue about the concept where the expert character acted as a teacher and the notice character as a motivated student. Yet this study does not investigate why this was the case. Further analysis of this would be beneficial. For example, for the story-based tasks, can prompts be written to promote this type of dialogue? If so, how does that impact the understanding demonstrated by the participants? Further, as students engaged in group work to write the dialogue, how do the group dynamics impact the dialogue they wrote?

Further research into students' areas of difficulties with the stories would also be important. For example, examining what students understood from the prompts and what aspects of the concepts were difficult for them to learn through the stories would be beneficial. Related to this, as two of the stories had students learning about the concepts through reading the stories, it would be beneficial to determine what aspects of the story-explanations helped or hindered their learning. This research would be very beneficial in improving the stories in the future and for determining how to use the stories beyond the specific context of a first-year business statistics course.

Finally, the story-based tasks in this study all used business contexts. Thus, it is unclear whether story-based tasks that are grounded in different contexts would result in the same learning and beliefs about the usefulness of statistics. Further research in story-based tasks using different, non-business contexts is warranted.

References

Albano, G., & Pierri, A. (2016). Digital storytelling in mathematics: A competence-based methodology. *Journal of Ambient Intelligence and Humanized Computing*, *8*(2), 301-312. doi: 10.1007/s12652-016-0398-8

Ambrose, S. A., Bridges, M. W., DiPietro, M., Lovett, M. C., & Norman, M. K. (2010). *How learning works: 7 research-based principles for smart teaching*. San Francisco, CA: Jossey-Bass.

American Statistical Association. (n.d.). ASA Newsroom. Retrieved February 15, 2019 from https://www.amstat.org/ASA/Newsroom.aspx

Anderson, A., Anderson, J., & Shapiro, J. (2004). Mathematical discourse in shared storybook reading. *Journal for Research in Mathematics Education*, *35*(1), 5-33.

Arnold, P., & Pfannkuch, M. (2014). Describing distributions. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Balakrishnan, C. (2008). *Teaching secondary school mathematics through storytelling* (Master's thesis, Simon Fraser University). Retrieved from http://www.peterliljedahl.com/wp-content/uploads/Thesis-Chandra-Balakrishnan.pdf

Beebe, S. A., Beebe, S. J., Redmond, M., & Geerinck, T. M. (2011). *Interpersonal communication: Relating to others* (5th Canadian Ed.). Toronto, ON: Pearson.

Ben-Zvi, D., & Garfield, J. (2004). Statistical literacy, thinking and reasoning: Goals, definitions and challenges. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical* *literacy, reasoning and thinking* (pp. 3-16). New York, NY: Kluwer Academic Publishers.

Blackburn, G. (2016). Effectiveness of eLearning in statistics: Pictures and stories. *E-Learning and Digital Media*, *12*(5-6), 459-480. doi: 10.1177/2042753016653704

Brown, E. C., & delMas, R. (2018). Developing students' causal understanding of sampling variability: A design research study. In M. A. Sorta, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018

Budgett, S., & Wild, C. J. (2014) Students' visual reasoning and the randomization test. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Burrill, G. (2005). Curriculum issues in statistics education. *The Mathematics Education into the 21st Century Project: Reform, Revolution and Paradigm Shifts in Mathematics Education* (pp. 59–64). Johor Bahru, Malaysia. Retrieved from http://math.unipa.it/~GRIM/21_project/21_malasya_burrill59-64_05.pdf

Capraro, R. M., & Capraro, M. M. (2006). Are you really going to read us a story? Learning geometry through children's mathematics literature. *Reading Psychology, 27*(1), 21-36. doi: 0.1080/02702710500468716

Canal, G. Y., & Gutiérrez, R. B. (2010). The confidence intervals: A difficult matter, even for experts. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Carlson, K. A., & Winquist, J. R. (2011). Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *Journal of Statistics Education, 19*(1), 1-23. Retrieved http://tandfonline.com/toc/ujse20

Carnell, L. J. (2008). The effect of a student-designed data collection project on attitudes toward statistics. *Journal of Statistics Education*, *16*(1). Retrieved from http://tandfonline.com/toc/ujse20

Carter, K. (1993). The place of story in the study of teaching and teacher education. *Educational Researcher*, *22*(1), 5-12.

Casey, B., Erkut, S., Ceder, I., & Mercer Young, J. (2008). Use of a storytelling context to improve girls' and boys' geometry skills in kindergarten. *Journal of Applied Developmental Psychology*, *29*(1), 29-48. doi: 10.1016/j.appdev.2007.10.005

Castro, E. S., Garcia, V. N., & Sanchez, E. (2018). Introduction of hypothesis testing through an informal inference approach in a Mexican high school. In M. A. Sorta, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018

Chan, S. W., & Ismail, Z. (2013). Assessing misconceptions in reasoning about variability among high school students. *Procedia-Social and Behavioral Sciences, 93*, 1478-1483. doi: 10.1016/j.sbspro.2013.10.067

Chance, B., delMas, R., & Garfield, J. (2004). Reasoning about sampling distributions. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning,* *and thinking* (pp. 295-323). Dordrecht, The Netherlands: Kluwer.

Chaphalkar, R., & Leary, C. (2014). Introductory statistics students' conceptual understanding of variation and measures of variation in a distribution. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Cho, H. (2012). *The use of cartoons as a teaching tool in middle school mathematics* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3517272)

Cho, H., Osborne, C., & Sanders, T. (2015). Classroom experience about cartooning as assessment in pre-service mathematics content course. *Journal of Mathematics Education at Teachers College, 6*(1), 45-53.

Clark, M. C., & Rossiter, M. (2008). Narrative learning in adulthood. *New Directions for Adult and Continuing Education, 119*, 61-70. doi: 10.1002/ace.306

Cobb, G. W. (1992). Teaching statistics. In L. A. Steen (Ed.), *Heeding the call for change: Suggestions for curricular action*. (pp. 3-34). Washington: Mathematical Association of America.

Cobb, G. W. (2007a). One possible frame for thinking about experiential learning. *International Statistical Review, 75*(3), 336–347. doi:10.1111/j.1751-5823.2007.00034.x

Cobb, G. W. (2007b). The introductory statistics course: A Ptolemaic curriculum? *Technology Innovations in Statistics Education, 1*(1), 1-15. Retrieved from http://eprints.cdlib.org/uc/item/6hb3k0nz

Cobb, G. W., & Moore, D. S. (1997). Mathematics, statistics, and teaching. *The American*

Mathematical Monthly, 104(9), 801-823.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher, 23*(7), 13-20.

Cobb, P., & Yackel, E. (1996). Constructivist, emergent and sociocultural perspectives in the context of developmental research. *Educational Psychologist, 31*(3/4), 175-190.

Cohen, L., Manion, L., & Morrison, K. (2011). *Research methods in education* (7th ed.). New York, NY: Routledge.

Confrey, J., & Kazak, S. (2006). A thirty-year reflection on constructivism in mathematics education in PME. In A. Gutiérrez & P. Boero (Eds.), *Handbook of research on the Psychology of Mathematics Education: Past, present and future* (pp. 305–345). Rotterdam, The Netherlands: Sense Publishers.

Cooper, L. L., & Shore, F. S. (2008). Students' misconceptions in interpreting center and variability of data represented via histograms and stem-and-leaf plots. *Journal of Statistics Education, 16*(2), 1-13. Retrieved from http://tandfonline.com/toc/ujse20

Creswell, J. W. (1994). *Research design: Qualitative and quantitative approaches*. Thousand Oaks, CA: Sage Publications.

Creswell, J. W. (2013). *Qualitative inquiry and research design: Choosing among five approaches* (3rd edition). Thousand Oaks, CA: Sage Publications.

