
Theresa A. Hickey
Seattle Pacific University

Follow this and additional works at: http://digitalcommons.spu.edu/soe_etd

Part of the Curriculum and Instruction Commons, Educational Assessment, Evaluation, and Research Commons, Educational Methods Commons, Educational Psychology Commons, Higher Education and Teaching Commons, Nursing Commons, Other Education Commons, Other Teacher Education and Professional Development Commons, Scholarship of Teaching and Learning Commons, Science and Mathematics Education Commons, Secondary Education Commons, Secondary Education and Teaching Commons, and the Social and Behavioral Sciences Commons

Recommended Citation
http://digitalcommons.spu.edu/soe_etd/28
Concept Maps and Feedback in Statistics Learning:
Exploring the Effect of Expert Map Feedback and Peer Feedback on Concept Map Structure

Theresa Hickey

Dissertation
Presented to the Faculty of the Graduate School of Education at Seattle Pacific University In Partial Fulfillment of the Requirements for the Doctor of Education Degree

Seattle Pacific University
2017
Concept Maps and Feedback in Statistics Learning:
Exploring the Effect of Expert Map Feedback and Peer Feedback
on Concept Map Structure

By Theresa Hickey

A dissertation submitted in partial fulfillment
Of the requirements for the degree of
Doctor of Education
Seattle Pacific University
2017

Approved by

(Nyaradzo Myududu, Ed.D., Chairperson)

(William Nagy, Ph.D.)

(David Denton, Ed.D.)

Program Authorized to Offer Degree

School of Education

Date

(DECEMBER 2017)

(Rick Eigenbrood, Ph.D., Dean, School of Education)
In presenting this dissertation in partial fulfillment of the requirements for the Doctoral degree at Seattle Pacific University, I agree that the library shall make its copies freely available for inspection. I further agree that extensive copying of this dissertation is allowable only for scholarly purposes, consistent with “fair use” as prescribed in the U.S. Copyright Law. Requests for copying or reproduction of this dissertation may be referred to University Microfilms, 1490 Eisenhower Place, P.O. Box 975, Ann Arbor, MI 48106, to whom the author has granted “the right to reproduce and sell (a) copies of the manuscript in microfilm and/or (b) printed copies of the manuscript from microfilm.”

Signature
Theresa Hickey
Date
12/12/17
Seattle Pacific University

Abstract

Concept Maps and Feedback in Statistics Learning:
Exploring the Effect of Expert Map Feedback and Peer Feedback
on Concept Map Structure

by

Theresa Hickey

Chairperson of Dissertation Committee: Nyaradzo Mvududu
School of Education

This mixed methods exploratory study examined the effects of two types of feedback – Peer and Expert Map – on Concept Maps used as learning tools in a statistics classroom. Of interest were possible effects of feedback on Concept Map structure (determined by structural scores), on structural classification (Discrete, Integrated), and on student choice of starting concept (General, Specific). Student perceptions of feedback and the role of mapping in statistics learning were elicited using a 12-item questionnaire. Two open-ended responses were coded by themes. All data were assessed separately, then merged to enrich findings.

Results from a mixed repeated measures ANOVA of Concept Map scores indicate a significant main effect of Time ($F(1, 26) = 4.92, p = .035, \eta^2_p = .159$). The main effect of Feedback Group was not significant at the $p = .05$ level ($F(1, 26) = 3.50, p = .073, \eta^2_p = .118$). Given the study’s low power, outright dismissal of this effect might result in a Type II error.
Questionnaire responses indicated that Concept Maps and feedback were positively perceived. Disaggregation by structural classification revealed that all students with Discrete maps agreed that feedback sessions aided conceptual understanding (100%), while those with Integrated maps were mixed (42% Agreed, 33% Disagreed). Students with Discrete maps agreed that feedback sessions helped them see new relationships (82%), while students with Integrated maps responded more neutrally (25% Agreed, 58% Neither, 17% Disagreed). Students with Discrete maps planned structure and relationships revisions (85%), while the group most focused on presentation revisions had Integrated maps to start (67%).

Analysis of starting concepts before and after feedback indicated students improved in their starting choices. Results from a 2-sided, exact McNemar test indicate a statistically significant difference in the proportion of students using General starting terms pre- and post-feedback, $p = .016$.

Regarding classifications, results from 2-sided exact McNemar tests indicate the proportion of Discrete and Integrated maps before and after feedback were not significantly different for the Peer group ($p = 1.0$), but might be considered significant for the Expert group ($p = .065$).
Acknowledgements

A heartfelt thank you to my brilliant and kind dissertation committee and to my amazing chair, who freely shared her patience, good humor, and statistics insights.

Very special thanks to my children Hannah, Eddie, and Sam, who were so uplifting and of whom I am very proud, and to my best friend Kent for his encouragement, good humor, and love.

I am so grateful.
# Table of Contents

List of Figures .............................................................................................................. v
List of Tables ................................................................................................................ vi
List of Appendices ........................................................................................................ v

Chapter One: Introduction .......................................................................................... 2
  Background ............................................................................................................... 2
  Focus ....................................................................................................................... 8
  Purpose .................................................................................................................... 8
  Approach .................................................................................................................. 10
  Significance of the Study ......................................................................................... 12
    Learning Theory .................................................................................................... 12
    Statistics Education ............................................................................................. 12
    Concept Mapping ................................................................................................. 13
  Research Questions ................................................................................................. 13
  Structure of the Dissertation .................................................................................... 14

Chapter Two: Review of Literature ............................................................................ 16
  Definitions ............................................................................................................... 16
    Concept Maps ....................................................................................................... 16
      Concepts .............................................................................................................. 17
      Linking Words .................................................................................................... 18
      Propositions ....................................................................................................... 18
      Focus Question .................................................................................................. 19
Feedback .................................................................................................................20
Theoretical Foundations ..........................................................................................23
Social Constructivism ............................................................................................23
Ausubel’s Learning Psychology .............................................................................25
Connections with Concept Maps ..........................................................................28
Overview of the Research Literature on Concept Maps ......................................28
Early Research Directions ....................................................................................29
More Recent Research Directions ......................................................................33
Summary ..................................................................................................................35
Current Considerations and Empirical Studies ..................................................36
Student Created Maps Versus Pre-Determined Maps ...........................................36
How to Measure Learning ......................................................................................40
  Objective Measures .............................................................................................41
  Comparing Maps Across Time ..........................................................................43
Use of Technology ..................................................................................................46
Additional Research Directions ..........................................................................50
  Feedback ..............................................................................................................50
  Changing Map Features ......................................................................................52
  Scoring ................................................................................................................52
Concept Maps and Statistics Learning .................................................................53
Chapter Three: Research Methodology ..............................................................55
  Research Design ..................................................................................................55
  Internal and External Validity .............................................................................56
Participants and Sampling..............................................................59
  Groups..................................................................................60
Instrumentation ..........................................................................61
  Expert Map............................................................................61
  Questionnaire ......................................................................62
Procedures..................................................................................66
Statistical Analyses......................................................................68
  Sources of Data ...................................................................68
  Quantitative Analysis..........................................................69
  Analysis of Question One....................................................70
  Analysis of Question Two .....................................................72
  Analysis of Question Three ..................................................73
Chapter Four: Results ...............................................................75
  Chapter Overview ................................................................75
  Research Question One Results............................................76
    Descriptive Statistics........................................................77
    Inferential Statistics ..........................................................80
    Results...............................................................................81
  Research Question Two Results ............................................83
    Likert Scale Responses ......................................................84
    Open-ended Responses ......................................................85
  Research Question Three Results ...........................................88
    Structural Classification.....................................................88
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Choice of Starting Concepts</td>
<td>92</td>
</tr>
<tr>
<td>Summary of Results</td>
<td>94</td>
</tr>
<tr>
<td>Chapter Five: Discussion of Results and Conclusions</td>
<td>98</td>
</tr>
<tr>
<td>Chapter Overview</td>
<td>98</td>
</tr>
<tr>
<td>Research Question One Discussion</td>
<td>98</td>
</tr>
<tr>
<td>Research Question Two Discussion</td>
<td>104</td>
</tr>
<tr>
<td>Research Question Three Discussion</td>
<td>108</td>
</tr>
<tr>
<td>Limitations</td>
<td>113</td>
</tr>
<tr>
<td>Implications for Theory</td>
<td>117</td>
</tr>
<tr>
<td>Implications for Practice</td>
<td>118</td>
</tr>
<tr>
<td>Conclusions and Further Research</td>
<td>120</td>
</tr>
<tr>
<td>References</td>
<td>121</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: The Elements of a Concept Map .................................................................17
Figure 2: Convergent Mixed Methods - Illustration of Process ...............................56
Figure 3: The Scoring System Created by Novak (1984) ........................................71
Figure 4: Histograms of Concept Map 1 Scores by Feedback Group ....................79
Figure 5: Line Graph - Effects of Time and Feedback Group on CMap Scores ..........82
Figure 6: Change in Concept Map Scores After Feedback Sessions .......................83
Figure 7: Responses to Question 6 by Classification ..............................................90
Figure 8: Responses to Question 7 by Classification ..............................................91
Figure 9: Coded Responses to Question 12 by Classification .................................92
Figure 10: Sample Line Graphs of Re-Scored Effects ............................................101
Figure 11: Merged Data – Question 12 results with Concept Map 1 scores ............108
List of Tables

Table 1: Demographic Information of Sample ..........................................................60
Table 2: Descriptive Statistics for Concept Map 1 and Concept Map 2 ......................77
Table 3: Descriptive Statistics for Concept Maps 1 and 2 by Feedback Group ..........78
Table 4: Tests of Normality by Feedback Group ......................................................81
Table 5: Frequency of Question 11 Responses by Theme and by Feedback Group ....86
Table 6: Frequency of Question 12 Responses by Theme and by Feedback Group ....87
Table 7: Structural Classifications of Concept Map 1 by Feedback Group ...............89
Table 8: Structural Classifications of Concept Map 2 by Feedback Group ...............89
Table 9: Choice of Starting Concepts for Concept Maps 1 and 2 – Overall ............93
Table 10: Choice of Starting Concepts for Concept Maps by Feedback Group .........94
List of Appendices

Appendix A: Student Questionnaire ................................................................. 134
Appendix B: Expert Concept Map ................................................................. 135
Appendix C: Concept Map 1 ........................................................................ 136
Appendix D: Discourse Supports for Group Work ....................................... 137
Appendix E: Concept Map 2 ........................................................................ 138
Chapter One

Introduction

Background

The call to change statistics education is not a new one. For the past 45 years, there has been a consensus among statistics educators that change was needed in the way statistics courses were focused and taught (Cobb, 1992; Garfield, Hogg, Schau, & Whittinghill, 2002). The problems driving this call have been complex and many, and among them are two long-running issues involving student conceptual understanding and application of knowledge (Tishkovskaya & Lancaster, 2012). Research consistently has indicated that even though students have successfully completed statistics courses, they often display poor conceptual understanding. Students also seem to have great difficulty applying what they have learned (Aliaga et al., 2005; Garfield & Ben-Zvi, 2007; Schau & Mattern, 1997; Tishkovskaya & Lancaster, 2012; Zieffler et al., 2008).

The need for change in light of these problems has become even more pressing given that statistics learning is being asked of a quickly growing number of students. Student enrollment in introductory statistics courses in two- or four-year colleges/universities in the fall of 2010 had increased 56% in just five years to 231,000 students (Blair, Kirkman, & Maxwell, 2013). In addition, because some sort of quantitative reasoning is required for completion of many majors, many of these students were from disciplines other than math. These non-math majors represented 35% of the total group, and their numbers had seen a similar rise of 50% in five years (Blair et al., 2013; Garfield et al., 2002). Statistics study at the secondary level has also shown great increases. One indicator of this can be seen in enrollment numbers for AP exams. In
2015-2016, over 2.5 million students participated in the AP Statistics Exam – a 95% increase in just ten years and a 32% increase in the past five years (College Board, 2016).

Facing rapid growth in the number of students taking statistics and the consensus that statistics instruction needed change, statistics educators came together to draft recommendations to improve statistics content and instruction. In 2005, the statistics education community formalized these calls for change into a set of guidelines, which charted a new direction for college level, introductory statistics courses (GAISE College Report ASA Revision Committee [GAISE], 2016). The *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) College Report addressed the topics that should be emphasized in introductory statistics education, and this report was widely endorsed by the American Statistical Association (ASA) and the broader statistics education community (Aliaga et al., 2005). It was updated in 2016 and continues to provide both focus and direction for statistics education (GAISE, 2016).

However, while this report made overarching goals clear, it did not dictate one instructional path to these goals. Educators have responded by approaching these goals in two distinct ways, indicating a clear divide in how to approach and teach statistics. One approach has prioritized content focused on rules, formulas, and statistical tests, and this approach has resulted in courses that involve direct instruction/lecture, mathematical emphases, and formulas, and that rely on a “consensus curriculum” (Zieffler, Park, Garfield, delMas, & Bjornsdotir, 2012). The other approach has prioritized the teaching of statistics through engagement with real-word applications. This approach often has involved learner-centered, constructivist methods, which might include the use of group work, simulations, and case studies (Tintle, VanderStoep, Holmes, Quisenberry, &
Swanson, 2011). These two distinct approaches have allowed for much debate across many academic areas – learning theory, psychology, mathematics, and statistics, to name a few – about how best to teach statistics and accomplish the proposed GAISE goals (Zieffler et al., 2008). As a result of these tensions, the field of statistics education itself remains an emerging one (Aliaga et al., 2005; Cobb, 1993; GAISE, 2016; Garfield et al., 2002).

One suggestion for change that has been discussed at length in the literature proposes moving the overall focus away from a concentration on mathematics, problem sets, and formulas, and toward a focus emphasizing statistical literacy, reasoning, and thinking (Aliaga et al., 2005; GAISE, 2016; Garfield & Ben-Zvi, 2007). In fact, the GAISE College Report (2016) listed the teaching of statistical thinking – through investigation and problem solving – as its first recommended focal area for statistics teaching and a focus on conceptual understanding as its second recommendation. The report also recommended a focus on active learning as its fourth recommendation (GAISE, 2016). Proponents of these changes have argued that a revised focus will address the problem of persistent student misconceptions, and that the strong conceptual understandings that result will improve problem solving as well (GAISE, 2016; Garfield & Ben-Zvi, 2007; Schau & Mattern, 1997).

The research base regarding this type of active learning in statistics has continued to grow, and thorough reviews of the existing literature such as those done by Garfield and Ben-Zvi (2007) and Zieffler et al. (2008) have offered strong summaries of what seems to be effective so far. In addition to these summaries, more recent research on instruction has investigated the use of real life examples drawn from places such as
weather reports, election polling, and sports (Mvududu & Kanyongo, 2011); guiding questions drawn from crime scenes and mystery stories to pique interest (Leavy & Hourigan, 2015); active learning by students in pairs or in small groups versus traditional lecture (Loveland, 2014); simulations and technology tools (Chance, Ben-Zvi, Garfield, & Medina, 2007; Rubin, 2007); and one-minute papers at the end of each class to articulate student understanding (Chiou, Wang, & Lee, 2014). In addition, not only did such constructivist methods seem to support statistics learning, they also have been associated with reduced student anxiety toward the subject (Mvududu, 2003; Pan & Tang, 2004; Sharma, 2015).