D'Andrea, L., & Waters, C. (2002). Teaching statistics using short stories: Reducing anxiety and changing attitudes. In B. Phillips (Ed.), *Developing a Statistically Literate Society. Proceedings of the Sixth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/

Dabos, M. (2014). A glimpse of two-year college instructors' understanding of variation in histograms. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

delMas, R. C. (2002). Statistical literacy, reasoning and learning: A commentary. *Journal of Statistics Education, 10*(3). Retrieved from http://tandfonline.com/toc/ujse20/

delMas, R. C. (2004). A comparison of mathematical and statistical reasoning. In D. Ben-Zvi & J. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 79-95). New York, NY: Kluwer Academic Publishers.

delMas, R. C., Garfield, J., Ooms, A., & Chance, B. (2007). Assessing students' conceptual understanding after a first course in statistics. *Statistics Education Research Journal, 6*(2), 28-58. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Dierdorp, A., Bakker, A., Eijkelhof, H., & van Maanen, J. (2011) Authentic practices as contexts for learning to draw inferences beyond correlated data. *Mathematical Thinking and Learning, 13*(1-2), 132-151. doi: 10.1080/10986065.2011.538294

Dubon, L. P., & Shafer, K. G. (2010). Storyboards for meaningful patterns. *Teaching Children Mathematics*, *16*(6), 325-329. Retrieved from http://www.jstor.org/stable/41199478

Dubreil-Frémont, V., Chevallier-Gaté, C., & Zendrera, N. (2014). Students' conceptions of average and standard deviation. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Edwards, T. G., Özgün-Koca, A., & Barr, J. (2017). Interpretations of box plots: Helping middle school students to think outside the box. *Journal of Statistics Education, 25*(1), 21-28. doi: 10.1080/10691898.2017.1288556

Egan, K. (1986). *Teaching as story telling: An alternative approach to teaching and curriculum in the elementary school*. London, ON: The Althouse Press.

Elia, I., van den Heuvel-Panhuizen, M., & Georgiou, A. (2010). The role of pictures in picture books on children's cognitive engagement with mathematics. *European Early Childhood Education Research Journal*, *18*(3), 275-297. doi: 10.1080/1350293X.2010.500054

Everson, M., Mocko, M., Carver, R., Gabrosek, J., Horton, N., Lock, R., … Wood, B. (2016). *Guidelines for Assessment and Instruction in Statistics Education College Report*. Alexandria, VA: American Statistical Association. Retrieved from http://www.amstat.org/asa/files/pdfs/GAISE/GaiseCollege_Full.pdf

Fiddler, F. (2006). Should psychology abandon *p* values and teach CIs instead? Evidence-based reforms in statistics education. In A. Rossman & B. Chance (Eds.), *Proceedings of the seventh International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/

Flyvbjerg, B. (2006). Five misunderstandings about case-study research. *Qualitative Inquiry, 12*(2), 219-245. doi: 10.1177/1077800405284363

Forster, M., & MacGillivray, H. (2010). Student discovery projects in data analysis. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Friesen, S., & Jardine, D. W. (2009). On field(ing) knowledge. In B. Sriraman & S. Goodchild (Eds.), *Relatively and philosophically earnest: Festschrift in honor of Paul Ernest's 65th birthday* (pp. 147-172). Charlotte, NC: Information Age Publishing.

Gal, I. (2002). Adults' statistical literacy: Meaning, components, responsibilities. *International Statistical Review, 70*, 1-25.

Gal, I., & Garfield, J. (1997). Curricular goals and assessment challenges in statistics education. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 1-13). Burke, VA: IOS Press.

Gal, I., Ginsburg, L, & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37-51). Burke, VA: IOS Press.

Garfield, J. (1995). How students learn statistics. *International Statistical Review, 63*(1), 25-34.

Garfield, J., Aliaga, M., Cobb, G. W., Cuff, C., Gould, R., Lock, R., ... Velleman, P. (2005). *Guidelines for assessment and instruction in statistics education (GAISE): College report*. Alexandria, VA: American Statistical Association. Retrieved from http://www. amstat.org/education/gaise

Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review, 75*(3), 372-396. doi: 10.1111/j.1751-5823.2007.00029.

Garfield, J., & Chance, B. (2000). Assessment in statistics education: Issues and challenges. *Mathematical Thinking and Learning, 2* (1-2), 99-125. doi: 10.1207/S15327833MTL0202_5

Garfield, J., delMas, R., & Zieffler, A. (2012). Developing statistical modelers and thinkers in an introductory tertiary-level statistics course. *ZDM Mathematics Education, 44*, 883-898. doi: 10.1007/s11858-012-0447-5

Gerofsky, S. (2004). *A man left Albuquerque heading east: Word problems as genre in mathematics education*. New York, NY: Peter Lang.

Giannakos, M., Chorianopoulos, K., & Jaccheri, L. (2012). Math is not only for science geeks: Design and assessment of a storytelling serious game. *12th IEEE International Conference on Advanced Learning Technologies* (pp. 418-9). doi: 10.1109/ICAL.T.2012.16

Gil, E., & Ben-Zvi, D. (2011). Explanations and context in the emergence of students' informal inferential reasoning. *Mathematical Thinking and Learning, 13*(1-2), 87-108.

Gordon, S. (2005). Understanding students' experiences of statistics in a service course. *Statistics Education Research Journal, 3*(1), 40-59. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Gougis, R. D., Stomberg, J. F., O'Hare, A. T., O'Reilly, C. M., Bader, N. E., Meixner, T., & Carey, C. C. (2016). Post-secondary science students' explanations of *randomness* and *variation* and implications for science learning. *International Journal of Science and Mathematics Education, 15*(6), 1039-1056. doi: 10.1007/s10763-016-9737-7

Green, T. F. (1971). *The activities of teaching*. New York, NY: McGraw Hill.

Guba, E. G., & Lincoln, Y. S. (1981). *Effective evaluation: Improving the usefulness of evaluation results through responsive and naturalistic approaches*. San Francisco, CA: Jossey-Bass.

Guimarães, G., Gitirana, V., Marques M., & dos Anjos, D. (2010). The concept of mean by primary school students. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Hagen, B., Awosoga, O. A., Kellett, P., & Damgaard, M. (2013). Fear and loathing: Undergraduate nursing students' experiences of a mandatory course in applied statistics. *International Journal of Nursing Education Scholarship, 10*(1), 27-34. doi: 10.1515/ijnes-2012-0044

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: A case of mathematics* (pp. 1-26). Hillsdale, NJ: Lawrence Erlbaum Associates.

Hiedemann, B., & Jones, S. M. (2010). Learning statistics at the farmers market? A comparison of academic service learning and case studies in an introductory statistics course. *Journal of Statistics Education, 18*(3), 1-21. doi: 10.1080/10691898.2010.11889584

Holmes, A., Illowsky, B., & Dean, S. (2016, August 28). *Introductory business statistics*. Retrieved from https://legacy.cnx.org/content/m53808/latest/?collection=col11969/1.5

Hood, M. (2013). Bricks or clicks? Predicting student intentions in a blended learning buffet. *Australasian Journal of Educational Technology, 29*(6), 762-776.

Jacob, B. L., & Doerr, H. M. (2014). Statistical reasoning with the sampling distribution. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The

Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Jones, M. G., & Brader-Araje, L. (2002). The impact of constructivism on education: Language, discourse, and meaning. *American Communication Journal, 5*(3), 1-9.

Ireland, S., & Watson, J. (2009). Building a connection between experimental and theoretical aspects of probability. *International Electronic Journal of Mathematics Education, 4*(3), 339-370. Retrieved from http://www.iejme.com

Kalinowski, P. (2010). Identifying misconceptions about confidence intervals. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Kaplan, J. J., Gabrosek, J. G., Curtiss, P., & Malone, C. (2014). Investigating student understanding of histograms. *Journal of Statistics Educ*ation, 22(2), 1-30.

Kaplan, J. J., Lyford, A., & Jennings, J. K. (2018). Effect of question stem on student descriptions of histograms. *Statistics Education Research Journal, 17*(1), 85-102. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Keller, G. (2018). *Statistics for management and economics* (11th ed.). Boston, MA: Cengage.

Kerby, A. T., & Wroughton, J. R. (2017). When do students' attitudes change? Investigating attitudes at midterm. *Statistics Education Research Journal, 16*(2), 476-486. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Kim, H. J., Alberts, K. S., & Thatcher, S. (2014). Teaching undergraduate students through statistical consulting. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in* *statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Konold, C., & Higgins, T. L. (2003). Reasoning about data. In J. Kilpatrick, W. G. Martin, & D. Schifter (Eds.), *A research companion to Principles and Standards for School Mathematics* (pp. 193–215). Reston, VA: NCTM.

Kreetong, K. (2014). Exploring informal inferential reasoning through data games. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Krishnan, S., & Idris, N. (2014). Students' misconceptions about hypothesis test. *REDIMAT, 3*(3), 276-293. doi: 10.4471/redimat.2014.54

Langrall, C. W. (2010). Does context expertise make a difference when dealing with data? In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Lavy, I., & Mashiach-Eizenberg, M. (2009). The interplay between spoken language and informal definitions of statistical concepts. *Journal of Statistics Education, 17*(1). Retrieved from http://tandfonline.com/toc/ujse20

Lem, S., Onghena, P., Verschaffel, L., & Van Dooren, W. (2013) On the misinterpretation of histograms and box plots. *Educational Psychology, 33*(2), 155-174.

Lerman, S. (1996). Intersubjectivity in mathematics learning: A challenge to the radical constructivist paradigm. *Journal for Research in Mathematics Education, 27*(2), 133-150.

Lock, K., & Meng, X. L. (2010). Real-life module statistics: A happy Harvard experiment. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Makar, K., Bakker, A., & Ben-Zvi, D. (2011). The reasoning behind informal statistical inference. *Mathematical Thinking and Learning, 13*(1-2), 152-173. doi:10.1080/10986065.2011.538301

Makar, K., & Rubin, A. (2009). A framework for thinking about informal statistical inference. *Statistics Education Research Journal, 8*(1), 82-105. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Mathews, D. & Clark, J. M. (2007). *Successful students' conceptions of mean, standard deviation, and the central limit theorem*. Retrieved from http://www1.hollins.edu/faculty/clarkjm/stats1.pdf

Merriam, S. B. (2001). *Qualitative research and case study applications in education* (2nd ed.). San Francisco, CA: Jossey-Bass.

Merriam, S. B. (2008). Case study research. In L. M. English (Ed.), *International encyclopedia of adult education* (pp. 93-96). New York: Palgrave McMillan.

Merriam, S. B. (2009). *Qualitative research: A guide to design and implementation*. San Francisco, CA: Jossey-Bass.

Merriam, S. B, Caffarella, R. S., & Baumgartner, L. M. (2007). *Learning in adulthood: A comprehensive guide* (3rd ed.). San Francisco, CA: Jossey-Bass.

Moore, D. S. (1997). New pedagogy and new content: The case of statistics. *International Statistical Review, 65*(2), 123-137.

Moore, D. S. (1998). Statistics among the Liberal Arts. *Journal of the American Statistical Association, 93*(444), 1253-1259. doi: 10.1080/01621459.1998.10473786

Moore, D. S., & Cobb, G. W. (2000). Statistics and mathematics: Tension and cooperation. *The American Mathematical Monthly, 107*(7), 615-630. doi: www.jstor.org/stable/2589117

Murtonen, M., Olkinuora, E., Tynjala, P., & Lehtinen, E. (2008). "Do I need research skills in my working life?": University students' motivation and difficulties in quantitative methods courses. *Higher Education, 56*, 599-612. doi: 10.1007/s10734-008-9113-9

Nespor, J. (1987). The role of beliefs in the practice of teaching. *Journal of Curriculum Studies, 19*(4), 317-328. doi: 10.1080/0022027870190403

Neumann, D. L., & Hood, M. (2009). The effects of using a wiki on student engagement and learning of report writing skills in a university statistics course. *Australasian Journal of Educational Technology, 25*(3), 382-398. doi: 10.14742/ajet.1141

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

Newbold, P., Carlson, W.L., & Thorne, B. (2010). *Statistics for business and economics* (7th ed.). Toronto, ON: Prentice Hall.

Nicol, C., & Crespo, S. (2005). Exploring mathematics in imaginative places: Rethinking what counts as meaningful contexts for learning mathematics. *School Science and Mathematics, 105*(5), 240-251. doi: 10.1111/j.1949-8594.2005.tb18164.x

Nolan, D., & Temple Lang, D. (2015). Explorations in statistics research: An approach to expose undergraduates to authentic data analysis. *The American Statistician, 69*(4), 292-299.

Nowacki, A. S. (2011). Using the 4MAT framework to design a problem-based learning biostatistics course. *Journal of Statistics Education, 19*(3), 1-24.

Onwuegbuzie, A. J. (1997). Writing a research proposal: The role of library anxiety, statistics anxiety, and composition anxiety. *Library and Information Science Research, 19*(1), 5-33.

Orta, A., & Sanchez, E. (2011). Influential aspects in middle school students' understanding of statistics variation. In M. Pytlak, T. Rowland, & E. Swoboda (Eds.), *Proceedings of the Seventh Congress of the European Society for Research in Mathematics Education*. Rzeszow, Poland: CERME.

O'Shea, J., & Leavy, A. M. (2013). Teaching mathematical problem-solving from an emergent constructivist perspective: The experiences of Irish primary teachers. *Journal of Mathematics Teacher Education, 16*, 293-318. doi: 10.1007/s10857-013-9235-6

Pajares, M. F. (1992). Teachers' beliefs and educational research: Cleaning up a messy construct. *Review of Educational Research, 62*(3), 307-332.

Pariseau, S. E., & Kezim, B. (2007). The effect of using case studies in business statistics. *Journal of Education for Business, 83*(1), 27-31. doi: 10.3200/JOEB.83.1.27-31

Petocz, P., Reid, A., & Gal, I. (2018). Statistics education research. In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 71-99). Cham, Switzerland: Springer.

Pfannkuch, M. (2011). The role of context in developing informal statistical inferential reasoning: A classroom study. *Mathematical Thinking and Learning, 13*(1-2), 27-46.

Pfannkuch, M., Arnold, P., & Wild, C. (2015). What I see is not quite the way it really is: Students' emergent reasoning about sampling variability. *Educational Studies in Mathematics, 88*, 343-360. doi: 10.1007/s10649-014-9539-1

Pfannkuch, M., & Budgett, S. (2014). Constructing inferential concepts through bootstrap and randomization-test simulations: A case study. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Pfannkuch, M., Regan, M. Wild, C., & Horton, N. J. (2010). Telling data stories: Essential dialogues for comparative reasoning. *Journal of Statistics Education, 18*(1), 1-38. Retrieved from http://tandfonline.com/toc/ujse20

Phelps, A. L., & Dostilio, L. (2008). Studying student benefits of assigning a service-learning project compared to a traditional final project in a business statistics class. *Journal of Statistics Education, 16*(3), 1-11. Retrieved from http://tandfonline.com/loi/ujse20

Philipp, R. A. (2007). Mathematics teachers' beliefs and affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning* (Vol. 1, pp. 257-315). Reston, VA: National Council of Teachers of Mathematics.

Philips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher, 24*(7), 5-12.

Pierce, R., & Chick, H. (2013). Workplace statistical literacy for teachers: Interpreting box plots. *Mathematics Education Research Journal, 25*, 189-205. doi: 10.1007/s13394-012-0046-3

Polkinghorne, D. E. (1988). *Narrative knowing and the human sciences*. Albany, NY: State University of New York Press.

Pratt, D., Ainley, J., Kent, P., Levinson, R., Yogui, C., & Kapadia, R. (2011). Role of context in risk-based reasoning. *Mathematical Thinking and Learning, 13*(4), 322-345.

Rafael, J. S. (1997). *Calculus reform from a constructivist perspective* (Master's thesis, University of Calgary). Available from ProQuest Dissertations & Theses Global database. (Order No. 0-612-24694-9).

Ramirez, C., Schau, C., & Emmioglu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal, 11*(2), 57-71. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Ranatunga, J., Strodl, U., & Sorin, R. (2014). The Dreamtime and imagination: Using Aboriginal stories and the arts in teaching and learning. *Proceedings of the 9th International Conference on Imagination and Education*. IERG: Vancouver, BC. Retrieved from http://ierg.ca/how-we-can-help/conferences/conference-proceedings/

Reaburn, R. (2010). Students' understanding of confidence intervals. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/icots/8/cd/home.html

Richards, H. M., & Schwartz, L. J. (2002). Ethics of qualitative research: Are there special issues for health services research? *Family Practice*, *19*, 135-139. doi: 10.1093/fampra/19.2.135

Roberts, N., & Stylianides, A. J. (2013). Telling and illustrating stories of parity: A classroom-based design experiment on young children's use of narrative in mathematics. *ZDM, 45*(3), 453-467. doi: 10.1007/s11858-012-0474-2

Rossiter, M. (2008). Narrative. In L. M. English (Ed.), *International encyclopedia of adult education* (pp. 418-422). New York: Palgrave McMillan.

Rossiter, M., & Clark., M. C. (2007). *Narrative and the practice of adult education*. Malabar, FL: Krieger Publishing.