Concept Maps are an additional educational tool that educators can use to support this change in focus. These maps involve the development and organization of conceptual understanding and require active learning (Novak, 2010; Novak & Cañas, 2007; Schau & Mattern, 1997). Educators across many disciplines have used them successfully to support planning, instruction, learning, and assessment – in nursing (Harrison & Gibbons, 2013), medicine (Cutrer, Castro, Roy, & Turner, 2011), physiotherapy (Joseph, Conradsson, Wikmar, & Rowe, 2017), engineering (Watson, Pelkey, Noyes, & Rodgers, 2016), psychiatry (Hay, Wells, & Kinchin, 2008), business (Chiou, 2009), pre-service teacher education (Poling, Goodson-Espy, Dean, Lynch-Davis, & Quickenton, 2015), and reading (Chang, Sung, & Chen, 2002), to name a few. Concept Maps have especially strong roots in science education (Nesbit & Adesope, 2006; Novak & Cañas, 2007). They have been associated with increased student engagement (Horton et al., 1993; Nesbit & Adesope, 2006) and positively perceived in many respects by students (Harrison & Gibbons, 2013; Witmer, 2015). When used to assess learning, they target higher order
elements of student understanding such as conceptual relationships and deeper, meaningful connections (Nesbit & Adesope, 2006; Novak, 2010; Novak & Gowin, 1984) and provide unique insights into student understanding (Shavelson, Ruiz-Primo, & Wiley, 2005). In addition, some Concept Maps have been found to correlate well with traditional assessments, although this finding is not consistently represented across studies (Acton, Johnson, & Goldsmith, 1994; Doorn & O’Brien, 2007; Schau & Mattern, 1997).

However, despite the associated benefits, Concept Maps have practical limitations that have restricted their use by both students and teachers. For students, the fact that Concept Maps can cover wide ranges of content has caused many to feel lost when encountering a complex map with many interrelationships (Daley, 2004; Schau & Mattern, 1997). Students also have found the mapping technique time-consuming to learn, consider, and create (Schau & Mattern, 1997; Schau, Mattern, Zeilik, Teague, & Weber, 2001). In addition, the lack of structure can overtax student thinking leading to cognitive overload and negative interference in the learning process (Chang, Sung, & Chen, 2001; Katayama & Robinson, 2000).

For teachers, Concept Maps have presented real implementation challenges. Maps can be difficult to score because there are multiple ways to correctly represent concepts, and there is often no single correct map structure (Shavelson, Lang, & Lewin, 1994; Wu, Hwang, Milrad, Ke, & Huang, 2012). A non-standardized structure also means each student will submit a unique map, which requires individualized scoring by the teacher and is time consuming. Having to score many unique maps also can lead to a lack of consistency between scores (Harrison & Gibbons, 2013; Wu et al., 2012). In addition,
Concept Maps are time consuming and inefficient for instructors to construct and use for assessing student learning (Chang et al., 2001).

Finally, one of the greatest practical struggles for teachers has involved the provision of feedback. Students must receive timely and relevant feedback on initial iterations of Concept Maps so that they can fix misconceptions and grow beyond an initial conceptual structure, but this feedback requirement demands a large investment of time and effort by the teacher (Chang et al., 2001; Wu et al., 2012).

There have been a few alternatives to traditional Concept Maps that have attempted to address these limitations. Fill-In Concept Maps (FICMs) provide a fully formed, teacher constructed Concept Map or “Expert Map” that has some concepts and relationships given and some concepts and relationships missing for the student to fill in. The entire structure has already been determined, and it becomes the students’ job not to create the map but to fill in empty concept circles or label missing relationships (Schau et al., 2001). While this type of map may be easier for instructors to score and give feedback on (Chang et al., 2001; Schau et al., 2001), the provided structure also eliminates the students’ role in articulating conceptual relationships. Highly scaffolded maps such as these move away from meaningful learning as they are more directed at matching a pre-existing structure and require little connection with student prior knowledge or knowledge structures (Chang et al., 2002).

Computer-Aided Concept Maps (CACMs) have tried to address these limitations as well. The research regarding the benefits of various computer systems is still emerging, but they have some immediate benefits over paper and pencil maps when it comes to scoring and feedback (Schau et al., 2001). Because computer systems use a pre-
existing map as a completion goal, they result in a structure that is identical among
students; this eliminates many of the scoring concerns found with traditional Concept
Maps (Chang et al., 2002). In addition, some CACMs can provide immediate feedback as
a student works to complete a map (Chang et al., 2002). This feedback is auto-generated
when students make incorrect entries, and it can be in the form of questions, prompts, or
links to supplementary readings (Chang et al., 2002). However, while convenient for the
instructor, this feedback has had mixed results in terms of effectiveness for learning
(Chang et al., 2002).

In the end, though, despite the benefits of offering more scaffolds than traditional
Concept Maps, CACMs have an end goal of a student map matching a pre-existing expert
map. While the feedback and provided hints result in immediate changes and help with
map completion, in the end, these maps were another learning tool which focused
conceptual representations into one pre-determined structure.

**Focus**

This exploratory study took a practical approach to the use of Concept Maps. It
focused on traditional paper and pencil maps, as opposed to computer-generated Concept
Maps, and it focused only on their role as learning tools. It further focused on the role and
use of feedback (Expert and Peer) in the refinement of Concept Map structure.

**Purpose**

The primary purpose of this study was to examine the effects of Expert Feedback
and Peer Feedback on the structural characteristics of Concept Maps created by students
in an undergraduate level statistics course. Concept Maps can reflect a picture of student
conceptual understanding and growth (Hay et al., 2008), and this study sought to explore
the effects of two types of feedback on the conceptual structures held by students and represented by their Concept Maps.

Research on theories of expertise in problem solving has seemed to indicate that experts structure knowledge differently than novices (Chi, Glaser, & Rees, 1982). This difference in knowledge structure leads experts to approach problem solving differently, more efficiently, and through the use of problem/solution representations that are not available to the novice (Chi et al., 1982). It may seem as if some of the difference in solution successes might be due to differences in recognizing cues within a problem (cues that suggest certain ways of solving), but Chi, Glaser, and Rees (1982) found that both novices and experts are similarly adept at recognizing existing cues. The difference, they found, was that experts were more proficient in making inferences about solutions given these cues (Chi et al., 1982). The researchers concluded that experts differed from novices in the range of representations they had access to based on their prior experiences within the relevant domain (Chi et al., 1982).

This study presents the use of an Expert Map as one type of feedback. The purpose of this Expert Map was not to have students attempt to memorize and match the Expert Map. The purpose was also not to have student understanding move from novice to expert after a single feedback session. Rather, in line with the findings of Chi et al. (1982), the purpose of the Expert Map was to make visible a wider range of conceptual representations and relationships so that students might consider how these representations compare with their existing understandings. If it is the case that students must struggle with and incorporate new ideas into existing frameworks of understanding in order for meaningful learning to have occurred (Novak, 2010), then the purpose of the
Expert Map was to make visible new conceptual relationships so that students might struggle with disparities between their versions and the Expert version as a means toward more meaningful understanding. This study also involved peer feedback and, as with the Expert Map, the purpose of this feedback was to expose students to new ways of envisioning relationships between concepts.

**Approach**

Student pre-feedback Concept Maps were compared to post-feedback Concept Maps and analyzed for changes in organizational characteristics and conceptual complexity. The Concept Maps provided three sets of data. All Concept Maps were given quantitative scores using a well-used scoring system. Maps were also given a structural classification – Discrete or Integrated. Finally, the researcher analyzed student choice of starting terms and recorded these frequencies.

The quantitative score was based on the components included in a map. This score gathered the number of concepts, linking words, and cross-links into a final numerical total. The scoring system used in this study was one of many that are present throughout the literature and was a variation of the scoring system proposed by Novak (2010). Differences in pre- and post-feedback quantitative scores indicated changes in the complexity of student Concept Maps, which were used as indicators of changes in the complexity of student conceptual organization. These differences then were analyzed for statistical and practical significance using a mixed repeated measures ANOVA. This part of the study attempted to determine whether significant differences existed between the scores of Concept Maps made after students received Expert Map Feedback and the scores of Concept Maps made after students received Peer Feedback.
Categorical structural classifications reflected the overall structure of the map using one of two designations – Discrete or Integrated. This categorization approach stemmed from the work of Kinchin, Hay, and Adams (2000), whose classifications consisted of Chain, Spoke, and Network. For this study, the categorization of Chain was excluded as it was too simplistic for the maps that were made. One benefit of the convergent mixed methods design chosen for this study was that it allowed for the merging of data for richer analyses. The structural classifications were merged with the questionnaire response data to more fully understand the intricacies of the responses. This also was the case with student choice of starting terms, which were counted, analyzed separately, and then merged for further analysis.

This study also explored student perceptions after the mapping and feedback process using a 12-item questionnaire. Students were asked questions about the role and perceived value of feedback in their learning and mapping processes. They also were asked about their perceptions of the usefulness of Concept Maps in the learning of statistics. Answers provided both quantitative data (10 responses using Likert response scales) and qualitative data (two short-answer responses). The Likert scale responses were analyzed according to frequency, and qualitative responses were compiled and analyzed for themes. Then, these responses were merged with the quantitative scores as well as the structural classifications with the hope of illuminating patterns and enriching understanding of responses.

The researcher created the short questionnaire (see Appendix A for Student Questionnaire). While there were existing student perception surveys that had documented high reliability and validity in the area of statistics education and anxiety –
for example, the Survey of Students’ Attitudes Toward Statistics-36 [SATS-36] (Cruise & Wilkins, 1980) and the Statistics Anxiety Rating Scale [STARS] (Schau, 2003) – the constructs measured in these surveys did not align with the current study’s purposes. The goal of the survey in this study was to elicit student perceptions on the value of feedback and the usefulness of Concept Maps given specific methods of feedback. The results were not meant to generalize to a larger population; instead, the results were designed to inform and illuminate the existing data.

**Significance of the Study**

This study brought together research strands from learning theory, statistics education, and Concept Mapping and extended the research base for each.

**Learning theory.** This study built on learning theory research specifically in the area of Constructivism. Concept Maps as learning supports allow students to actively create meaning, integrate learning with prior knowledge, and practice meaningful (as opposed to rote) learning (Novak, 2010; Novak & Cañas, 2007).

This study also advanced research on the role of feedback on conceptual understanding and meaningful learning. It specifically addressed the effect of Concept Maps/Expert Maps used as feedback to aid conceptual development in statistics learning, a specific use and area that is not fully explored within the literature.

**Statistics education.** This study has import for the field of statistics education in its exploration of how best to support learning in statistics courses. The curriculum of the selected course closely followed the textbook content and heavily emphasized a rules-, mathematics-, and formulas-approach, and the teaching methodologies used in the classroom involved primarily direct instruction (Moore, Notz, & Fligner, 2015). All
statistical applications were focused within the realm of science. However, although it
was more traditional in its approach, this course still attempted to teach statistics literacy,
and, as noted previously, the research base is still emerging on how best to support
conceptual understandings (Mvududu, 2003; Schield, 2004; Tintle et al., 2011; Zieffler et
al., 2012). This study adds to newly emerging research in statistics education by
exploring the use and effects of Concept Maps with feedback as learning tools. It may be
that these learning tools help bridge the gap between the two seemingly disparate
statistics directions.

**Concept Mapping.** This study focused on Concept Maps as learning tools. It
specifically focused on the role of Expert Map Feedback and Peer Feedback in advancing
student conceptual structures in a statistics course, elements which have not been
explored in this manner. It also provided a new possibility for the practical use of
Concept Maps in classroom settings through its exploration of a time-saving, easy-to-
implement feedback strategy for improving student learning. Finally, it answered a call
for more research to explore Concept Map use in small groups and to explore the ways in
which students use maps to learn (Nesbit & Adesope, 2006).

**Research Questions**

This study used a Convergent Mixed Methods design to attempt to answer three
research questions:

Question 1: Is there a significant difference between the scores for Concept Maps
made after students receive Expert Map Feedback and the scores for Concept
Maps made after students receive Peer Feedback?
Question 2: Do student perceptions of the usefulness of Concept Maps for statistics learning and of the value of feedback they experienced differ based on the type of feedback they experienced - Expert Map Feedback versus Peer Feedback?

Question 3: Are there significant differences between groups when structural classification and student choice of starting terms is considered?

The researcher collected both quantitative and qualitative data in one time period, analyzed the data separately, then merged relevant portions of the data with the goal of using all findings to understand the results and practical implications of the research more fully (Creswell, 2013; Creswell & Plano Clark, 2018). This design allowed for an enriched exploration of the relationship between feedback and student learning through its inclusion of both types of data as well as the merging of relevant parts of each data set. The different types of data both informed and expanded on one another.

This approach supported one pragmatic, underlying purpose, which was to explore more practical means of using Concept Maps in the classroom. While quantitative scores and analyses of differences provided information on understanding as expressed through Concept Maps, they did not address student perceptions of this learning tool. This additional perspective was valuable in light of the underlying practical nature of the study and the fact that students, in the end, would be the desired users of this learning support.

**Structure of the Dissertation**

This dissertation is divided into four subsequent chapters: Review of Literature, Research Methodology, Results, and Discussion of Results and Conclusions.
The Review of Literature chapter (Chapter Two) contains definitions along with a presentation of broad themes in the existing literature of Learning Theory, statistics education, and Concept Mapping, as they relate to this study. The Review of Literature also provides more specific research background and support specifically related to this study through the use of representative studies (both qualitative and quantitative).

The Research Methodology chapter (Chapter Three) comprises a description of the methodology used in this study. It details the research design, the participants and sampling process, validity and reliability measures, procedures, and statistical analyses.

The Results chapter (Chapter Four) includes a detailed summary of the study results and is organized according to research question. The summary presents results in both narrative and table format, and content includes descriptive statistics, assumptions and respective tests, and a summary of significant findings.