Saldanha, L., & McAllister, M. (2014). Using re-sampling and sampling variability in an applied context as a bias for making statistical inferences with confidence. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching.* Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Schau, C., & Emmioglu, E. (2012). Do introductory statistics courses in the United States improve students' attitudes. *Statistics Education Research Journal, 11*(2), 86-94. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Sherwood, C. W. (2018). Storytelling and teaching statistics. In M. A. Sorta, A. White, & L. Guyot (Eds.), *Looking back, looking forward. Proceedings of the Tenth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://iase-web.org/Conference_Proceedings.php?p=ICOTS_10_2018

Sierpinska, A. (1994). *Understanding in mathematics*. Bristol, PA: The Falmer Press.

Simon, M. A. (1995). Reconstructing mathematics from a constructivist pedagogy. *Journal for Research in Mathematics Education, 26*(2), 114-145.

Skemp, R. R. (1978). Relational understanding and instrumental understanding. *Arithmetic Teacher, 26*(3), 9-15. (Reprinted from *Mathematics Teaching, 77*, 20-26, 1976)

Smith, D. O. (2014). *Eliciting elementary students' informal inferential reasoning through* *storytelling* (Doctoral dissertation, Western Michigan University). Retrieved from http://scholarworks.wmich.edu/dissertations/315/

Snee, R. D. (1999). Discussion: Development and use of statistical thinking: A new era. *International Statistical Review, 67*(3), 255-258.

Songsore, E., & White, B. J. G. (2018). Students' perception of the future relevance of statistics after completing an online introductory statistics course. *Statistics Education Research Journal, 17*(2), 120-140. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2007). Students' misconceptions of statistical inference: A review of the empirical evidence from research on statistics education. *Educational Research Review, 2*(2), 98-113.

Sotos, A. E. C., Vanhoof, S., Van den Noortgate, W., & Onghena, P. (2009). How confident are students in their misconceptions about hypothesis tests? *Journal of Statistics Education, 17*(2). Retrieved from http://tandfonline.com/loi/ujse20

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks, CA: Sage Publications.

Stake, R. E. (2005). Qualitative case studies. In N. K. Denzin & Y. S. Lincoln (Eds.), *The Sage handbook of qualitative research* (3rd ed., pp. 443-466). Thousand Oaks, CA: Sage Publications.

Steen, L. A. (1988). Calculus today. In L. A. Steen (Ed.), *Calculus for a new century: A pump, not a filter* (pp. 10-13). Washington, D.C.: Mathematics Association of America.

Stevens, S. P., & Palocsay, S. W. (2012). Identifying addressable impediments to student learning in an introductory statistics course. *INFORMS Transactions on Education, 12*(3), 124-139. doi: 10.1287/ited.1120.0085

Swanson, T., VanderStoep, J., & Tintle, N. (2014). Student attitudes towards statistics from a randomization-based curriculum. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Tintle, N., Chance, B., Cobb, G., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2015). *Introduction to statistical investigations* (Preliminary ed.). Hoboken, NY: John Riley.

Tintle, N. L., Rogers, A., Chance, B., Cobb, G. W., Rossman, A., Roy, S., Swanson, T., & VanderStoep, J. (2014). Quantitative evidence for the use of simulation and randomization in the introductory statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Tintle, N., Topliff, K., VanderStoep J., Holmes V. L., & Swanson T. (2012). Retention of statistical concepts in a preliminary randomization-based introductory statistics curriculum. *Statistics Education Research Journal, 11*(1), 21-40. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Tishkovskaya, S., & Lancaster, G. A. (2012). Statistical education in the 21st century: A review of challenges, teaching innovations, and strategies for reform. *Journal of Statistics Education*, *20*(2), 1-56. Retrieved from http://tandfonline.com/loi/ujse20

Thompson, A. G. (1992). Teachers' beliefs and conceptions: A synthesis of the research. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 127-146). New York, NY: MacMillan.

Toor, A., & Mgombelo, J. (2015). Teaching mathematics through storytelling: Engaging the 'being' of a student in mathematics. In K. Krainer & N. Vondrova (Eds.), *Proceedings of the Ninth Congress of the European Society for Research in Mathematics Education* (pp. 3276-3282). Retrieved from hal.archives-ouvertes.fr/hal-01289881

Triola, M. F. (1998). *Elementary statistics* (7th ed.). Reading, MA: Addison-Wesley.

Tsao, Y. L. (2006). Teaching statistics with constructivist-based learning method to describe student attitude towards statistics. *Journal of College Teaching & Learning, 3*(4), 59-64.

van den Heuvel-Panhuizen, M., van den Boogaard, S., & Doig, B. (2009). Picture books stimulate the learning of mathematics. *Australasian Journal of Early Childhood, 34*(3), 30-39.

von Glaserfeld, E. (1982). An interpretation of Piaget's constructivism. *Revue Internationale de Philosophie, 36*(4), 612-635.

von Glaserfeld, E. (1990). An exposition on constructivism: Why some like it radical. In R. B. Davis, C. A. Maher, & N. Noddings (Eds.), *Constructivist views of the teaching and learning of mathematics, Journal for Research in Mathematics Education Monograph, 4* (pp. 19-29). Reston, Virginia: NCTM.

Watkins, A. E., Bargagliotti, A., & Franklin, C. (2014). Simulation of the sampling distribution of the mean can mislead. *Journal of Statistics Education, 22*(3), 1-21. Retrieved from http://tandfonline.com/loi/ujse20

Watson, J. M. (2008). Exploring beginning inference with novice grade 7 students. *Statistics Education Research Journal, 7*(2), 59-82.

Weinberg, A. (2014). Characteristics of students' probabilistic reasoning in a simulation-based statistics course. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics* *education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Wild, C. J., & Pfannkuch, M. (1999). Statistical thinking in empirical enquiry. *International Statistical Review, 67*(3), 223-265.

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What is statistics? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 5-36). Cham, Switzerland: Springer.

Williams, A. S. (2015). Statistics anxiety and worry: The roles of worry beliefs, negative problem orientation, and cognitive avoidance. *Statistics Education Research Journal*, *14*(2), 53-75. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Wroughton, J., McGowan, H., Weiss, L., & Cope, T. (2013). Exploring the role of context in students' understanding of sampling. *Statistics Education Research Journal*, 12(2), 32-58. Retrieved from http://iase-web.org/Publications.php?p=SERJ

Yin, R. K. (2014). *Case study research: Design and methods* (5th ed.). Thousand Oaks, CA: Sage Publications.

Zaidan, A., Ismail, Z., Yusof, Y. M., & Kashefi, H. (2012). Misconceptions in descriptive statistics among postgraduates in social sciences. *Procedia - Social and Behavioral Sciences, 46*, 3535 – 3540. doi: 10.1016/j.sbspro.2012.06.100

Zazkis, R., & Liljedahl, P. (2009). *Teaching mathematics as storytelling*. Rotterdam, The Netherlands: Sense Publishers.

Zieffler, A., delMas, R., Garfield, J., & Brown, E. (2014). The symbiotic, mutualistic relationship between modelling and simulation in developing students' statistical reasoning about inference and uncertainty. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics*. Voorburg, The Netherlands: International Statistical Institute. Retrieved from http://icots.info/9/proceedings/home.html

Zieffler, A., Garfield, J., delMas, R., & Reading, C. (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, *7*(2), 40-58.

Zieffler, A., Garfield, J., & Fry, E. (2018). What is statistics education? In D. Ben-Zvi, K. Makar, & J. Garfield (Eds.), *International handbook of research in statistics education* (pp. 37-70). Cham, Switzerland: Springer.

Zimmerman, W. A. & Austin, S. R. (2018). Using attitudes and anxieties to predict end-of course outcomes in online and face-to-face introductory statistics courses. *Statistics Education Research Journal*, *17*(2), 68-81.

Appendix A – Story-based Tasks and Supplemental Tasks

This appendix provides an example of one of the four story-based tasks used in the intervention, *The Dragon Lady,* which covered the topic sampling distribution of sample means. In the stories, the portion in orange is where the students wrote dialogue. Also included are the reflection and follow-up task for this story. This story-based task was written in conjunction with the instructor for the course in which the study was conducted. As co-author of the story-based task, the instructor has given written permission to publish this story here (see permissions letter in Appendix C).

43. Story-based Task #3: *The Dragon Lady*

Chapter One

Jed was ecstatic. He'd just returned from Dragon's Den with a massive investment in his new venture. Jed and his business partner Sheldon had created a solar powered electric scooter that promised to make hover boards yesterday's news. They had a patent, they had a rich partner from the Den, and now they had a cheque for 100 grand.