The Discussion of Results and Conclusions chapter (Chapter Five) presents key findings and is organized according to research question. Discussion of findings includes discoveries, connection of findings with existing research results, and any implications for practice and for theory. This chapter also provides a discussion of limitations, suggestions for further research, and the overall conclusion.
Chapter Two
Review of Literature

Definitions

**Concept Maps.** Concept Maps are pictorial tools for organizing and displaying concepts and their interrelationships (Davies, 2011). They were developed by Novak (2010) in 1972 in the course of a research project in science which sought to visualize changes in student understanding of science concepts. Concept Maps became one way for students to display their conceptual understandings at different points in time while also allowing researchers to observe changes in student knowledge (Novak & Cañas, 2007). Although the shapes and connecting lines may seem at first glance to indicate simple student drawings, Concept Maps manifest learning at a level much deeper than a drawing and work in unique ways with learners to show conceptual understanding (Davies, 2011; Novak, 2010). Concept Maps involve students, prior knowledge, concepts, relationships, context, and structure, among other elements; and they were one facet of an entire theory on learning developed by Novak (2010) in the 1970s and refined again for educators and businesses in 2010. They involve interconnections, active learning, and constructing meaning. The final result not only reveals knowledge structures, but it empowers (Novak, 2010; Novak and Cañas, 2007). Concept Maps consist of four primary elements: concepts, linking words, propositions, and a focus question. Novak and Cañas (2007) created the Concept Map in Figure 1 to illustrate these learning tools, and it provides a useful reference for the following discussion of elements.
**Concepts.** A concept was defined very specifically by Novak (2010) as “a perceived regularity or pattern in events or objects, or records of events or objects, designated by a label” (p. 42). Concept learning represents a deeper learning than the learning represented by a name or a definition. While a definition provides a representational understanding, students reach deeper conceptual understanding when they learn what a word means or learn the pattern that it stands for. Novak (2010) used the word “carnivore” to illustrate the difference (see p. 43). A student might read the word “carnivore” in a sentence about lions and understand the definition of carnivore when it is given in relation to this animal. Simply understanding the word in this context was representational understanding. However, if the student then could see a pattern in the idea of “carnivore” that allowed a connection of this idea with other animals, then the pattern of the concept of carnivore had been understood; this represented conceptual
understanding (Novak, 2010, p. 43). In Figure 1, concepts are represented by the words in rectangles; for example, “Knowledge,” “Teaching,” and “Symbols” are concepts.

**Linking words.** Linking words are the words that are used between concepts in a Concept Map to indicate relatedness. The presence of these relationship words marks the difference between other mental representations such as mind maps – which do not include linking words – and Concept Maps. More important, linking words are primary indicators of the depth of student conceptual understanding (Novak & Cañas, 2007). Students who poorly understood concepts or the relationships between them would struggle to find appropriate linking words because it was these linking words that indicated a more complex understanding.

In addition, Concept Maps involve cross-links, which are linking words that connect concepts that are in separate conceptual groupings. Cross-links illustrate an understanding of the bigger organizational framework in which the concepts operate (Novak, 2010). In Figure 1, linking words are represented by the words that connect two rectangles; for example, “aids” and “needed to see” are linking words. Also in Figure 1, there are cross-links such as “is a basis for,” which connects the concept of “perceived regularities” with the related but separate thread of “creativity.”

**Propositions.** Propositions consist of the grouping of concept-linking word-concept, and Novak (2010) described them as “the fundamental units of meaning” (p. 26). These groupings could reflect a rich and deep understanding, as determined by the specificity and sophistication of the linking words, or they could reflect a superficial understanding as indicated by overly simplistic or no linking words. The quality of propositions depends in large part on the links made between concepts. When students
struggle with finding conceptual connections, it means that they do not understand concepts or relationships in depth. This superficial understanding would be reflected in their Concept Map propositions (Novak, 2010). In Figure 1, “Concepts…connected using…Linking Words” reflects a proposition, as does “Organized knowledge…necessary for…Effective Learning.”

**Focus question.** Focus questions are provided to guide the development of a Concept Map, and there are some question formats that are more effective than others at doing this (Novak & Cañas, 2007). Generally, Concept Map quality has improved when focus questions require explanation as opposed to simple classification (Novak & Cañas, 2007, p. 34). Questions that began with, “Why do…?” “How can…?” or “How would you explain…?” would lead to conceptually stronger maps than questions that began with “What are?” (Novak & Cañas, 2007, p. 34). While focus questions have been crucial for determining the direction of Concept Maps, they rarely have been provided in the map itself. This was true even of Novak’s initial maps; Figure 1, for example, was published in 2007 without the focus question, while a more recent edition included it (Novak, 2010; Novak & Cañas, 2007). Updated discussions now suggest the inclusion of the focus question on all maps for purposes of clarity (Novak, 2010).

The general process for creating a Concept Map began with the focus question, then moved to students determining relevant concepts, creating propositions, and then organizing all together – concepts, links and cross-links – into hierarchies, beginning at the top with the broadest conceptual categories and proceeding down with more specificity (Davies, 2011). Concept Maps did not necessarily have to be arranged hierarchically, but they did have to be structured around relationships, unlike Mind Maps,
which could be more free-form and exploratory (Davies, 2011). It is important to note that Concept Mapping is a process, involving iterations and adjustments, and revision was always necessary to strengthen and refine conceptual relationships (Novak & Cañas, 2007).

**Feedback.** The purpose of formative feedback is to improve learning. This has been done by conveying information to a learner that will help the learner make adjustments and improve (Shute, 2008). Results from extensive research on feedback have indicated that feedback has the potential to significantly affect learning (Hattie & Timperley, 2007; Shute, 2008). Researchers are still exploring how feedback works, but it does seem clear that many other variables – related to the task, the teacher, and the student, among others – can greatly affect outcomes (Hattie & Timperley, 2007; Shute, 2008). For this reason, despite the quantity of existing research on feedback, there have been few consistent conclusions (Hattie & Gan, 2011; Hattie & Timperley, 2007; Kluger & DeNisi, 1996; Shute, 2008). However, there have been some general characteristics that seem to apply to successful feedback, even across studies that hold differing assumptions about learning and about how feedback should operate in the classroom (Hattie & Gan, 2011; Hattie & Timperley, 2007). Because the research base is extensive, this review has focused only on the limited issues regarding feedback that were pertinent to the study at hand.

In a review exploring feedback and instruction, Hattie and Gan (2011) listed eight characteristics of successful feedback gleaned from the existing literature, four of which apply to this study: 1) feedback should clearly indicate the criteria needed to move forward to achieve learning goals; 2) it should cue the student to the task as well as to
strategies necessary for the task and for self regulation; 3) it should be “calibrated” so that the learner is engaged at an appropriate level – matching the learner’s level or slightly above; and 4) peer feedback not only engages students in active learning, it provides a way for them to take ownership of their learning process (p. 250). These characteristics alone would not ensure successful feedback, however, because there should also be some mechanism in place to ensure that students hear and process given feedback. However, these characteristics provide a useful structure for the concept.

It is important to note that feedback can function differently depending upon the epistemological assumptions about learning that are held by the researcher or the teacher in the classroom (Hattie & Gan, 2011). This study was premised on the socio-cultural constructivist perspective discussed by Vygotsky (1978) in his theory of social interaction. As a result, it focused on the ways in which feedback was informed by a socially mediated construction of knowledge. The use of peers as a source of feedback, then, played an especially important role, as peer groups contribute to learning that is socially mediated and created. This type of feedback incorporates the use of meaningful dialogues and promotes negotiation and co-construction of meaning (Hattie & Gan, 2011). In addition, one of the characteristics of quality feedback, according to Hattie and Gan (2011), was the way in which it promoted active participation by its members and encouraged student ownership of learning. Peer feedback created a dynamic learning situation where students explored and created meaning together (Hattie & Gan, 2011).

In addition, feedback can be considered a scaffold, which connects with students in their zones of proximal development and helps them to move toward greater independent learning (Hattie & Gan, 2011; Vygotsky, 1978). The Expert Map that was
provided allowed students to choose whether they would like to access this information and at what levels. The peer feedback developed based on student discussions with one another and through their questions to each other. This aligns with the “calibration” of successful feedback to a level that matched learner levels or was slightly higher (Hattie & Gan, 2011).

There are also various forms that feedback can take, from the simplest form – “verification” – which provides right/wrong results to the most complex form – “strategic tutoring” – which provides right/wrong results and individually tailored hints about how to proceed (Shute, 2008). The use of peer groups, discussion, and in some groups the use of an Expert Map situated the feedback used in this study near the highest level of complexity. Not only did this feedback include error analysis, but peers may have provided elaborate feedback that was student specific, provided hints or clues about how to restructure ideas, and engaged in discussions that targeted specific misconceptions (Shute, 2008).

Finally, the timing of feedback is another element that might affect feedback success, and while the research base has been extensive in this area, as with feedback in general the results have remained inconclusive (Hattie & Timperley, 2007; Kulik & Kulik, 1988). This may be due in part to differing effects based on the complexity of the assignment. For instance, complex assignments might require detailed feedback, which takes time to generate but which has the potential to strongly impact student learning. Simple assignments might allow for simple and quickly generated feedback, which might be immediate but which might have less impact on learning. Additionally, there should be some mechanism in place to ensure that students read and respond to feedback. Some
patterns have emerged, though, and they have indicated that for applied studies involving real classroom learning, immediate feedback can positively affect learning (Kulik & Kulik, 1988). In a meta-analysis of feedback timing involving 11 applied feedback classroom settings, student scores on criterion measures indicated greater achievement when students received immediate feedback (Kulik & Kulik, 1988). The effect size was .28, \( t(10) = 4.81, p < .01 \), indicating that student achievement increased by .28 of a standard deviation when immediate feedback was provided (Kulik & Kulik, 1988). These findings, while encouraging, were limited by the small number of cases and, as a result, the strength of the conclusions was limited (Kulik & Kulik, 1988).

**Theoretical foundations**

Concept Maps are learning tools that make conceptual understandings visible. Theoretically, they have drawn support from constructivist epistemology, specifically social constructivism, and from Ausubel’s learning psychology. As epistemologies and learning theories are complex, the following discussions are brief and focused only on elements directly relevant to this study.

**Social constructivism.** Constructivism’s foundational idea has been that knowledge is constructed by learners who take in new information and modify it in an interplay with their existing prior knowledge. It is an active process, and requires learners to engage with learning, not simply observe. Social constructivism is one type of constructivism, heavily influenced by the work of Vygotsky (1978), and as the name suggests it emphasizes a social dimension to knowledge construction. In this type of constructivism, learners construct meaning, but they have constructed it together with others in social and cultural contexts, and then they have internalized this socially created
meaning (Chaiklin, 2003). Social constructivism has been considered both an external and internal process because learning happens together and meaning is socially constructed, but learners then internalize these ideas through an integration into their own prior learning structures (Novak, 2010). This has been one idea that theoretically supports the use of Concept Maps as learning tools in the classroom (Chaiklin, 2003; Novak, 2010).

Another element that has supported this use involves a learner’s ideal range for learning. Vygotsky (1978) called this range the “zone of proximal development,” and from this, the idea of scaffolding learning developed (Chaiklin, 2003). The zone of proximal development was described by Vygotsky (1978) in his theory on child development, and it refers to a zone which is created by the difference between what a learner can learn alone and what a learner can learn collaboratively or with assistance (Chaiklin, 2003). The zone was an area that existed as an evaluation of a learner, and it spoke to a range of perceived capabilities (Chaiklin, 2003). In practical application, the goal for an educator would be to develop a sense of a learner’s zone of proximal development so that learning activities could be structured to challenge this learner in the most appropriate manner (Chaiklin, 2003).

Scaffolding was one means of supporting learning, and it was a practical way to help learners progress through their respective zones. It refers to learning supports that an instructor or collaborator has put into place to help students build necessary learning and skills, with the thought that the supports would slowly be removed to allow for students to then learn on their own (Novak, 2010). Concept Maps are a type of learning scaffold that allows learners to operate within their zones of proximal development. These maps
have the potential to help learners meaningfully learn new material, and they are tools that can be used to empower learners (Novak, 2010).

**Ausubel’s learning psychology.** Ausubel’s learning psychology works in conjunction with the general constructivist position, but it is important to note that Ausubel’s theoretical focus connected not with regard to the construction of knowledge through discovery processes but rather with regard to the level of meaningfulness experienced in learning (Ausubel, 1966b). Ausubel’s early theories appeared in the article, “In Defense of Verbal Learning,” published in 1961 and republished in 1966, and were formalized in the book, *The Psychology of Meaningful Verbal Learning* (1963). Ausubel refined these thoughts in subsequent articles in journals of educational psychology and educational theory (Novak, 2010). Many of the early arguments appeared alongside defenses of what were termed “verbal learning,” ideas which were set in direct opposition to Bruner’s theories on “discovery learning” (Ausubel, 1966b). Verbal learning referred to a type of learning where the content had been organized ahead of time and presented to learners, usually through lecture, and Ausubel termed the learning done by students as “reception learning” (Ausubel, 1966b). Ausubel argued that educators and theorists had over-focused on the idea of discovery learning, and that verbal learning was as important (Ausubel, 1966b). Critics had portrayed verbal learning as “parrot-like recitation and rote memorization of isolated facts” (Ausubel, 1966b, p. 87) and Ausubel acknowledged that these problems did exist in poorly run classrooms, but also noted the very real problems that had accompanied discovery learning. Verbal learning (through lecture and presentation by the teacher and the reception of ideas by the
student) could be both powerful and effective, and Ausubel argued that the learning method should be separated from the meaningfulness of the learning (Ausubel, 1966b).

For Ausubel, more important and distinct from the verbal/discovery dichotomy was the idea of the meaningfulness of the learning itself (Ausubel, 1966b). Ausubel (1966a) asserted that there were two types of learning, meaningful learning and rote learning, and that the difference involved the construction of understanding along with the learner’s commitment to learning (Novak & Cañas, 2007). Meaningful or rote learning could occur when using any type of learning, and with this argument, Ausubel attempted to establish the greater importance of the meaningfulness of the learning over the method of learning (Novak, 2010).

Meaningful learning involved learners choosing to learn in a way that brought new information into meaningful relationship with what the learners already knew. Learners had to actively bring new ideas into interplay with pre-existing knowledge, and this would integrate new knowledge into their existing conceptual understanding. Rote learning was in contrast to this. With rote learning, learners acquired information in individual and unrelated bits, and they did not integrate new ideas into what they already knew (Novak & Cañas, 2007).

Materials learned in these two ways were organized differently in a learner’s cognitive structure (Ausubel, 1966a). Meaningful learning resulted in material that was integrated into a learner’s preexisting structure of concepts and understandings, whereas rote learning resulted in “discrete and relatively isolated entities which were only relatable to cognitive structure in an arbitrary, verbatim fashion” (Ausubel, 1966a, p. 104). The goal was for new material to interact with a learner’s prior knowledge, and
eventually be incorporated into the learner’s existing structure of understanding, a process Ausubel (1966b) termed “subsumption.” This cognitive approach assumed that learners organized knowledge in a hierarchical way, with general concepts incorporating more specific ones, with relationships existing between all to further refine meanings (Ausubel, 1966a).

It is important to note that in Ausubel’s theory, it was not only the approach to learning that determined whether learning was meaningful or rote; the student had to bring an intention to learn. This meant that the student had to do more than simply attempt to get an answer or memorize new content. If this intent were missing, then new content would not be assimilated into existing understandings (Ausubel, 1966b). For this reason, even a student faced with a discovery task had the potential to finish with no more than rote learning (Ausubel, 1966b). On this point, Ausubel’s theory differed from Bruner’s theory of discovery learning (Ausubel, 1968).

Ausubel’s central ideas of meaningful learning and rote learning connect in an important way with constructivist epistemology. The foundation of constructivist epistemology warrants that knowledge is constructed by the learner and that this construction involves both knowledge and learner affect; Ausubel’s meaningful learning needed to connect in a meaningful way with learners’ conceptual frameworks, and learners had to actively construct this connection (Ausubel, 1966b).

Ausubel’s learning theory and Novak’s later adaptation of this theory for the rationale behind Concept Maps both centered upon the idea of constructing knowledge (Novak & Cañas, 2007). Novak (2010) even argued that knowledge creation was the same as meaningful learning, because it involved learners bringing new concepts into
existing knowledge networks and constructing meaning and relationships between them (Novak & Cañas, 2007).

**Connections with Concept Maps.** Concept Maps align with the tenets of social constructivist epistemology and were designed in accordance with Ausubel’s theory of meaningful learning (Novak, 2010). For this study, Concept Map creation and feedback occurred in a collaborative setting to encourage the creation of shared meaning as well as the connection of learning with individual prior knowledge structures. The provision of an Expert Map for feedback not only allowed students to operate within their zones of proximal development, it allowed them to study new material and incorporate it into their unique understandings. Finally, the activity of Concept Mapping had the potential to engage learners, which could increase their commitment to learning, an engagement required according to constructivist epistemology and discussed by Ausubel as necessary for meaningful learning.

**Overview of the Research Literature on Concept Maps**

Educators have used Concept Maps to support teaching, learning, and assessment, and research results consistently have shown that the use of these maps has the potential to benefit learners because of the way they involve students and engage them in identifying conceptual relationships (Horton et al., 1993; Nesbit & Adesope, 2006). However, the research base has exhibited some limitations. Historically, researchers studied the effectiveness of Concept Maps primarily on achievement, but little clarification was provided about how this construct was being defined and measured. In addition, research seemed to be done in broad strokes, with large groups made up of very different populations grouped together for results. Current research trends have been
moving in the direction of greater specificity of focus and definition. In addition, researchers have expanded their interests to focus on maps themselves, analyzing elements such as structure and function. New questions have explored effects from adjusting map elements, involving technology, or experimenting with feedback, among other things, and these ideas are reshaping the direction of research in new and interesting ways. This section presents a summary and critical analysis of the research that has emerged on Concept Maps. Two meta-analyses, done 13 years apart, provide a convenient means of summarizing the broader research themes, and a selection of representative empirical studies addresses some research directions in more detail.

**Early research directions.** Early research studies tended to focus heavily on Concept Map use in science learning, a use consistent with the earliest implementation of Concept Maps by Novak (Novak, 2010). Many of these early studies reported positive effects of using these maps in the classroom, but the quality of studies tended to be inconsistent, at best, and many details needed for replication or comparison were not reported (Horton et al., 1993). An early and widely referenced meta-analysis by Horton et al. (1993) explored the research addressing the effects of Concept Mapping as an instructional tool on student achievement, and these limitations became clear as the research base was aggregated and analyzed as a whole.

Researchers began with 133 studies located through a thorough search of the literature from the prior 10 years. However, out of 133 studies, they accepted only 18 for their meta-analysis, eliminating the remainder because they had not met their inclusion criteria (Horton et al., 1993). To be included in this meta-analysis, studies had to have occurred in actual classrooms, they had to have used Concept Maps as an instructional
tool and measured the effectiveness using a quantitatively measured outcome, they had to have included a control group in their design against which to compare the outcomes from the treatment group, and they had to have either provided an effect size or produced sufficient data for the calculation of one (Horton et al., 1993). According to the authors, the control group criterion eliminated the majority of the cases not used (Horton et al., 1993). Results were grouped according to the type of dependent variable measure: achievement or affect.

To determine effect sizes, researchers used the Glass’s Δ equation. This effect size measure involves the difference in mean score between the control group and the experimental group, divided by the standard deviation of the control group; and according to the authors, it was the best available effect size measure at the time (Horton et al., 1993).

From the meta-analysis results, researchers concluded that Concept Maps had generally positive results on student achievement, raising achievement an average of 0.46 standard deviations. This increase reflects a move from the 50th to the 68th percentile. These results confirmed their assumptions about the effectiveness of Concept Maps as instructional tools (Horton et al., 1993).

However, a number of issues plagued this early meta-analysis. First, the range of participant grade levels was relatively wide considering the small number of included studies. Studies ranged from fifth graders to college seniors, and each grade level had relatively few representations. For example, there were two studies from fifth grade, two from middle school, nine from high school, and five from college. In addition, it was unclear how many of these studies took place in the United States; three of the studies
took place in Nigeria and one in Taiwan, but the remainder were grouped as “United States and Canada.” There was no rationale given for this grouping and no identifying numbers regarding how many belonged to each country. This was an important point needing clarification, as there are many cultural and systemic differences between these two countries.

The academic discipline in which the studies were run was generally more consistent, with most involving some sort of science, although results also included studies involving social studies and reading comprehension. The sciences included Biology, Ecology, Earth science, Chemistry, Physics, and Physical science. Treatments of Concept Map creation also varied, with some created individually, in groups, by the instructor, by students, and by both students and instructor. Maps were created in class for eight studies, but the remaining seven studies either did not specify where or specified only that they were created out of class. Researchers had hoped to examine 11 variables for meta-analysis results, but most studies provided little to no information on the desired variables; researchers could only include two in their final analysis.

Finally, and most seriously, the measure of “achievement” was not indicated for any of the studies. Studies were either grouped as having had a dependent measure of “achievement” or “attitude,” but no further elaboration of these constructs was provided. Current research has indicated that these constructs must be clearly defined in order for meaningful conclusions to be drawn. The researchers had access to this information, as the coding sheet used to record study details and given in an appendix clearly asks for descriptions of the outcomes. Studies were analyzed for information such as “Format of outcome measure (multiple choice, essay, etc.),” “Test &/or scale name, or unpublished,”
and “Type of outcome (achievement, attitude, etc.)” – but no information about these details was provided, suggesting that a general classification of either achievement or attitude was sufficient (Horton et al., 1993).

Effect sizes showed great variation. For example, average effect sizes differed depending upon the subject matter of the study, with biology students using Concept Maps achieving at the 72\textsuperscript{nd} percentile on average while chemistry or physical science students achieved at the 56\textsuperscript{th} percentile (Horton et al., 1993). In addition, average effect sizes differed by country, with studies taking place in the United States and Canada showing average effect sizes of 0.49 and those in Nigeria and Taiwan showing an average effect size of 1.0. The researchers cautioned readers to consider the small number of studies in this meta-analysis when considering results, but other than this, the authors concluded that there was medium positive effect size on achievement from Concept Map use (Horton et al., 1993).

This meta-analysis indicates two primary problems seen throughout the early literature on Concept Maps -- not only was there a wide difference in the way researchers identified and measured and reported the dependent variable of “achievement” when dealing with Concept Maps, but there were multiple design issues and omissions of data (Horton et al., 1993). Researchers noted extensive omissions that would lead to replication issues, with studies lacking detail regarding length of treatment, which participants actually created the Concept Maps (teachers or students or both), and whether terms were supplied or not, to mention a few. Not discussed but also relevant to results is how maps were assessed or scored. There currently are multiple methods for scoring, each of which focuses on different elements and leads to different results (Rye &
Rubba, 2002; Taricani & Clariana, 2006; Watson et al., 2016). In light of these limitations, Horton et al. (1993) finished their analysis with a clear call for changes in the way future studies on Concept Maps were designed, conducted, and reported.

**More recent research directions.** A more recent meta-analysis by Nesbit and Adesope (2006) provided an updated view of research surrounding Concept Maps. This meta-analysis reflects a body of research that has incorporated many of the changes identified by Horton et al. (1993), and yet it is a still-emerging body of research literature. A thorough search of the literature yielded over 500 peer-reviewed research studies related to Concept Maps and mapping, and most had occurred in the 13 years since the Horton et al. (1993) review (Nesbit & Adesope, 2006). By this point in time, three research strands had emerged. One continued the research begun by Novak (2010) into student creation and use of Concept Maps, both individually and in groups (Nesbit & Adesope, 2006). Another strand considered the use of pre-made Concept Maps, a use that required students to fill-in missing concepts or linking words/relationships (Schau et al., 2001). A third strand explored Concept Map creation, use, and revision with technology (Novak, 2010; Novak & Cañas, 2007), and included such topics as ease of use, automatic and immediate feedback and/or correction (Chang et al., 2001; Wu et al., 2012).

In addition to the research strands regarding Concept Maps, related research explored theories of exactly why these learning tools seemed so effective. Nesbit and Adesope (2006) provided more complete summaries of these research directions, including theory and research on dual coding and conjoint retention (i.e., that learning is aided when maps are coded by the brain as pictures and accompany text), verbal coding (i.e., that learning is aided when maps are coded by the brain as texts), and a range of
learning theories regarding features of mapping that may have led to greater achievement, from student choice in processing order to summarization to main point extraction (Nesbit & Adesope, 2006).

Different types of learners and groupings were considered as well, with studies focused on low-achieving learners, learners with limited prior knowledge, students with low verbal ability, and collaborative and cooperative learning situations (Nesbit & Adesope, 2006). The range of studies and topics in this meta-reflection reflected a great maturation of the literature base over the course of 13 years.

Researchers narrowed their focus to 122 studies, which were narrowed even further to 55 studies, involving 5,818 participants (Nesbit & Adesope, 2006). When determining effect sizes, researchers first ensured that studies were statistically independent, and they devised a special coding scheme to ensure this independence even in studies when two treatment groups were compared against the same control group. They calculated Cohen’s $d$ for the standardized mean difference effect size, and from this calculated Hedges $g$, the unbiased estimate of the standardized mean difference effect size (Nesbit & Adesope, 2006). Around each effect size, they calculated a 95% confidence interval to determine statistical significance (Nesbit & Adesope, 2006).

Another difference indicative of a maturing research base was the attention paid to different outcome measures of the dependent variable “achievement” (Nesbit & Adesope, 2006). The Horton et al. (1993) meta-analysis considered this an implicitly understood construct, and they grouped all of the study results that did not involve affect under the classification of achievement. This grouping may have seemed less appropriate to researchers 13 years later who encountered a different research base. The widening
research base encountered by Nesbit and Adesope (2006) allowed for greater specificity within this category and as a result, more specificity of results. Researchers divided the construct into two types of outcomes – those involving students studying Concept Maps and those involving students constructing Concept Maps. Within these groups, results were categorized into two general groups determined by the purpose of the construct measure outcome – for retention only or for retention and transfer – and by the way in which the construct was tested (e.g., free or interview recall, objective, short answer, or mixed item types) (Nesbit & Adesope, 2006). These classifications were in contrast to the Horton et al. (1993) general outcome classifications of either achievement or attitude.

In general, the Nesbit and Adesope (2006) results were similar to those found by Horton et al. (1993), this even though the more recent meta-analysis included more focused research studies and a number of newly explored research directions. Despite these differences, researchers consistently found that Concept Maps were associated with achievement gains, even across a wide range of learning levels, subject areas, and settings (Nesbit & Adesope, 2006). More specifically they found that activities involving Concept Maps were more effective for helping students learn and retain knowledge than typical classroom activities such as reading and listening to lectures (Nesbit & Adesope, 2006). Researchers attributed some of this effect to the increased engagement that resulted from the classroom use of Concept Maps (Nesbit & Adesope, 2006).

**Summary.** The research base on Concept Mapping continues to mature. Not only has it been deepening within established research areas, but it has been broadening as well. In addition, the growing use of technology within education has changed not only how maps are used, but how they function. These and other issues have prompted a
further, detailed discussion of current research, which is organized in the next section according to various research considerations facing current researchers. The following section also uses empirical research to illustrate some of the relevant issues.

**Current Considerations and Empirical Studies**

**Student created maps versus pre-determined maps.** The literature results have repeatedly shown that Concept Maps offer strong scaffolds for student learning and have the potential to positively affect learning. They allow students to make visible their understanding of concepts and relationships, and the structure can be unique to each student (Novak, 2010). Despite these positive outcomes, educators have refrained from using them for a variety of practical reasons, many of which involve time-constraints. One reason, for example, involves the time it takes to teach students how to create maps; another involves the time it takes educators to score them (Shavelson et al., 1994). It also takes time to provide feedback (Schau et al., 2001). These latter reasons have arisen because of the way that Concept Maps are unique, drawing on individual student conceptualizations and prior knowledge structures; because of this, each map requires individual attention by educators.

Some researchers have attempted to alleviate these time concerns by pre-constructing maps and asking students to fill in missing concepts or relationships (Schau et al., 2001). Under this method, student answers became uniform, and instructors could score maps and provide feedback very quickly; results were so uniform that even computers could provide the necessary scoring and feedback (Chang et al., 2001; Derbentseva, Safayeni, & Cañas, 2007; Wu et al., 2012). While these did offer time savings, there were definite trade-offs, and the use of student constructed maps over pre-
determined maps marks one divide in the literature (Novak, 2010; Novak & Cañas, 2007). The gist of the argument has roots both in ontology and epistemology – what is knowledge, and how do we come to know? With traditional student created Concept Maps, what is known has been evidenced through graphical representations of what students understand. The answers are less clear with pre-constructed, fill-in-the-blanks maps. For these, what is known has been indicated through a one-word answer that has been derived to match a teacher-created conceptual structure. While non-integrated ways of knowing are present and valued in the current educational system, for example with skill-based learning and subject-specific courses, the trends that have favored these do not tend to align with constructivist ways of knowing.

An analysis of one representative study by Schau, Mattern, Zeilik, Teague, and Weber (2001) helps to illustrate these considerations. In this study, researchers attempted to determine whether pre-constructed maps could be valid representations of learning. To answer this, they gathered student scores on an objective test and student scores on what they called Select and Fill-in Concept Maps (SAFIs), then sought to determine whether a correlation existed between the two scores (Schau et al., 2001). Researchers explored two related studies within their article, but for the purposes of this discussion, only the first warrants analysis.

In this first study, researchers sought to examine the validity of Concept Maps as measures of student understanding in science (Schau et al., 2001). The rationale given for the study was to determine if Concept Maps could be appropriate measures of student connected understanding. The authors recognized the extensive use of Concept Maps in classrooms, especially in science, and they explained three major limitations of Concept
Map use for science assessments, two of which mirrored the ones previously discussed: maps were time consuming for students to learn, and there was no standardized method of scoring (Schau et al., 2001). To address these concerns, researchers chose to use SAFI Concept Maps, where a pre-configured map done by an expert (the teacher) was given to students with some missing concepts or linking words (Schau et al., 2001). In these maps, students were tasked with looking at the overall relationships and filling in the missing answers. Researchers then compared these scores with scores from an objective measure they created – a multiple choice test – to determine what type of relationship, if any, existed.