Jed had approached the Den because, though his fledgling business was a growing success (he already had one contract with a sporting goods chain and one more pending with a national bicycle retailer), his costs were too high and he needed some help overcoming lingering skepticism in the market about the reliability of solar powered mobility products. In exchange for the cash injection, Jed had sold 20% of his business to Michele Romanow, a tech millionaire from Calgary, and had agreed to work on getting their manufacturing costs down by at least 10%. His Dragon (and now business partner), had also sent an expert consultant named Reema Chaitani to help him with his costs.

As soon as she arrived, Reema got straight to work getting to know the operations, interviewing the assembly line operators and talking to Jed's partner, Sheldon, the engineer who actually invented the solar powered deck and drivetrain. While Jed is the voice and face of the organization, he knows little about the technical side. And while Sheldon knows everything there is to know about the technical side, he has a phobia of business operations—hence the need to bring in Reema, a consultant often called in by Romanow when small start-ups need to untangle the mysteries of building a business and cutting costs.

Jed had been dreading his first meeting with Reema, convinced she was going to start mucking with their invention. He sat up straighter as she entered the room. Impeccably dressed and looking confident as hell, she cut an intimidating figure.

"Sheldon has filled me in on the technical side of the production process," she began, "but can you tell me a bit about the scooter itself?"

Jed brightened. He loved talking about the scooter. He loved talking. Period.

"The **Insert name of scooter here** is game changing!" he began. "The entire deck of the board is a solar panel, with another slimmer panel running down from the handle bars to the front of the board. I don't know if you've seen some of the other scooters (see: http://solarelectricscootersinc.com/video-2/). They're slow and heavy. Plus they only go about 30 kms per charge and max out at about 25 kilometers per hour. Not this one. It's sleek and fast--like a Tesla. It can run for hours in sunny weather without needing to be plugged in. But if it's cloudy, no problem--it takes just two hours to fully charge in a normal electrical socket."

"So what's your secret?" Reema asked.

"Oh, didn't Sheldon go over that? It's the huge power output we get from our revolutionary deck panel, combined with the other panel on the handlebar stem. There's no other scooter that comes even close to that kind of power. This thing will go 50 kmph under ideal conditions-- as long as we don't start messing with the design to cut costs…"

"Slow down, Jed." Reema interrupted. "I don't want to dampen your enthusiasm, so let me clarify why I'm here. And trust me, it's not to mess with your design – I'm strictly here to help you cut costs. I've been pouring over your books for two days straight and found your largest controllable cost driver. Your design looks pretty flawless, but Sheldon's quality control process is flagging a massive number of defects."

"I don't follow," Jed frowned. "Sheldon has never said anything about defects."

"I didn't say you have a massive number of defects, Jed. I said your quality control process seems to think you do. And because of that, Sheldon and his team has been spending way too much time recalibrating the scooters that fail inspection…."

"Recalibrating?" Jed interrupted.

"Yeah. He has to manually tweak the electronics a bit to get each scooter's performance standards in line with its performance requirements. Unfortunately, it's time consuming and expensive. And unnecessary, I might add."

Jed grimaced. "Well, shit."

"Don't panic, Jed." assured Reema. "I think Sheldon has been way too paranoid about this. You guys have been successful because you rigorously monitor the quality of your product. You're right to want to avoid the exploding hoverboard fiasco, and your existing and future clients demand a safe and perfectly functioning product. It's your invention—well Sheldon's invention, really--that will get those kids to school with all their toes still intact, right?"

“Exactly.” Listening not being high on his skill set, Jed was all ears now, having a given one of the boards to his own 14-year-old. “This company is going to change the world, which is precisely why we can’t cut corners on monitoring quality.”

“I agree. But I’m suggesting there might be a better way to approach quality control—one that is just as precise but less likely to result in having to recalibrate so many scooters.”

“So what do you have in mind?”

“I’m still working on it. But let me take you through my investigations so far. Here, look at this spreadsheet I put together this morning.” Reema slid her laptop over to Jed. **(See *Maximum KMPH Scores* on the Excel file.)**

“What am I looking at?” Jed scratched his head.

“Let me explain. The heart of the board is the two solar panels, right?”

“Strong enough to withstand the power of the sun *and* a 250-pound man,” Jed added.

Reema smiled. “Right. But Sheldon tests a random scooter coming off the assembly line about every 15 minutes throughout the day, every day. That’s a lot of testing. Plus, whenever he finds one that doesn’t fall within his rigorous parameters, he grinds the whole production process to a halt to recalibrate the scooter’s electronics.”

“I had no idea.” Jed sighed.

“That spreadsheet you’re looking at has the results of 320 randomly sampled scooter taken off the production line at 15-minute intervals. Each time your operators take a scooter off the line for testing, it’s checked for its maximum attainable speed. As you said, under ideal conditions the solar panels should produce enough power to accelerate a scooter to a maximum of 50 kilometers per hour. Not sure how you guys pulled that off, by the way—or it that’s even safe!--that's a lot of power," she sidetracked. "Anyway, obviously there is some variation in this, but there shouldn't be much." Reema paused. "Am I getting too technical for you?"

"No, I think I'm still with you." Jed scratched his head again, already resenting her vastly superior technical knowledge after only a few days.

"Ok," Reema continued. "You guys have a contractual obligation with your clients that the peak speed for each batch of scooters you ship will average 50 kmph…"

"Or your money back," Jed interrupted, something he did often.

"But here's the problem, Jed. Sheldon has been interpreting the contract to mean that *every* scooter must be capable of 50 kmph. But, that's not what the contract says. The contract focuses on the average, not on every scooter."

"What's the difference?"

"A big one, Jed. One that could bankrupt your company."

"Well that's a bit dramatic isn't it?"

"Not really, Jed. Not all your scooters are going to hit that 50 kmph score exactly. That's demanding a pretty unrealistic level of precision, especially for a new product in a fledgling industry. What your contract actually demands is that the overall *average* peak speed of each batch is 50 kmph and that this *average* falls within a pre-determined tolerance interval. What's got Sheldon pulling his hair out is that a lot of your individual scooters don't fall within your tolerance interval."

"What do you mean by batch?"

"Really, Jed? You don't know how many scooters you cram into a shipping crate?"

"Hey, I'm the marketing guy. I try to keep my hands clean and my Armani wrinkle free." Jed winked.

“Fine. Every batch, by which I mean shipping crate, holds 30 scooters.”

“Okay, got it. And what’s a tolerance interval?”

“It’s the allowable deviation from that average of 50. The fine print actually allows for a tolerance interval between 48 and 52 kmph. Make sense?”

Jed frowned, not wanted to admit his confusion. “Uhm, where did those numbers come from?”

“That was probably negotiated at some point by your contract lawyer, with Sheldon’s input of course. But it doesn’t really matter. You’re stuck with them for now. The fact is your tolerance interval is pretty tight: it doesn’t give you much wiggle room. So, for any given batch, the average maximum speed should fluctuate by no more than 2 kmph on either side of the mean—the mean being 50. So, as long each batch has a *mean* maximum speed that is somewhere between 48 and 52 kmph, it will be within the tolerance interval. Make sense?”

“Ok, I think I get it. But why is it a problem if the maximum speed is too high. That’s good isn’t it?

“Yes and no. You don’t want them getting too high. If they consistently reach speeds over 50 kmph, we risk legislators demanding motorcycle licenses. Plus too much power causes other problems. Remember the exploding hoverboards and the flaming Samsung phones?”

Jed chuckled. Then got worried. “So how do we fix this?”

“As I said earlier, you guys are spending a fortune on quality control. I’m saying we need to cut back on the testing and stop recalibrating so many scooters. That gets expensive, remember?”

“So you keep reminding me.” Jed frowned.

"Right. But now let me show you this." Reema clicked a few more times on the laptop then slid the laptop over to Jed.

1. Use MegaStat to produce the descriptive statistics for the speeds of the individual scooters. Insert the results here. Label them: "Table 1: Descriptive Statistics of KMPH Scores (320 Units)" and "Figure 1: Histogram of KMPH Scores (320 Units)".

"Uhm, what am I looking at?" he asked.

"It's the descriptive statistics and a histogram for all 320 scooters."

"Ah. Ok." Jed was scratching his head furiously.