Researchers used multiple choice questions from released National Assessment of Educational Progress (NAEP) questions as their objective measure of general conceptual understanding (Schau et al., 2001). The sample consisted of seventh graders ($n = 236$) and eighth graders ($n = 395$) from five schools in New Mexico, and researchers had students fill in SAFI maps during the course of one of their science classes (Schau et al., 2001). Students also took the multiple choice test. Researchers then analyzed the correlation between the scores on the tests and on the SAFI maps and found strong correlations between them (seventh grade, $r = .74$; eighth grade, $r = .77$). This provided evidence that Concept Maps correlated strongly with objective measures of science achievement and as a result might be appropriate and reliable measures for science understanding (Schau et al., 2001). However, it might have been the case that these fill-in maps correlated strongly with the one type of science achievement tested here, which was knowing in the form of success with multiple choice test questions, but would not
correlate in the same manner with measures intended to elicit science understanding in other ways.

In addition, researchers noted that the SAFI format for Concept Maps seemed to alleviate the limitations of time and scoring associated with traditional student created Concept Maps (Schau et al., 2001). SAFIs differed from student created Concept Maps in other ways, as well, one of which was that students were presented with a pre-existing conceptual structure instead of being asked to represent what they knew. But instead of finding this difference a disadvantage, researchers addressed this in their overall discussion as somewhat of a strength; they claimed one benefit of SAFIs was that students were provided with explicit content (Schau et al., 2001).

The use of pre-structured versus student-made Concept Maps indicates an important divide in the research literature. Some researchers have considered SAFIs to be Concept Maps, but the pre-determined format of SAFIs has been seen by some to be at direct odds with the fundamental features of Concept Maps. Traditional Concept Maps are designed to involve students in active constructions of maps, and these maps are supposed to represent their individual conceptual understandings (Novak, 2010; Novak & Cañas, 2007). Maps done in this way are unique, and one beneficial element of this format is its inclusion of student prior knowledge (Novak, 2010). This allows for greater connections and more meaningful learning on the part of the student.

In contrast, SAFI maps present a completed map, representing one way of conceptually visualizing concepts. Instead of indicating this difference and noting the ways in which it was at odds with the fundamental purposes of Concept Maps, researchers assumed the structural difference made a minimal difference. In fact, the
researchers almost explained the pre-constructed feature of maps as beneficial, claiming that another advantage of SAFIs was that students were able to “see their visual representation of the domain as they complete[d] their maps” (Schau et al., 2001, p. 156). But students using this format did not see “their” representations; they merely filled in the blanks and completed associations that had been put together by someone else. They saw “a” representation, but it was not unique to them. Cañas, Novak, & Reiska (2012) recognized this issue and recommended against maps that over-scaffold, arguing that they provide too much structure and require rote knowledge to complete (Cañas et al., 2012). However, this recommendation has not prevented researchers from using SAFIs, and the research base for them has continued to grow.

In the end, the fundamental questions regarding how to measure knowledge remain, and researchers must choose how to proceed. Those who choose traditional student-constructed Concept Maps support the belief that student knowledge is best indicated through student constructions of what they know. Those who choose pre-determined maps support the belief that student knowledge can be indicated through objective measures. Filling in the blanks for a SAFI is akin to answering a question on an objective test, and there certainly is precedent for this type of measure. Both pathways offer research opportunities, and current studies have continued to expand both bases.

**How to measure learning.** One application of Concept Maps within the literature has involved their use as learning tools, but how best to assess their impact on learning reflects one divide in this line of research. One branch of research has analyzed the effects of Concept Maps on learning through the use of a separate but related measure. The other branch has analyzed this through an examination of Concept Maps themselves...
looking at how they change over time. Researchers have tended to name this impact various things – e.g., achievement, learning, conceptual understanding, knowledge – according to the research direction they have taken. No matter the approach, researchers have chosen multiple ways to measure the effects of Concept Maps, and readers of the literature should carefully assess the chosen methods, because within every study of this type there remains the question of exactly what construct is being measured.

**Objective measures.** Many researchers exploring Concept Map effects on student learning have used tests created from textbooks, course materials, or standardized tests to indicate growth in knowledge and understanding (Chiou, 2009; Doorn & O’Brien, 2007). A study examining the effects of Concept Maps on the learning of business and economics statistics students illustrates the use of this type of measure (Chiou, 2009).

This study involved 51 university sophomores in a business and statistics education course in Taiwan. Students were randomly assigned into three groups, one of which served as the control group. The researcher used a single factor, between subjects experimental design, with one group using collaboratively constructed Concept Maps (CCMING), one group using individual Concept Maps (ICMING), and one group serving as a control using traditional textbook exercises (TTE) (Chiou, 2009). There were 17 students in each group. Learning was assessed through an objective test created by the researcher. Chiou (2009) used the textbook’s question database to create a pre-test, which consisted of content from chapters one and two, and a post-test, which consisted of content from chapters three through seven. Students created and submitted initial Concept Maps in their groups during the third week of class. During this week they also took the pre-test measure, which provided quantitative scores. Students then worked on their maps
individually (ICMING), collaboratively (CCMING), or not at all, working instead on textbook problems (TTE) during class. At the end of the course, final Concept Maps were created and submitted. Then, students took a post-test, which yielded quantitative scores. Learning was reflected not by analysis of the Concept Maps, but by the change in test scores on the test (Chiou, 2009).

Descriptive statistics indicate pre-post score differences for each group: CCMING (58, 79), ICMING (57, 66), and TTE (56, 51). The researcher used these scores to determine the effect size of the Concept Map learning strategy: CCMING had an effect size of 1.22, ICMING had an effect size of 0.43, and TTE had an effect size of −0.28. The researcher concluded from these results that Concept Mapping had strong potential to be an effective learning tool. Given that regular classroom instruction using no educational innovations yields an average effect size of .40, the effect sizes for the Concept Mapping groups do seem educationally important (Hattie, 2012). This is especially true for the collaborative Concept Map group (CCMING), which had a very large effect size of 1.22.

A one-way analysis of covariance (ANCOVA) was run using pre-test scores as the covariate and post-test scores as the dependent variable. Results indicated that the main effect was significant ($F(2, 47) = 18.43, p < 0.01$); there was a statistically significant difference in post-test scores after controlling for pre-test scores (Chiou, 2009). Chiou (2009) also separated the post-test into multiple choice and calculation questions, and conducted one-way ANOVAs on these groups followed by Bonferroni post hoc comparisons. There were statistically significant differences between all groups. When it came to the multiple choice questions, CCMING scores were statistically significantly greater than those from ICMING, $p = 0.012$, and TTE, $p < 0.001$. Scores
from ICMING were statistically significantly greater than TTE scores, \( p = 0.039 \). Chiou’s (2009) analysis of the calculation questions indicated that CCMING and ICMING groups both outperformed the TTE group. The score difference between CCMING and TTE results were statistically significantly different, \( p < 0.001 \), as were the score differences between ICMING and TTE, \( p = 0.041 \). The differences between CCMING and ICMING were non significant. These positive effects of Concept Mapping in the classroom align with previous research on Concept Maps. The researcher concluded that Concept Maps had very strong potential to positively affect learning in statistics classrooms.

**Comparing maps across time.** Some researchers have asserted that accurate assessments of learning should be derived not from scores on objective tests, but from indicators of individual growth over time (Hay, 2007; Hay et al., 2008). Researchers supporting this line of research have believed that learning is shown through changes in understanding over time, and that Concept Maps offer one means of measuring this change (Hay et al., 2008). This approach has called for the comparison of Concept Maps done by students over time and has treated any changes in conceptual representation of ideas as learning (Hay, 2007; Hay et al., 2008). Novak (2010) asserted that changes in Concept Maps could signify learning and even show different levels of this learning. He distinguished between rote and meaningful learning, and suggested that each manifests differently in student work (Novak, 2010, p. 53). Concept Maps would be one means of making student learning visible.

A representative study by Hay et al. (2008) illustrates this research approach. The study involved 18 health care workers who were taking a course in mental health at the Institute of Psychiatry, King’s College London (Hay et al., 2008). The course took place
in 12 six-hour sessions over a period of six months, and the course content focused on dual diagnosis and issues of substance abuse and mental illness (Hay et al., 2008). The purpose was to measure learning using changes in Content Maps. The course began with a lesson and practice on creating Concept Maps, after which students created Content Maps to reflect their understanding of the content matter of the class. Students also took a short (15 question) assessment to measure prior learning. These two measures were repeated at the end of the course and changes between each were analyzed.

Researchers used a two-part method to analyze maps, a method created by Kinchin et al. (2000) and validated in a subsequent study by Hay (2007). The process involved a repeated evaluation of the structure of maps and then an analysis of the pattern of changes made; results were qualitative in nature (Hay, 2007). First, maps were classified by the complexity of their structure as either chain (least complex), spoke, or network (most complex). Second, pre-post Concept Maps were compared, and changes between two maps were analyzed for patterns of change.

The change or addition of various elements would move a map from one level to the next (Hay, 2007). For instance, additions of new concepts might indicate a student had moved from non-learning to rote learning; or, the addition of meaningful links between newly added concepts and existing parts of the prior concept structure might indicate a move from rote learning to meaningful learning (Hay, 2007). These changes were used as indicators of meaningful learning, surface learning, or non-learning (Hay, 2007; Kinchin, Hay, & Adams, 2000). The change of map structure, from less complex to more complex, would also signify learning. For example, an initial Concept Map with a simple chain structure compared against an end of course Concept Map with a network
chain structure would be an indicator of meaningful learning. The argument behind this method was that structural movement from simple chains to spokes and networks would reveal more complex conceptual understandings, and this change reflected learning (Hay, 2007).

Using this assessment system, researchers noted only 2 of 18 students created maps that researchers considered illustrations of meaningful learning. There were four students who showed non-learning through pre-post representations that were essentially unchanged, five cases that showed rote learning, and seven cases where the assessment system combined categories and labeled the learning “rote/meaningful” (Hay, 2007). This last category indicated maps where students had added meaningful new concepts and had made some strong connections between concepts but had also made superficial or no connections in places (Hay, 2007).

The study finished with in-depth case study analyses of six of the students and their pre-post class Concept Maps (Hay, 2007). Researchers discussed the value in seeing “sequential snapshots” of their students’ conceptual understandings, and noted that using these snapshots allows instructors to improve the quality of student learning (Hay, 2007, p. 234). In terms of limitations, researchers discussed the labels of learning – non, rote, and meaningful – as one possible limitation as these labels may not have revealed the true depths of student learning (Hay, 2007). One limitation they did not mention was the time-intensive nature of analyzing Concept Maps for learning in this manner. However, despite these limitations, this study is important to the literature in two ways: it provides an alternate way to measure learning through its comparison of pre-post Concept Maps, and it provides meaningful qualitative results. The case study analyses provide rich
Use of technology. The use of technology with Concept Maps has led Concept Map research in a number of new directions. The creation of Concept Maps using computer tools initially led to research comparing the effects of maps made by hand with maps made by computer. Some findings included that students have an easier time making corrections to their maps when they create them using the computer versus using pencil and paper, and that computers offer an efficient means of map creation (Chang et al., 2001; Novak & Cañas, 2007). The efficiency of computer systems quickly led researchers to investigate whether this efficiency could be applied to the areas of Concept Map feedback and scoring, and these lines of research (often combined in studies) continue to evolve (Chang et al., 2001; Cutrer et al., 2011; Joseph et al., 2017; Kordaki & Psomos, 2015; Wu et al., 2012).

A study done by Chang, Sung, and Chen (2001) offers an early version of feedback and correction support with computer generated Concept Maps. This system offered students two ways to create Concept Maps – a “construct by self” mode and a “construct on scaffold” mode (Chang et al., 2001). Students constructing on scaffold were presented a completely formed Concept Map that had some missing concept names and links. These scaffolded maps matched the format used by Schau et al. (2001) in what they termed SAFI maps. Students filled in answers and then checked their map for immediate feedback, or they filled in what they were able, and for concepts that were unknown to them, they clicked on hint buttons and received helpful hints. The hints were prompts, usually in the form of questions, such as, “Meiosis results in …?” (Chang et al.,
which, when completed, formed correct propositions for the map. Another scaffold opened up after students had spent more than 30 minutes on a map. This scaffold presented the completed expert map for students to study, and then allowed students to return to their own maps for completion (Chang et al., 2001). Finally, students could correct their maps by clicking on the “evaluation” button; in this process, the student map was compared against the expert computer map, and any deviations in concepts or links were marked (Chang et al., 2001).

The study involved 48 seventh grade students in a biology class in Taipei. Researchers randomly assigned students to one of three learning method groups: construct on scaffold, construct by self, and paper and pencil. They then measured the effects of these methods through a comparison of pre-post measures. Researchers used scores from a biology achievement test as post-test measures and compared this against student biology achievement scores from the previous semester, which they used as pre-test measures.

Researchers conducted an ANCOVA controlling for variance from pre-test scores and found post-test scores for groups were statistically significantly different ($F(2, 44) = 3.79, p < 0.05$). A post hoc comparison using the LSD method indicated that the scores from the construct on scaffold group were significantly better than those of the construct by self group ($p = 0.029$) and the paper and pencil group ($p = 0.016$). Findings also indicated that the difference between the construct by self and the paper and pencil groups was non significant (Chang et al., 2001).

Wu, Hwang, Milrad, Ke, and Huang (2012) further developed a computer Concept Mapping system that focused on feedback. They called their system “a concept
map approach with instant feedback mechanism” (Wu et al., 2012, p. 217). As the name suggests, this system provided students with instant feedback, with the understanding that students would use this feedback immediately to make changes to their maps (Wu et al., 2012). Students had a list of concepts, then were presented with a fill-in Concept Map to complete. When students finished, they submitted their map, and the computer scored it against an expert map, then provided both the score and feedback to students for corrections (Wu et al., 2012). Feedback consisted of indications that answers were incorrect, comments about the problem like, “There is a missing link for Concept A” (Wu et al., 2012, p. 220), links to supplementary material regarding the incorrect answer, and a score. As students made corrections, their scores increased and the amount of feedback decreased (Wu et al., 2012).

Researchers conducted their study on 81 final year nursing students in a clinical nursing course in Taiwan. The course focused on problem solving approaches in the field, and the specific content was “heart failure” (Wu et al., 2012). The study used a quasi-experimental design for non-equivalent groups. The experimental group \( n = 39 \) used the instant feedback computer mapping system; the control group \( n = 42 \) used a computer mapping system and relied on teacher feedback, which was provided “days or weeks” later (Wu et al., 2012, p. 225). Most important for this discussion was the measure of knowledge from the use of the Concept Map program. Researchers created a test from the course content being practiced on the Concept Maps for the post-test.

The pre-test measured general prior knowledge on heart failure, and the mean scores of each group, when compared, indicated no significant differences between group prior knowledge \( (t = .74, p > 0.05) \). Results from the post-test indicated that the
experimental group did significantly better than the control group \((t = 3.26, p < 0.05)\) (Wu et al., 2012).

This study is relevant for the way that it further developed earlier studies on Concept Mapping with computers and also for the way in which it illustrates a major limitation of this line of research. The limitation involves the issue of exactly how learning is being measured, an issue previously discussed in the section on “Student created maps versus pre-determined maps.” The measurement of learning in this study directly involved the content being practiced in the Concept Map task, which meant that students who did the mapping process again and again with the same map until they made a perfect score were simply practicing the memorization of given concepts and relationships (Wu et al., 2012).

In addition, this course was described as one that would address problem solving skills in the field, but this content was not integrated in this study; the content of the Concept Maps addressed the second course purpose which was to provide information about heart failure (Wu et al., 2012). This is another limitation found throughout research in this area, which is that it remains unclear whether the learning benefits from Concept Maps extend to transfer of knowledge or application (Nesbit & Adesope, 2006). The differences in group results from this study might have indicated learning, but the results seem to have been more representative of rote learning than meaningful learning as they did not involve the student and prior knowledge structures in any meaningful way. This was also a concern regarding the Chang et al. (2001) results.

Research has seemed to indicate that computer systems can efficiently deploy scores and types of feedback, but this efficiency has come at a cost. The computer system
interactions have been based on a universal map, and feedback and scoring have
developed from this map. As a result, some of the same issues of over-structure and rote
learning that were previously discussed with SAFI maps apply here as well. In the end,
these computer generated feedback and correction systems have depended on a primary
map and have required that all student maps match this structure and content.

**Additional research directions.** There are many other research directions
currently being pursued regarding Concept Maps. Many are only tangentially relevant to
this study, although some bear mentioning.

**Feedback.** The use of feedback with Concept Maps has not been well
documented in the literature and seems to be a research direction that is emerging more
robustly in the area of computer created Concept Maps than in the area of traditional
Concept Maps (Ifenthaler, 2011). While it seems to be understood that some sort of
direction is helpful for students once they finish their maps, this direction has most often
been interpreted as correction (Novak, 2010; Novak & Cañas, 2007). In computer created
Concept Map studies, feedback has consisted most often in the indicating of incorrect
responses accompanied by hints on how to correct or links to resources for further study
(Chang et al., 2001; Wu et al., 2012). Another avenue was illustrated by a study by
Joseph, Conradsson, Wikmar, and Rowe (2017), which used traditional Concept Maps to
test the application of a specific feedback theory (Joseph et al., 2017).

One variation on this last use of feedback and Concept Maps can be seen in a
study by Cutrer, Castro, Roy, and Turner (2011), which involved computer created
Concept Maps and explored the use of an expert map for feedback. This study involved
46 pediatric residents from Baylor College of Medicine. Participants were randomly
assigned to either the control group (which received a control lecture) or to the experimental group, which received an expert Concept Map to study for feedback. For the study, researchers asked all students to create an initial Concept Map to assess prior knowledge. Students then either received an expert Concept Map to use for study for five minutes or a short control lecture for the same amount of time. After this, all maps – student and expert – were collected, and students were asked to create new maps to indicate their conceptual understanding of the material. Changes between the two maps were assessed through quantitatively scoring each of the maps and analyzing pre-post score differences (Cutrer et al., 2011).

Score differences between groups from pre-intervention maps were non-significant, indicating that groups were entering into the study with equivalent prior knowledge ($M_{\text{experimental}} = 143.64, M_{\text{control}} = 140.30, p = 0.803$) (Cutrer et al., 2011). However, researchers found statistically significant differences when comparing pre-post scores; participants in the experimental group had significantly higher scores than those in the control group ($p = 0.001$) (Cutrer et al., 2011). Researchers concluded that while traditional lectures provided content to students, these lectures were not effective at providing a “big picture” for students and had not provided the structure that students needed to learn deeply. Concept Maps seemed to provide this conceptual structure, and when students revised their understandings using an expert’s organization for the content, they showed significant changes in their understanding; their maps reflected more complexity and greater conceptual connections (Cutrer et al., 2011). These results were promising, but more research is needed in the area of expert maps used in a consultative manner by students to build understanding. More research is needed regarding different
types of feedback as well, and all research should make an effort to clearly define the term “feedback” using parameters and/or examples. As noted in the introduction, this exploratory study hopes to expand the research in this area.

Changing map features. An emerging line of research has been focused on the elements of Concept Maps and the mechanisms by which these tools affect learning. Research in this area has focused on the effects of changing the nature of the focus question (Derbentseva et al., 2007; Miller & Cañas, 2008b) and on how the type of structure provided by a pre-determined map influences the nature of student linking terms (Derbentseva, Safayeni, & Cañas, 2004; Derbentseva, Safayeni, & Cañas, 2006; Derbentseva et al., 2007). Research also has explored the effects of various structural requirements on learning. This line of research has examined a continuum of map structures, including maps that have no boundary instructions and simply a focus question, maps that provide some structure such as a list of concepts to use, and maps that are pre-determined and restrict student participation to filling in missing items (Cañas et al., 2012).

Scoring. Novak (2010) established the first quantitative scoring system with the creation of Concept Maps in 1972, and since that time there have been many variations developed and tested within the literature. Most have involved deriving a quantitative score from the analysis of a map’s structure, content, or both. Some have been weighted by elements, for example giving greater numerical importance to crosslinks than concepts, and others have been weighted by the preciseness with which they match an expert map (Cline, Brewster, & Fell, 2010; Novak, 2010; Novak & Gowin, 1984; Rye & Rubba, 2002; Taricani & Clariana, 2006). Some also have focused on the qualitative
assessment of maps (Kinchin et al., 2000; Miller & Cañas, 2008a). There were too many methods to detail here, but the study by Anohina and Grundspenkis (2009) provides a comprehensive list of methods for examination.

**Concept Maps and statistics learning.** The research base has continued to expand for the use of Concept Maps in statistics education, and there are many possible research directions. Concept Maps seem to have great potential as learning supports in this field as they involve students in active learning, and they have helped both students and teachers identify and address misconceptions.

In one high school level summer statistics course, having students routinely map what they learned proved to be engaging, valuable for promoting discussions, and successful for tapping into student prior learning (Izumi, 2013). By the end of the course, student maps had changed in ways that reflected new understandings and conceptual relationships (Izumi, 2013). In another class, a university-level introductory statistics class, the instructor used Concept Maps as a wrap-up assignment, to help students see the “big picture” (Witmer, 2015). The instructor explained two purposes; first, the instructor wanted students to consider all of the ideas they had discussed and then make sense of them in their own ways and using their own connections; second, the instructor wanted to see student thinking (Witmer, 2015). From these student maps, the instructor was able to see the concepts that students thought were important, how students connected concepts, and how students made organizational sense out of all of the ideas they had learned (Witmer, 2015). Schau and Mattern (1997) described this as connected understanding. They argued that students must avoid random bits of isolated statistics knowledge and instead understand the many interrelationships between concepts in order to successfully
reason and problem solve (Schau & Mattern, 1997). Concept Maps can help students move toward connected understanding (Schau & Mattern, 1997).

Other studies such as Chiou’s (2009) have focused less on Concept Maps as learning tools and more on how they might reflect what “knowing statistics” should look like. This study made use of an objective statistics test, with both conceptual and computational questions, as a reference and then analyzed the ways in which Concept Maps aligned with this representation of “knowing statistics” (Chiou, 2009).

Finally, one well researched area of statistics education has involved statistics anxiety. It has been thoroughly documented in the literature that student affect plays an important role in student success in statistics learning (Macher, Paechter, Papousek, & Ruggeri, 2012; Onwuegbuzie, 2000). Research has begun exploring the ways in which Concept Maps might influence student affect, especially anxiety. One qualitative study used case studies of eight university students in an introductory statistics course to explore the effects of Concept Mapping on students’ affect and on student perceptions of the impact of Concept Mapping on their ability to relate to complex statistical concepts (Trehan, 2015). Narratives from this study indicated that students had mixed responses to Concept Maps as learning tools – for example, they appreciated the way these maps facilitated the integration of new material into their thinking, but remained uncertain about any benefits they may have regarding application of concepts to problem solving (Trehan, 2015).

There is still much to be done in the way of researching Concept Map use in statistics learning, but early research results have mirrored the benefits seen with Concept Maps in other disciplines.
Chapter Three

Research Methodology

This chapter comprises a description of the research methodology that was used in this study. It details the research design, validity and reliability considerations, the participants and sampling process, instrumentation, procedures, and statistical analyses.

Research Design

This study used a convergent mixed methods research design to attempt to answer three research questions:

Question 1: Is there a significant difference between the scores for Concept Maps made after students receive Expert Map Feedback and the scores for Concept Maps made after students receive Peer Feedback?

Question 2: Do student perceptions of the usefulness of Concept Maps for statistics learning and of the value of feedback they experienced differ based on the type of feedback they experienced - Expert Map Feedback versus Peer Feedback?

Question 3: Are there significant differences between groups when structural classification and student choice of starting terms is considered?

The researcher collected both quantitative and qualitative data in one time period, analyzed the data separately, then merged relevant portions of the data with the goal of using all findings to understand the results and practical implications of the research more fully (Creswell & Plano Clark, 2018). This design allowed for an enriched exploration of the relationship between feedback and student learning through its inclusion of both types of data as well as the merging of relevant parts of each data set; the different types of data
both informed and expanded on one another. The process used in this study is illustrated in Figure 2.

![Figure 2. Convergent mixed methods - Illustration of process. The data sources and merging process for this mixed methods research study.](image)

This design supported one pragmatic, underlying purpose, which was to explore more practical means of using Concept Maps in the classroom. While quantitative scores and analyses of differences provide information on understanding as expressed through Concept Maps, they do not address student perceptions of this learning tool. This additional perspective was valuable in light of the underlying practical nature of the study and the fact that students, in the end, would be the desired users of this learning support.

**Internal and External Validity**

The researcher sought to strengthen the power of the study by controlling for certain variables and factors that might affect the internal and external validity of the study. Attempting to control for factors that might affect validity is always important for
the strength of a study, especially for studies like this that use multiple means of collecting data and lack random selection (Gall, Gall, & Borg, 2007).

Probably the greatest threat to internal validity in this study regarded Experimental Treatment Diffusion (Gall et al., 2007). This involves the sharing of one group resource with another group during a study. As a result, the power of the resource being measured in one group is diffused (Gall et al., 2007). The researcher understood that the Expert Map (see Appendix B) might be considered a desirable resource by members of the groups that did not have access to it. To control for the sharing of this resource, the study ran on the same day, group members were asked not to talk outside of their groups for the duration of the study, and the groups that received the Expert Maps received them in an unobtrusive fashion in an attempt to draw little attention to the additional resource. In addition, the name of “Expert Map” was never used. Instead, the Expert Maps were introduced using versions of the following statement: “Here is another example that someone did if you want to look at it as you talk.”

The extraneous value of history was also relevant for this study (Campbell & Stanley, 1963). This variable involves the passage of time and the influence of other elements (e.g., additional learning, attitudinal shifts, and learning environment changes) (Gall et al., 2007). Since Concept Maps concern learning and conceptual relationships, the passage of time and the inclusion of further reading or statistics work had the potential to change conceptual understanding and influence the outcomes of the study. Because this study focused on the role of feedback on map creation and structure, the researcher sought to limit the confound of time and other history elements through repeated mapping during the same course period.
In terms of external validity, the researcher attempted to control for the Hawthorne effect – that is, the situation of having participants change their behavior because of the knowledge that they are being watched – by having the course teacher co-implement the study and by including activities within the regular classroom learning experience (Gall et al., 2007). The researcher limited any novelty and disruption effects by running the study during a regularly scheduled class, which gave the appearance that it was a standard class exercise.

Finally, the researcher tried to control for external validity issues due to the measurement of the dependent variable – an effect where the measurement of the dependent variable is overly influenced by the treatment (Gall et al., 2007). The researcher attempted to control for this in three ways: through feedback instructions, through restricting physical revisions to maps, and through changing the Concept Mapping topic and concepts on the second map. First, the study included clear instructions regarding how to give feedback, instructions that required students to ask questions of one another regarding why a map element was constructed the way it was. The focus was on discussing ideas rather than on revision of structure. These discourse supports were in the form of question stems. In this way, the researcher hoped to encourage meaningful discussion by all with a goal of having students compare and explore the relational differences expressed by different maps. Second, students were asked to make their initial maps using colored pens, and these were switched with pens of a different color after the initial creation of maps. In this way, students could indicate possible changes to their maps in a way that would leave their initial version intact. This was designed to focus the group time on discussions about concepts and
interrelationships rather than on copying or revision, but to allow students to make notes on their maps if desired. Finally, the topic of the second map was different from the first but related conceptually. This prevented students from simply memorizing the structure of the Expert Map or of a peer map and replicating this structure on their second map.

**Participants and Sampling**

This study took place in a private university in a major city in the western United States. A convenience sample was used, consisting of students in a 10-week, undergraduate-level, introductory statistics course that met during the fall quarter. Students enrolled in the course as part of their required university course load, and because of this, they did not constitute a randomly selected group for the study.

The course consisted of 120-minute class meetings, held in person, two times per week for each of the ten weeks. Data were gathered during the first hour of one regularly scheduled class period. The course provided accepted math credits for many majors, and students could enroll after having passed the university mathematics placement exam or after having taken a 100-level college math readiness course. This particular class was specifically focused on statistical applications in the sciences. All students in the class ($N = 28$) participated in this study.

The demographic breakdown of the sample is provided in Table 1. The course was designed for students interested in science. Those majoring in Nursing comprised the greatest group (32.1%), followed by Biology majors (14.3%). The class was composed primarily of Juniors (42.9%) and Sophomores (35.7%).
Table 1

Demographic Information of Sample

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Frequency</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic or Latino or Spanish Origin</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>Asian</td>
<td>4</td>
<td>14.3%</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>White</td>
<td>19</td>
<td>67.9%</td>
</tr>
<tr>
<td>Two or more races</td>
<td>3</td>
<td>10.7%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gender</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>19</td>
<td>67.9%</td>
</tr>
<tr>
<td>Male</td>
<td>9</td>
<td>32.1%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Grade Level</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Freshman</td>
<td>2</td>
<td>7.1%</td>
</tr>
<tr>
<td>Sophomore</td>
<td>10</td>
<td>35.7%</td>
</tr>
<tr>
<td>Junior</td>
<td>12</td>
<td>42.9%</td>
</tr>
<tr>
<td>Senior</td>
<td>3</td>
<td>10.7%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Declared Major</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Biochemistry</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>Biology</td>
<td>4</td>
<td>14.3%</td>
</tr>
<tr>
<td>Computer Science</td>
<td>3</td>
<td>10.7%</td>
</tr>
<tr>
<td>Dietetics</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>Exercise Science</td>
<td>2</td>
<td>7.1%</td>
</tr>
<tr>
<td>Health and Fitness</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>Physics</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>Mathematics</td>
<td>2</td>
<td>7.1%</td>
</tr>
<tr>
<td>Music</td>
<td>1</td>
<td>3.6%</td>
</tr>
<tr>
<td>Nursing</td>
<td>9</td>
<td>32.1%</td>
</tr>
<tr>
<td>Physiology</td>
<td>2</td>
<td>7.1%</td>
</tr>
<tr>
<td>Other</td>
<td>1</td>
<td>3.6%</td>
</tr>
</tbody>
</table>

Groups. The researcher received a class list, assigned every student a randomly generated number, then randomly assigned students into one of two feedback groups, referred to throughout as Peer and Expert Map. Both groups consisted of 14 members,
broken into smaller groups. Each feedback group had four groups of three students and one group of two.