2. Have Reema and Jed discuss what the descriptive statistics tell us about the speed of an individual scooter. Focus on specific, relevant measures.

"Alright that makes more sense. But what does that have to do with the contract?"

Reema sighed. "Remember the tolerance interval we talked about earlier? Not only is it a contractual obligation; it's also important for quality control. When ensuring quality control, rather than focusing on individual scooters, you should be focusing on batches of scooters, just like your contract does. In other words, your contract recognizes that individual items will always deviate somewhat from the norm, but Sheldon has been forgetting this and doing a lot of expensive and unnecessary recalibrating, thinking his production process was out of control."

"I don't know. Sheldon is pretty smart…"

"I agree, but his statistical reasoning and his reading of contracts needs a bit of polish."

"I suppose," Jed mumbled.

"So based on what we're seeing, your process isn't quite where it should be, but it's not that far off either, certainly not as far off as Sheldon seems to think it is. Unfortunately, the probability that any given scooter coming off the line would need recalibration is really high."

“Probability?”

“Yes. I’ll show you, theoretically, just how large the probability is that Sheldon will recalibrate a randomly chosen scooter under his system of testing. We know it’s over 50% based on the quartiles, but let’s look a bit more closely. To do this, let’s assume the scooters’ maximum speeds follow a normal distribution—I realize that’s not quite accurate, but it’s close enough for what I’d like to show you. Let’s produce the normal distribution on Excel using the same parameters Sheldon is using. Looking at the descriptive statistics for the 320 scooters, we have a mean **insert the mean from the descriptive statistics** and a standard deviation of **insert the standard deviation from the descriptive statistics.** You still with me?”

“Uhm, I’m trying.” Jed was looking a bit pale.

“I think Sheldon and I can actually get those numbers down a bit once we tune up his aging machinery and re-train some of his operators, so I’m going to anticipate an eventual mean of 50 and a standard deviation of 2.5. So what I want to show you now is a *theoretical distribution* based on just two numbers—the mean and standard deviation. Keep in mind, because this is a theoretical distribution, it’s not going to give us a perfect impression of what is really happening, but it should be close. Once we have the theoretical distribution, we’ll check to see what the probability is of a scooter failing a quality control test.”

“Ahh…”

“Don’t worry, Jed, I’ll guide you through it.”

3. Use MegaStat to find the probabilities required to answer this question. Use the snipping tool to paste the resulting distribution below as “Figure 2: Probabilities for tolerance interval”. Have Jed or Reema clearly state the probability that a scooter will fail the probability test.

“Ok, so what am I looking at?” Jed asked.

4. Write the dialogue to explain the theoretical probability of any particular scooter falling outside the tolerance interval. Your explanation needs to include why, if Sheldon's interpretation of the contract was correct, this would be really bad. Keep in mind that this probability will be based on the assumption that the population of scores is normally distributed, which we know is not necessarily true.

Jed looked terrified. "Oh no! How are we ever going to meet the contract when we have to recalibrate so many scooters? I've sunk my whole life savings into this project. No wonder Sheldon is going bald."

"Calm down. Remember this is what Sheldon has been looking at, but I'm trying to tell you he's wrong. Deep breath, Jed. Deep breath." Reema took in a deep breath.

Chapter Three

"Ok. Now that you've stopped hyperventilating let me tell you my plan before we go any further, Jed."

"I'm all ears."

"First, we need to focus on the actual contractual observations of the *average* score of those samples, not on the individual scores, as Sheldon has been doing."

Reema paused to make sure that Jed was doing ok. "Let me show you a histogram of the *average* scores. We'll use samples of 30 because that's how many scooters are in each crate, taken from the 'parent sample' of 320."

"Huh?"

"Don't worry, Jed. And stop scratching your head! What we will do is take thousands of samples of 30 over and over again from the parent sample, measure the *average* of each of those samples and then plot those averages on a histogram. This will give us a better tool for checking quality control."

"I thought you said we needed to stop over testing. Now you're talking about thousands…"

"No, no. We won't actually do this in practice. We will do this as a simulation. Once we have that, you should better understand where I'm going with this."

"That sounds like an insane amount of work." Jed sighed.

"Actually, there are applets that can do this in seconds." Reema started clicking again. "Look, here's one I've used in the past. Unlike the theoretical distribution we made earlier, this applet can use the actual data (the 320 scores we looked at earlier) and simulate a distribution based on repeated samples taken from those 320 values."

"And that's better?"

"Not necessarily. But to create a theoretical distribution here we would need the population data, which would be the peak speed for every scooter you've ever created … and will ever create. That's impossible and costly. So instead we are going to use the empirical distribution which is based on the sample data that we have, which we call the 'parent sample'. Since our sample is fairly large and has been collected using a sound sampling technique, we can be confident that the distribution of the sample is pretty close to the distribution of the population. The **state the relevant statistical law here** tells us this."

"Ok. So what is this empirical distribution that you're going to create?" Jed was looking very confused.

"That *we're* going to create Jed." Reema looked reassuringly at Jed. "It's actually pretty simple, Jed. We will start with real data—in this case our sample of 320 observations. We will then take a small random sample from this 'parent sample.' It makes sense to take a sample of 30 because that's the size of each batch of scooters. We will measure the mean of that sample of 30, record it and then replace the 30 observations back into the parent sample. Then we will take another sample of 30 and record its mean. We will do this over and over again until we have thousands of sample means recorded. The resulting distribution is called the 'sampling distribution' because it was created from sampling over and over again. It is *empirical* because we used sample data. If we wanted a *theoretical* sampling distribution we would need to take *all* possible samples of size 30 from the parent sample. There are 1.32 x 10^{42} samples of size 30 from our parent sample. That's a tredecillion samples. You up for it?"

Jed ignored the question. "Does this process have a name?"

"This process of using what we have (the sample data) to try to figure out stuff about something we don't have (the population data) is called 'bootstrapping'. Specifically, we are bootstrapping what the theoretical sampling distribution will look like based on an empirical sampling distribution that was created from sample data. Bootstrapping is simply the process of building an empirical sampling distribution by re-sampling over and over again from a parent sample. Each sample must be random and the same size, of course."

"I think I've got this."

5. Have Jed and Reema discuss what a sampling distribution is. In particular, comment on how the data in the sample and the data on the sampling distribution are different and how the process of collecting a sample is different from the process of creating a sampling distribution.

"Well done, Jed. Meaning that, unlike our previous histogram, which was built from the performance scores of *individual* scooters, our empirical sampling distribution will be built from the *mean* performance scores taken from batches of 30 scooters sampled *ten thousand times*."

"You're shitting me, right?'

"Just watch me, Jed."

6. Go to:

http://www.rossmanchance.com/applets/OneSample.htmlhttp://www.rossmanchance.com/applets/OneSample.html

Use the 320 measurements from the Excel sheet as your parent sample and then take at least 10,000 samples of 30 for this simulation.

a) Click on "Clear" and then "Use Data"

b) Copy the raw data (320 measurements) from Excel. The data must be in one column and must contain a ONE WORD label in the first cell.

c) Paste the raw data from one column, including the label in the first cell, into the window and the click "Use Data" again

d) Click "Show Sampling Options"

e) Type in the desired number of samples (10,000) and the sample size (30).

f) Click on "Draw Samples"

Snip all three histograms from the bottom of applet's window and paste them below. Label it "Figure 3: Bootstrapped Sampling Distribution".

Jed scratched his head. "So now what are we looking at?"

"The first histogram on the left is the same as Figure 1. It's simply a histogram of 320 values. It looks a bit different but that's just because the width of each bar is different. As you can see, it confirms that the output is not normally shaped."

"Ok, I got that part. But what are those blue pimply things?"

"Oh, they simply represent where the 30 items of the last (10,000th) sample came from. It just reinforces the fact that each sample is truly random."

"And what's the second histogram?" Jed asked.

"That's simply a histogram of those 'pimples', as you called them. It represents the *very last* sample of 30 taken from the parent sample. It doesn't really tell us anything, other than to further demonstrate that small samples aren't very good at representing the population. As you can see, it looks nothing like the parent sample."

"And the third?"

"Now that's our secret weapon, Jed. That's our **empirical sampling distribution of sample means**."

"Ok." Jed tried to sound like he understood.

"Bear with me. Because this is a bit technical, I'm going to summarize again." she said reassuringly. "Earlier I showed a histogram (Figure 1) with all 320 measurements (our parent sample). It was **insert shape**."