All methods were submitted to the Institutional Review Board of the university and approved.

**Instrumentation**

**Expert Map.** The researcher worked with the instructor to create the Expert Map given during feedback to ensure that the content matched what was being learned in class and aligned with the purposes of the study. All terms were chosen from the same chapter of the student textbook. Although the term “Expert” has been widely used to convey specialized knowledge, in this context this was not the intended meaning. Rather, in this study the term “Expert” was used in conjunction with the feedback map, and the Expert Map was intended solely to represent one clear, organized, and sound representation of the material. And, even though the Expert Map was intended to illustrate another possible way to organize concepts and conceptual relationships, it was not intended to be copied or to serve as a “final” version or the one “correct” version. Additionally, students never heard the term “Expert Map.” The final Expert Map consisted of 10 concepts, 18 propositions, and eight cross-links (see Appendix B for Expert Concept Map). The 10 concepts consisted of five general concepts (e.g., “bias”) and five specific concepts (e.g., “do not call list”), and the map organization moved from more general concepts off the starting concept to more specific concepts. The propositions consisted of phrases (concept - linking words - concept) that were designed to indicate relationships (e.g., Convenience sample might result in undercoverage.) The cross-links were intended to show the interrelated nature of some of the more general concepts. For example, the
concept “bias” connected to five different concepts, as opposed to the more specific term “do not call list” which connected only to two. When scored using the scoring system, the Expert Map earned a score of 210 points.

**Questionnaire.** The researcher created and used a 12-item questionnaire (see Appendix A for Questionnaire). While there were existing student perception surveys that had documented high reliability and validity in the area of statistics education and in the area of statistics anxiety – for example, the Survey of Students’ Attitudes Toward Statistics-36 [SATS-36] (Cruise & Wilkins, 1980) and the Statistics Anxiety Rating Scale [STARS] (Schau, 2003) – the researcher found that the constructs measured in these surveys did not align with the current study’s purposes. The goal of the survey in this study was to elicit student perceptions on the value of feedback and the usefulness of Concept Maps given specific methods of feedback. The results were not meant to generalize to a larger population; instead, the results were designed to inform the quantitative data generated by the first two research questions.

A search of the literature revealed questionnaires that asked about student perceptions with Concept Maps, but none that specifically asked the questions of interest in this study connecting feedback methods and statistics learning. While a researcher-created scale would not have undergone the reliability and validity checks of a measure that has been in use, it did have other benefits. Most notably, a scale that addressed the specific feedback questions linked with statistics learning would provide more valuable context when analyzing the results of the Concept Maps. The goal of the measure was to capture the underlying perceptions of students regarding feedback, Concept Maps, and
statistics learning, and the hope was that the more focused questionnaire would capture these concepts better than a generalized survey on anxiety or statistics learning.

To check for readability, two independent parties reviewed the survey. One reviewer was a teacher at the secondary level, and one reviewer was a recent college graduate. Neither was associated with the university or students involved in the study. Both checked the language used and read questions out loud. The reviewers were asked to think aloud while reading each question and to respond to the question, “What does this mean to you?” The questionnaire language for all questions was kept simple, and all terms were designed to be familiar and easily understood by all class members. The researcher eliminated vague terms and constructed all sentences so that they were in active voice. Care was taken to avoid double-barreled questions so that each question attempted to measure only one concept.

Additionally, the questionnaire was kept short, to guard against survey fatigue, and question order was intentionally chosen to ensure as much as possible that answers to previous questions did not impact future answers. For this reason too, demographic questions were placed at the end of the questionnaire.

The researcher chose a Likert scale for responses, and selected the language and number scale after carefully considering the benefits and limitations of different options. Traditional Likert scales consist of response choices that have been organized on a continuum. These choices can range from “Strongly disagree” to “Strongly agree,” or they can consist of other language appropriately chosen to fit the questions, language such as “Never” to “Always” (reflecting frequency) or “Not at all” to “A lot” (reflecting quantity). The number of choices also varies, with some scales offering upwards of seven
choices and others as few as two. Traditional scales offer five choices (Nadler, Weston, & Voyles, 2015). This questionnaire consisted of Likert scale responses that ranged from “Strongly disagree” to “Strongly agree,” and offered five scale options from which to choose.

There are differing views about whether Likert surveys should offer four choices or five response choices, an odd number or an even number. Research has been mixed regarding one “best” option, and in the end, the consensus has seemed to be to choose the format that best fits the given questions and that takes a clear look at the benefits and limitations of the number chosen.

Five point response scales differ from four point response scales in that they offer a center value, one that is suggested to be neutral on the scale. This value had benefit to the researcher in that it made a larger set of scale choices and as such increased the variability of the scale. Offering a larger number of responses (all of which are labeled) also has been shown to increase the reliability of the scale (Weng, 2004). For the survey taker, it also provides an answer for those who might genuinely not have an opinion regarding a question or who have perceived no effect or change. This type of response fits the labels often given to this middle choice – “Neutral” or “Neither agree or disagree” (Nadler et al., 2015).

However, proponents of a four-point scale often are critical of this middle, fifth option. They claim that even though the title of the choice says “Neutral” or “Neither agree or disagree,” survey takers often use this option to reflect a variety of other responses, ranging from being unsure to not caring. Critics argue that this option does not
have a consistent, clear meaning to all and because of this, the meaning of the response cannot accurately be understood (Nadler et al., 2015).

Critics also suggest that survey takers might also use this response in instances where further careful thought would be necessary to discern a true preference. Because of this it might allow respondents to answer quickly without the careful consideration necessary for more subtle questions (Adelson & McCoach, 2010).

In addition, critics of the five-point scale argue that a center response allows survey takers to have an “out” from expressing their opinion. This desire to not answer the question might especially affect topics that are sensitive.

However, despite these criticisms, there are concerns regarding offering a four point scale. These scales are often referred to as “forced choice” scales, and while this might seem to be a benefit in terms of simplifying or clarifying answers in an analysis, forcing a respondent to choose might result in answers that are not truly reflective of what a respondent thinks. So, despite the seeming clarity of a decided choice, the answers might not accurately reflect opinions. In addition, when faced with answer options that won’t accurately reflect their ideas, some respondents might become frustrated (Nadler et al., 2015). Research also seemed to indicate that fewer response items in a scale results in lower scale reliability (Weng, 2004).

In the end, the goal was to best match the type of questions and number of response options to the survey and the audience so that the most representative information could be gained. For this questionnaire, a five-point response option was provided to ensure that students had the opportunity to express the full range of their thinking. This was an exploratory study, and because of this, the researcher did not want
to force students into expressing an effect or an opinion when none really existed. A five-
point response option not only increased the variability of the scale, but it provided a full
set of responses for further study.

Survey questions consisted of 10 Likert scale responses and two short-answer
responses. The Likert scale responses were organized in a five-point scale (1 = Strongly
disagree, 2 = Disagree, 3 = Neither agree nor disagree, 4 = Agree, 5 = Strongly agree).
The researcher structured some questions generally based on questionnaire themes that
emerged from prior research surrounding Concept Map use in the classroom, but no
questions were taken verbatim (Cutrer et al., 2011; Harrison & Gibbons, 2013; Witmer,
2015; Wu et al., 2012). This was done to provide a possible means of connection between
this study and prior studies using Concept Maps for learning with feedback and to further
situate study results in the existing literature. Students were asked to complete all 12
questions. The completion rate for questions 1 through 10 was 100% and was similarly
high for the open-response Questions 11 (100%) and 12 (96%). The Cronbach’s alpha
was 0.797, which indicated a high level of internal consistency for the questionnaire with
this sample of students. Generally, a Cronbach’s alpha of 0.700 or greater is desired (Gall
et al., 2007).

**Procedures**

All activities occurred during existing statistics course time as part of expected
coursework. Before the intervention, students were asked to watch an introductory, nine-
minute narrated PowerPoint on Concept Maps which detailed what they are, how to
construct them, and tips for use in learning. The video included an example of a map
being made, and at the end, students were asked to practice by making a Concept Map on
one topic related to their reading homework. Analytics from the video indicated that 15 of the 28 students (54%) accessed the video, but the actual number of students who viewed the video may have been different as students reported working together on their homework and some students may have had it playing but might not have listened.

On the day of the intervention, students came to class having read the assigned course materials and finished their textbook homework problems. The researcher and teacher introduced the exercise of Concept Mapping, and the researcher gave a short review of the elements that comprise a Concept Map. The language for this review and the example provided were taken from the PowerPoint video.

The teacher and researcher then distributed the first Concept Map sheet, which included the Focus Question and the Concepts to be used, and distributed a special color blue pen for students to use in their work (see Appendix C for Concept Map 1). Students had 12 minutes to create this first map.

The pens were collected and a new color pen was provided for students to use. All students were given brief instructions to guide group work, which described what feedback should look like in this group activity. A short list of discourse prompts and sentence starters was provided for use during group work, and the researcher modeled the use of these prompts (see Appendix D for discourse supports). Group work expectations were explained before students met in groups. They were asked to keep their discussions within their group only and to work without teacher help or clarification. In addition, students were instructed that they could make adjustments or changes to their maps using the different color pen, but that the focus should be on the discussion of the concepts rather than on revision. Then, the small groupings were shown on the document camera,
and students were instructed to meet with their groups and discuss their maps. At this time, the researcher casually gave each of the five groups the Expert Map accompanied by the statement, “Here is another example that someone did if you want to look at it as you talk.”

The feedback exercise ran for eight minutes. Students talked in their groups about the maps they had made and provided feedback to one another.

After eight minutes, initial student maps and Expert Maps were collected. Students were then asked to construct a second Concept Map (see Appendix E for Concept Map 2). Students had 12 minutes to create this new map, and then all maps were collected.

Finally, the questionnaire was given to determine student perceptions of Concept Map creation, feedback, and usefulness for learning and conceptual organization (see Appendix A for Questionnaire).

Statistical Analyses

Sources of data. The Concept Maps provided three sets of data – one set of continuous data and two sets of categorical data. First, the Concept Maps were scored, and these continuous scores provided quantitative data that was analyzed using IBM SPSS Statistics 25. The Concept Maps also were given one of two structural classifications – Discrete or Integrated. Maps were classified as Discrete in structure if they contained three or fewer cross-links, and they were classified as Integrated in structure if they contained more than three cross-links. Finally, Concept Maps were analyzed for the terms that were chosen to start the maps. All terms that directly connected to the starting concept were tallied and recorded as either General or Specific.
This categorical descriptor came from the list of available concepts. This list had been structured so that half of the available concepts were general and half were specific.

Additional data came from the Questionnaire. The questionnaire consisted of 10 statements accompanied by a five-point Likert scale for responses, which ranged from “Strongly agree” to “Strongly disagree.” It also consisted of two open-ended questions (see Appendix A for Student Questionnaire). The Likert scale responses provided data that were analyzed for frequency. The open-ended responses were analyzed for emergent themes, and these themes were coded (Creswell & Plano Clark, 2018). Question 11 themes and codes were: 1- Elements, 2-Relationships, and 3-Big Ideas. Question 12 themes and codes were: 1-Presentation, 2-Structure, 3-Relationships, and 4-Big Ideas. The codes for both questions were arranged in an order that proceeded from specific to general.

First, the data were analyzed and/or organized separately, and then in accordance with the mixed methods design, the researcher merged different groups of data in an effort to corroborate findings or glean further insights (Creswell & Plano Clark, 2018). For example, the Discrete/Integrated structural classifications were merged with the questionnaire responses to explore whether the responses on the questionnaire might differ based on the structural sophistication of the Concept Maps.

Quantitative analysis. For the quantitative analyses of the Concept Map scores, the significance level (α) was set at 0.05 for all hypotheses tests, a level that is generally used in educational research and appropriate to avoid increased risk of Type I or Type II errors (Gall et al., 2007). An a priori power analysis using online power analyzer G*Power 3 (Faul, Erdfelder, Lang, & Buchner, 2007) established that a study using a
repeated measures ANOVA and given an alpha level of 0.05, a desired power of 0.80 (80%), and desired medium effect size, would require 82 participants. The sample size of 28 was not adequate for answering both research questions within the power and effect size boundaries desired. However, the study was constrained by the reality of an available class, and to help offset the loss of power given the small sample size, the design was adjusted to include information from other analyses with the hope of more fully understanding the results despite a lack of power.

In addition to statistical significance of results, the researcher determined practical significance of results through the use of the partial eta squared ($\eta_p^2$) effect size measure (Field, 2013). Effect size measures provide a standardized way to measure the magnitude of an effect, and they give important information regarding the practical importance of findings (Field, 2013). In addition, the American Psychological Association (APA) recommends their inclusion in research results (American Psychological Association [APA], 2009). The researcher used the following, well established guidelines for interpreting partial eta squared values: $\eta_p^2 = .01$ (small effect), $\eta_p^2 = .06$ (medium effect), and $\eta_p^2 = .14$ (large effect) (Field, 2013).

**Analysis of Question One.** A mixed repeated measures ANOVA was used to analyze the data from this study to answer Research Question One. The research goal for this question was to determine if there were statistically significant differences in Concept Map scores based on group membership.

In this design, there were two factors or independent variables – Feedback Group and Time. There were two levels to the variable “Feedback Group” that were named “Peer” and “Expert Map.” The variable “Time” also had two levels, 1 and 2. In this
design, the Within Groups variable was Time and the Between Groups variable was Feedback Group. The mixed repeated measures ANOVA offered results for the main effects and the interaction effect between Feedback Group and Time (FeedbackGroup*Time).