"Right."

"Now, as I said earlier, what we've done is to take a random sample of 30 values from a 'parent sample' of 320 values. We then measured the mean of those 30 values, recorded that mean and then threw those 30 values back into the parent sample. Then we did this over and over again, 10,000 times. You still with me?"

“I think so,” Jed resisted the urge to scratch his head.

“Once we had these 10,000 means, we plotted them all on a curve. And presto! We have what’s called a sampling distribution of means.” Reema smiled proudly. “Remember, it’s an *empirical* sampling distribution because it is based on sample data rather than population data -- 10,000 sample means taken from your ‘parent sample’ of 320 scooters.

7. Have Reema and Jed discuss why a sample of 10,000 means will result in an empirical sampling distribution that will be a good estimate of the theoretical sampling distribution.

“And, on top of all that, look at those beautiful curves!” She pointed at the sampling distribution on the screen, smiling again.

“You’re weird,” but Jed was smiling too. “I’m still not sure what we’re supposed to do with this, though.”

“Remember when we calculated how many individual scooters would fall outside the tolerance interval?”

“Too well. It still frightens me.”

“But as I said many times, we should not be worrying about individual scooters’ performance scores. We should be worrying about the average score per batch.”

“Right, how could I forget?”

“Well, this sampling distribution of sample means is the tool that will allow us to assess the quality of each batch of scooters and, more importantly, predict the probability that the average score (from a batch of 30) will fall outside the tolerance interval, which is what Sheldon should have been doing all along.”

8. As Reema, discuss with Jed the features of the sampling distribution of means on the right side of Figure 3. Compare and contrast those features (shape, centre and spread) with those of the first histogram on the left side of Figure 3.

Chapter Four

"Nice work, Jed. Now, have a look at the standard deviation on the 'parent sample', continued Reema.

"Yeah, it's about **insert standard deviation**."

"Right. But now look at the standard deviation on the sampling distribution. Notice anything?"

"Yeah. It's **Pick either "larger", "smaller" or "equal to"**. That's weird. I thought we used the same data to produce both histograms."

"The same data, yes, but not the same measurement, Jed. The sampling distribution of sample means is a histogram of *mean* scores (each one from a sample of 30, remember?), whereas the histogram of the 'parent sample' (Figure 1) is a histogram of 320 *individual* scores. Plus, the sampling distribution contains data that has come from bootstrapping (re-sampling over and over again). There are 10,000 means on that distribution, remember?"

"Sorry, you'll need to explain the different standard deviations more thoroughly. My head is spinning again." Jed slouched.

"Nope."

"What?"

Reema winked. "I'd rather you try. I want to see if you've learned anything."

9. Have Jed and Reema discuss why the standard deviations are so different.

"So now that you understand how an empirical sampling distribution is created and what it measures, we can start using it as a tool to check if we are meeting our contractual obligations."

"Once again, I'm all ears," Jed offered rare smile, thankful that they appeared to be nearing some sort of conclusion to this statistics lesson.

Reema continued. "So let's say, for example, we randomly selected 30 units for inspection, what are the chances we'll fail our contractual obligations to our clients based on the criteria for the mean kmph score?"

"Sorry, what?" asked Jed.

"What are the chances that a random sample of 30 will have a *mean* score that falls outside of your tolerance interval?"

"It's not 50% is it?" Jed offered.

"Not even close. We need to examine these probabilities—just as we did earlier—but with a focus on the likelihood that any particular *mean* would fall outside the interval.

10. Go back to the applet and, using the "Count Samples ≥" and "Count Samples ≤" windows, calculate the probabilities that the sample mean would fall outside the tolerance interval. Paste your results below. Have Jed and Reema discuss these probabilities. (Keep in mind that you are focusing the probability that the mean score will fall above or below the tolerance interval identified earlier.)

    Note: Everyone's answers will be slightly different because we are dealing with an empirical sampling distribution. But your answer should be very small.

"Ok, wait, Reema. If I'm understanding you correctly, Sheldon has been recalibrating over 50% of the scooters unnecessarily?"

"You got it, Jed."

Chapter Five

After recovering from the shock that Sheldon was doing something so wrong, Jed sat quietly for a moment digesting all that he'd learned. "I hate to ask, but something has been bugging me. What if the crates had only 20 scooters in them? Or 50? Would the same thing happen? What if the parent sample was all bunched on one side with a few values on the other?"

Reema looked positively gleeful. "I'm so glad you asked. There's a statistical theorem called 'The Central Limit Theorem'. It's almost magical…"

Shoulders slumped, Jed immediately regretted opening his mouth.

11. After going over the activity with your instructor, have Jed and Reema discuss the properties of a sampling distribution of the means. In particular, explain what happens to the mean, the standard deviation and the shape of the sampling distribution for different sample sizes. This question is the crux of this assignment. Spend most of your time answering this question. If you really get this, you've really got this concept.

"Nice work Jed. Now let me show you how this will help your quality control *and* cut costs."

"Yes please. But first, it's time for a coffee."

Chapter Six

"You've got to be freaking kidding me!" Sheldon yelled.

Alarmed and a little frightened, Jed pushed his chair back further. Sheldon *never* yelled. "Sorry Sheldon, but Reema went over everything with me forwards and backwards…"

Sheldon threw himself into the leather recliner in Jed's office, nearly shaking with rage. "So I've been wasting hours recalibrating every second randomly tested scooter when I didn't have to?" He growled through clenched teeth.

"That the short of it, yes. I wish Reema were here to explain; she's so much better at it, but I'll have to do…"

12. As Jed, explain to Sheldon how he has been misreading the contract. Then explain how he should use a sampling distribution to properly do quality control. Compare the probability of an individual scooter being within a tolerance interval with the probability of the average of 30 being in the tolerance interval to drive your point home. Explain why this revised method will save money.

44. Reflection Task

Solutions to the assignment will be posted after [the story-based task] is due. The goal of this part of the assignment is to reflect on what you've learned from the assignment.

Before completing this component, do the following, but do not submit it. See below for what you submit.

1. Compare your assignment to the solutions. The solutions will include the correct answer and possible areas of misunderstanding. For each question, give yourself a mark out of 3 on each question.

Description	Mark
Your answer is very similar to the answer provided.	3
Your answer is similar to the answer provided, but is either slightly incomplete (i.e., missing one or two detail) or contains one or two errors in reasoning. Note: If the details or errors are substantial, then it is a 1.	2
Your answer contains part of the answer provided, but is either incomplete (i.e., missing many details or one major component) or contains multiple errors in reasoning or one major error.	1
Your answer is nothing like the answer provided or is blank.	0

This will give you a sense of how well you understood the assignment.

2. Evaluate your understanding of the material covered in the assignment:
   a. What did you understand?
   b. What did you struggle with?

c. For the areas that you struggled with, do you understand them better now that you've seen the solutions? If not, what are you going to do to get help to understand the area better?

Submission: (this is what you will actually submit to your instructor)

1. *What did you understand?*
   a. Copy and paste a question on the assignment that you were successful on. **(1)**
   b. Copy and paste your answer to the question provided in a). Do not change your answer from what you submitted in Part 1. **(3)**
2. *What did you struggle with?*

   For this question, you are NOT allowed to use a question where you were only asked to insert input from MegaStat.

   a. Copy and paste a question on the assignment that you were **not** successful on. **(1)**
   b. Copy and paste your **incorrect** answer to the question. Note: Do not change your answer to the question from what you submitted in part 1. **(1)**
   c. Having looked at the key, explain what was incorrect in the answer to your question and why it was incorrect. **(3)**
   d. For the question provided in c), provide your revised answer. Make sure you answer the question completely. Note: Be careful! You cannot just write down the solution provided or the definition from a textbook/website. You need to write it in your own words and demonstrate *your* understanding of the solution. If your answer too closely resembles the solution provided or a textbook/website answer, it may be considered plagiarised (see the MRU calendar for a thorough discussion of academic misconduct and its penalties). **(4)**

3. This story sought to demonstrate how statistics can be used to solve a real-world problem. Provide an example of how you think you can use the statistical concepts from the story either in your everyday life or your future career. **(2)**

45. Follow-up Task #3

RoboTune (RT for short), a manufacturer of electric vehicles, is currently trying to break into some of the BRIC countries: Brazil, Russia, India, and China. Alvin Monk, the CEO of RT has recently returned from a visit to China, where he met with several top-level ministers who have the power to make or break the Monk's endeavor. China is particularly interested in electric vehicles as a means of reducing air pollution.

Among Monk's challenges is to create sufficient demand, especially difficult for a manufacturer of premium priced vehicles trying to enter markets where disposable incomes are well below that of the U.S., where the cars are manufactured. But before he can start focusing on consumer demand, he must convince the Chinese government that his product is worth the infrastructure investment of building thousands of charging stations across the country. On top of this, the Minister of Transport, Li Xiaopeng, is not yet convinced that the vehicles will perform as promised.

During their China visit, Monk's team presented performance figures based on American testing of over 300 vehicles and have argued that their RT Model A will average 365-390 kms per charge (under normal driving conditions) at a cost of only $3.00 US (about 20 Chinese Yuan) per full-charge. Moreover, Monk has guaranteed that at least 95% of the time any random shipment of Model A's will exceed 370 km's per charge, on average.

Monk offered to send two shipments of 30 (60 cars in total) Model A's to China to allow them to conduct their own tests. The minister agreed but insisted that the shipment be randomly selected from vehicles already in use in the US. After the vehicles had been shipped, testing began under the tight control of the Ministry of Transport. After examining the raw data of the results, Mr. Li was not impressed.

Km's Per Charge

342	362	366	373	384	401
345	362	366	374	386	404
345	363	368	374	388	411
347	363	368	375	390	417
348	364	370	375	393	418
354	364	370	376	394	418
356	365	372	377	394	418
358	365	372	379	395	421
358	365	372	379	398	423
361	365	372	383	399	427

1. Mr. Li is not impressed with these results. Specifically, he is concerned with those features of the data that can be seen without doing any calculations. He believes the results indicate that Mr. Monk's performance guarantees are exaggerated and the battery life of these vehicles is wildly out of control.

   Hint: Mr. Li is looking at the raw data from the perspective of someone who knows little about statistics and is basing his initial impression on merely a cursory glance at the raw data above. He has not run any descriptive statistics. Explain how Mr. Li reached his conclusions, based on examining the raw data (Do not run any descriptive statistics at this point.).
2. Mr. Monk, who has spent some time working with statistics, disagrees. Generate all relevant output that might be helpful in supporting Mr. Monk's case. Use the descriptive statistics function on MegaStat. Paste the output below.

3. Looking at only the relevant measures from your descriptive statistics output, explain why the results do actually support Mr. Monk's performance claims. Specifically, examine the measures of centre and the empirical rule in addressing this. NOTE: A process that is 'in control' would be expected to follow a normal distribution.
4. Mr. Monk is going to have to educate Mr. Li on the wonders of sampling distributions. Use the Rossman Chance applet (http://www.rossmanchance.com/applets/OneSample.html) to generate a bootstrapped sampling distribution from the raw data (so we all have similar results, set the *n* as 30 and the number of repetitions as 100,000). Paste your results below.
5. Mr. Li has pointed to the histogram on the left and argues that the large standard deviation is further proof that there is too much variation in battery life. He argues that Monk's "give or take" value should be only around 10 kms above and below the mean. Monk says that Mr. Li is looking at the wrong measure, arguing that the standard error is the correct "give and take" value to use in this context. Explain why Monk is correct—and Mr. Li is incorrect--in using the standard error in assessing Mr. Monk's guarantees about battery performance. Be sure to carefully interpret the standard error in this context. (Use the standard error from your bootstrapped sampling distribution above or calculate the theoretical standard error manually.)
6. Mr. Li is also concerned that Mr. Monk's claim that **at least** 95% of the time any random shipment of Model A's will average of over 370 kms per charge may be exaggerated. Use the applet's "Greater than" function and your bootstrapped sampling distribution to demonstrate why Mr. Li should not be concerned. Paste the applet results below and clearly explain your reasoning.

7. Mr. Li is now more or less comfortable with Mr. Monk's guarantees, but his curiosity has been piqued in further examining the applet's output. Explain for Mr. Li why and how the shape of the histogram on the left differs from that on the right. He is very curious.

Appendix B - Instruments

This section provides the instruments used in the study. In particular, the pre- and post-intervention written response items, and the question guide used in the interview are provided.

46. Pre-intervention Written Response Items

Thank you for agreeing to participate in this study that is seeking to understand the effectiveness of using stories to teach statistics. These first set of questions are about your beliefs about the usefulness of statistics prior to taking the course.

Participation in the research is completely voluntary. Even though you have agreed to participate, you can refuse to answer any questions and stop the collection of data at any time. You may withdraw from the study altogether without consequence or explanation at any time prior to June 15, 2017, by contacting me at [email address]. If you decide to withdraw, please let me know so that I will stop contacting you. If you decide to withdraw from the study, all of the information collected in the study will be destroyed.

1. Please include your name.
2. Will statistics be useful in your life and career? Explain your answer.
3. Please provide an example of how you or someone else has either recently used statistics in everyday life or at work. In your example, please explain how statistics was used.
4. Please provide an example of how you believe statistics will be useful in your future career. Please clearly state what your planned future career is. If you cannot think of a way that it will be useful to you, please provide an example of how you think it will be useful for someone in business.

47. Post-intervention Written Response Items

Thank you for agreeing to participate in this study that is seeking to understand the effectiveness of using stories to teach statistics. These questions are about your beliefs about the usefulness of statistics after taking the course.

Participation in the research is completely voluntary. Even though you have agreed to participate, you can refuse to answer any questions and stop the collection of data at any time. You may withdraw from the study altogether without consequence or explanation at any time prior to June 15, 2017, by contacting me at [email address]. If you decide to withdraw, please let me know so that I will stop contacting you. If you decide to withdraw from the study, all of the information collected in the study will be destroyed.

1. Please include your name.
2. Will statistics be useful in your life and career? Explain your answer.
3. Please provide an example of how you or someone else has either recently used statistics in everyday life or at work. In your example, please explain how statistics was used.
4. Please provide an example of how you believe statistics will be useful in your future career. Please clearly state what your planned future career is. If you cannot think of a way that it will be useful to you, please provide an example of how you think it will be useful for someone in business.

The following questions are about the stories in the class. The stories covered in the class were Bob's Bikes, Can Dolphins Communicate?, The Dragon Lady and *Can They DIG It?*.

5. In what ways do you think the stories impacted your beliefs about the usefulness of statistics?
6. What story stood out the most for you and why?

7. Which story did you learn from the most? What did you learn from that story?
8. Which story did you learn from the least? Why?

48. Semi-structured Interview Question Guide

Learning and stories

- Which story do you remember the most? (If they provide many, ask them to focus on the one that stands out the most for them)
    - Why was it so memorable?
    - How did it help you learn?
    - What did you learn from the story?
    - What did you find engaging in the story?
    - What do you remember about the story?
    - Which story did you relate to the most? Why?
    - What statistical ideas were covered in the story?
- What statistics concepts stand out for you? Why? (If they provide many, ask them to focus on the one that stands out the most for them)
    - Tell me about the concept.
    - How did the story help you learn the concept (if at all)?
- The purpose of written dialogue for the stories was to give you the opportunity to explain your learning. How do you think this process impacted your learning?
    - How did it help?
    - How did it hinder?
- The take home assignments had three parts: the story, reflection and Part 3.
    - What were your experiences with the process?

- While studying for exams or quizzes, how did you use the stories to study?
- One of the reasons for using stories is because they are memorable. Did you ever find when you were thinking about the course content that you would think back to the story to help you remember it? If so, please provide an example. If not, why not.

Beliefs

- Can you give me an example of how what you've learned in this course has been or could be useful in your everyday life?
    - How has your view of statistics changed during the course?
    - The stories were written to provide examples of how statistics is used. How, if at all, did they change your beliefs about the usefulness of statistics?
- How do you think statistics will help you in your future career?
    - How do you think statistics helps businesses?

Appendix C – Permission Letter

Dear [redacted]

As you are the co-author of the story-based task entitled "The Dragon Lady," I would like to include the above-mentioned story in my PhD dissertation. The dissertation will be published in the "University of Calgary Theses Repository – The Vault".

If you agree to provide me with permission, please sign this permission letter and return one copy to me by email (a scanned version is fine).

I appreciate your consideration of my permissions request.

Sincerely,
Collette Lemieux

By signing below, I warrant that I have the right to grant the permission requested in this letter, and that I provide you with that permission.

Signature:
[redacted]
Date:
Jan 9, 2020