The dependent variable was Concept Map score as measured using a variation of the traditional scoring method detailed by Novak & Gowin (1984). This scoring method has been widely used in the Concept Map research and has been analyzed for validity and internal consistency (Watson et al., 2016). This method was chosen over others for its ability to yield consistent scores between Concept Maps, for its time efficiency while scoring (one to three minutes per map), and because it required only one scorer due to the fact that it relied only on straightforward counts of map features, without interpretation (Watson et al., 2016) (see Figure 3 for scoring system).

\[
\text{Concept Map score} = (#C - #CL) + (HL \times 5) + (#CL \times 10)
\]

where
- \#C = number of Concepts used
- \#CL = number of Cross-Links
- HL = Greatest Hierarchy Level

*Figure 3.* The scoring system created by Novak (1984). This system is widely used and has many variations in the literature. The Concept Map scoring system used in this study weighted the terms equally at x10 and included an additional term of “Branches.”

The data were analyzed to ensure that the conditions necessary for repeated measures ANOVA were met. This involved checks that the data were continuous, each level of each variable was (relatively) normally distributed, and that there were no outliers. Outliers were checked using boxplots, while normality was checked using histograms and P-P plots. Skewness and kurtosis was analyzed through histograms and
output statistics. Additionally, normality was checked using the Shapiro-Wilk test (Field, 2013). The data generally met the required conditions, with a small variation in normality for one level of one variable, but the variation minimally affected the data and the decision was made to leave the initial data intact. Sphericity concerns did not apply to this design. Levene’s test was used to check homogeneity of variance.

**Analysis of Question Two.** The researcher analyzed data for the second research question using two methods since exploring this question required both quantitative and qualitative data analysis. The second research question explored whether student perceptions of the usefulness of Concept Maps for statistics learning and of the value of feedback they experienced differed based on the type of feedback they experienced. To answer this, the researcher analyzed responses for the first 10 questions on the questionnaire using frequencies for each Likert scale response. The responses were disaggregated by group, and tables and histograms were examined for patterns.

Additionally, the two open-ended questions (Questions 11 and 12) were analyzed for the emergence of themes. Three themes emerged for Question 11 and four themes emerged for Question 12. These were given coded values of one to three and one to four, respectively. The codes for both questions were arranged in an order that proceeded from specific to general, and were meant to be categorical references only. There was no scale associated with the chosen numbers. The coded data were then merged with the quantitative data from the Concept Map scoring and analyzed for patterns. Combining data in this way was meant to further inform the quantitative results that were obtained through scoring the Concept Maps (Creswell & Plano Clark, 2018).
**Analysis of Question Three.** Research Question Three explored the differences (if any) between groups when structural classification and student choice of starting terms was considered. For structural classification, all Concept Maps were analyzed based on overall structure and given one of two structural categorizations – Discrete or Integrated. The map structures were designated “Discrete” if they consisted of discrete branches off the starting concept, if the branches were connected by few or no crosslinks, and if there were few or no crosslinks between concepts within a branch (#Crosslinks ≤ 3). Concept Map structures were classified as “Integrated” if they consisted of discrete branches off the starting concept but the branches were connected by crosslinks and there were crosslinks between concepts within a branch (#Crosslinks > 3). These two structural classifications arose after consideration of the work by Kinchin et al. (2000), which classified structures as Chain, Spoke, or Network. These three classifications increased in complexity and involved considerations similar to those described for Discrete and Integrated. The use of the Chain structure was excluded as it was overly simplistic for the Concept Maps from this study.

Research Question Three also analyzed Concept Maps based on the concepts that students chose to connect to the starting concept. While not an idea represented in the literature regarding structural analyses of maps, this researcher suggests it can fit in a structural analysis because a student’s first concept choice very strongly drives structural considerations (Novak, 2010). When designing the selection of ten concepts to be used in each of the maps, the researcher created concept banks that consisted of five general terms (terms such as “bias” and “inference”) and five terms that were specific and could serve as examples (terms such as “wording effects” and “random digit dialing”). The
organization of Concept Maps typically radiates away from the center concept in a move from general to specific (Novak, 2010). A map that continued from the given concept using general words first would be considered to have a stronger conceptual presentation than a map that began with an example, then used a general concept at the end of a branch, a placement usually reserved for the most specific terms (Novak, 2010).

The researcher analyzed only those concepts that directly attached to the center concept and recorded how often each of the general words was chosen and how often each of the specific words was chosen. Responses were recorded by feedback group.

Finally, the structural classifications were merged with questionnaire responses with the purpose of illuminating more patterns within the data and/or corroborating findings from separate analyses.

Chapter Four presents an in-depth account of the results from and analyses of these three research questions. This includes a full presentation of descriptive and inferential statistics in both table and narrative form, as well as a discussion of the ways in which the data met required assumptions for statistical tests.
Chapter 4

Results

Chapter Overview

This chapter presents the results and the analysis of results from the Concept Map and feedback study. The researcher sought answers to three research questions:

Question 1: Is there a significant difference between the scores for Concept Maps made after students receive Expert Map Feedback and the scores for Concept Maps made after students receive Peer Feedback?

Question 2: Do student perceptions of the usefulness of Concept Maps for statistics learning and of the value of feedback they experienced differ based on the type of feedback they experienced - Expert Map Feedback versus Peer Feedback?

Question 3: Are there significant differences between groups when structural classification and student choice of starting terms is considered?

The results of the study derived from two sources – two separate Concept Maps created by students and answers to a short questionnaire. The Concept Maps were given quantitative scores and structural classifications, and were analyzed for starting concepts. The results from a short questionnaire provided to students after their feedback and mapping experience provided quantitative data in the form of responses on a 5-point Likert scale and qualitative data from two open-ended questions. All data were explored and analyzed separately, and then qualitative data from the open-ended questions were merged with the Concept Map scores and classifications for greater depth of understanding (Creswell & Plano Clark, 2018).
Research Question One Results

With the first research question, the researcher sought to determine whether a statistically significant difference existed between the Concept Map scores of those who received Peer Feedback and those who received Expert Feedback. This question was explored using quantitative data in the form of scores from two Concept Maps. The first map was made prior to feedback and the second map was made after one of two types of feedback (Peer or Expert Map). The scoring system used to quantitatively score the Concept Maps was a version of the one created by Novak (1984). This scoring approach has been widely used in the Concept Map research, has been analyzed for validity and internal consistency, and is time efficient and requires only one scorer (Watson et al., 2016). It is a system that awards scores based on the presence or absence of structural characteristics, with higher scores indicative of stronger conceptual representations. For this project and topic, scores on the first set of maps ranged from 140 to 230, and scores from the second set of maps ranged from 110 to 240. The Expert Map given to the Expert Map feedback groups received a score of 210 points when scored in this way. The basis for the scoring formula is provided in Figure 3 (see Chapter 3 on Research Methodology).

Both Concept Maps were made by students on the same day, during a regular statistics class. The focus questions and primary content were different for both maps, but they addressed statistical concepts from the same chapter in student textbooks. To analyze these data, the researcher used a mixed repeated measures ANOVA and considered significance of effects at an alpha level of 0.05 (Gall et al., 2007).
**Descriptive statistics.** The data consisted of scores from all students present in the class \((N = 28)\). There were no missing scores. Students worked independently for the first Concept Map. After completion, students met in feedback groups. Each Feedback Group type had four groups of three students and one group of two students \((n_{\text{peer}} = 14; n_{\text{expert}} = 14)\). Students had been randomly assigned to these groups. Half of these groups received an Expert Map to supplement their discussions, and all students received discourse supports. The focus was on discussion of this first map, not on revision. When the group discussion time ended, students were then asked to individually create a second Concept Map. The presence or absence of the Expert Map during feedback determined the two levels of the independent variable “Feedback Group.” The levels are referred to here as Peer and Expert Map. Table 2 provides initial descriptive statistics for the pre- and post-feedback maps of all students before consideration of feedback group.

Table 2

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SEM</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMap 1</td>
<td>28</td>
<td>183.21</td>
<td>4.426</td>
<td>23.42</td>
<td>.047</td>
<td>-.887</td>
</tr>
<tr>
<td>CMap 2</td>
<td>28</td>
<td>195.71</td>
<td>5.454</td>
<td>28.86</td>
<td>-.861</td>
<td>1.27</td>
</tr>
</tbody>
</table>

The mean scores for the class for Concept Maps 1 and 2 were different, with initial mean scores \((M = 183.21, SD = 23.42)\) slightly lower than post-feedback mean scores \((M = 195.71, SD = 28.86)\). At first glance, the mean difference of 12.5 points between Concept Map scores when grouped in this initial manner suggested a modest improvement in scores as a class between the first and second mapping attempts.
The descriptive statistics also provided skewness and kurtosis values for these variables. For a distribution to be similar to a normal distribution, the levels of skewness and kurtosis should be between 0 and 1, with values closest to 0 indicating greater normality (Field, 2013). In these samples, skewness values were less than a value of one, indicating that the data for both had relative symmetry about the center point and approximated the symmetry of a normal distribution (Field, 2013). The data for both Concept Maps 1 and 2 had high kurtosis values. While the kurtosis value for Concept Map 2 exceeded the suggested value of one, the histograms and Q-Q plots did not show severe differences from normal. It is important to note that these statistics can be affected by small sample sizes (Field, 2013).

Additional information regarding the shape and spread of the data was available once the data were disaggregated by Feedback group. Table 3 provides these results.

Table 3

*Descriptive Statistics for Concept Maps 1 and 2 by Feedback Group*

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>Mean</th>
<th>SEM</th>
<th>SD</th>
<th>Skewness</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>CMap 1 (peer)*</td>
<td>14</td>
<td>178.57</td>
<td>7.02</td>
<td>26.27</td>
<td>.151</td>
<td>-1.47</td>
</tr>
<tr>
<td>(expert)*</td>
<td>14</td>
<td>187.86</td>
<td>5.36</td>
<td>20.07</td>
<td>.409</td>
<td>-.143</td>
</tr>
<tr>
<td>CMap 2 peer</td>
<td>14</td>
<td>185.71</td>
<td>8.88</td>
<td>33.22</td>
<td>-.413</td>
<td>.995</td>
</tr>
<tr>
<td>expert</td>
<td>14</td>
<td>205.71</td>
<td>5.42</td>
<td>20.27</td>
<td>-.864</td>
<td>.355</td>
</tr>
</tbody>
</table>

*These are the groups students later would be assigned to. Groups were announced and met after the completion of Concept Map 1.

Kurtosis values for the initial map were high for the Peer group, with values exceeding the recommended threshold of plus or minus one for Concept Map 1.
(kurtosis<sub>peer</sub> = -1.47) and nearing it for Concept Map 2 (kurtosis<sub>peer</sub> = .995). The skew values remained at levels less than one. While these kurtosis values provided information about the shape of the data – the high negative values indicated a distribution that was heavy in the tails – it is important to note again that the values for both skewness and kurtosis can be greatly affected by sample size, with greater differences found in small samples (Field, 2013). Given that there were also no indications of outliers when boxplots were checked, that plots revealed no severe deviations from normality, and given the small sample size, the decision was made to leave the data untransformed. The differences between groups on the first set of Concept Map scores are illustrated in Figure 4.

![Histograms of Concept Map 1 scores by Feedback Group. These histograms indicate scores on initial Concept Maps, before meeting in one of two feedback groups.](image)

From Concept Map 1 to Concept Map 2 in the Peer Feedback group, the mean score changed from $M = 178.57$ to $M = 185.71$ points, a difference of 7.14 points. In the Expert Map Feedback group, the mean score changed from $M = 187.86$ to $M = 205.71$
points, a difference of 17.85 points. These differences indicated that while both groups improved the structure of their maps and consequently increased their group’s average map scores, the students in the Expert Map group seemed to experience a greater overall increase in scores. The standard deviations between groups were relatively similar for the first Concept Maps ($SD_{peer} = 26.27$, $SD_{expert} = 20.07$), but for the second Concept Maps, there was greater variability in the peer group scores ($SD_{peer} = 33.22$, $SD_{expert} = 20.27$). This difference was visually noted in a comparison of the boxplots as well.

**Inferential statistics.** The choice of a mixed repeated-measures ANOVA for statistical analysis of the data involves similar condition checks as for an ANOVA, with the exception of the condition of independence of observations. This condition is violated in repeated measures because the data to be analyzed originates from the same individuals. The standard condition checks are discussed first, and all statistical checks are confirmed with the use of visual data when possible.

Boxplots and histograms when analyzed by Feedback Group revealed no outliers. The assumption of normality applies to the residuals of the data, and an analysis of the standardized residuals revealed patterns similar to those discussed in the prior section. Histograms, Q-Q Plots, and statistical tests of residuals all indicated characteristics that approximated a normal distribution albeit with kurtosis in one group as discussed previously. The Shapiro Wilk test of normality was appropriate given the small sample size. The results of this analysis are provided in Table 4.
Table 4

Tests of Normality by Feedback Group

<table>
<thead>
<tr>
<th>Group</th>
<th>Kolmogorov-Smirnov</th>
<th>Shapiro-Wilk</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Statistic</td>
<td>df</td>
</tr>
<tr>
<td>CMap 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>expert</td>
<td>.156</td>
<td>14</td>
</tr>
<tr>
<td>peer</td>
<td>.189</td>
<td>14</td>
</tr>
<tr>
<td>CMap 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>expert</td>
<td>.188</td>
<td>14</td>
</tr>
<tr>
<td>peer</td>
<td>.211</td>
<td>14</td>
</tr>
</tbody>
</table>

* This is a lower bound of the true significance.

For all groups, the Shapiro Wilk statistic showed p-values which were greater than \( \alpha = .05 \). This indicated that there was insufficient evidence to reject the null hypothesis and that the data for these groups were likely to have come from a normal distribution.

Levene’s Test of Homogeneity of Variance was used to test that the dependent variable variances were the same across all levels of the variables. Results were non significant for both Concept Map 1 scores (\( F(1, 26) = 2.42, p = .132 \)) and Concept Map 2 scores (\( F(1, 26) = 1.61, p = .216 \)), which indicated that the condition of homogeneity of variance was not violated in either group.

Results. To return to the first research question, the researcher sought to determine whether a statistically significant difference existed between the Concept Map scores of those who received Peer Feedback and those who received Expert Feedback. The mixed repeated measures ANOVA provided results for a Within Subjects Effect (Time), a Between Subjects effect (Feedback Group), and an interaction effect for Time*Feedback Group. The interaction effect for Time*Feedback Group was non
significant \((F(1, 26) = .904, p = .351, \eta_p^2 = .034)\). The observed power for this result was .15. Additionally, the main effect of Feedback Group, the Between Subjects variable, was non significant \((F(1, 26) = 3.50, p = .073, \eta_p^2 = .118)\). The observed power for this result was .44. This finding indicated that although there exists a difference in scores between groups, this difference was not statistically significant at the \(p = .05\) level. However, the probability value was close to \(p = .05\) and the power level of the test was very low, which could indicate that the effect was significant, but the test lacked power to identify. As a consequence, dismissing this result might result in a Type II error.

Finally, the main effect of Time, the Within Subjects variable, was significant \((F(1, 26) = 4.92, p = .035, \eta_p^2 = .159)\). The observed power level was .57. These results suggested that the construction of Concept Maps over time had a significant effect on Concept Map scores and accounted for approximately 16% of the apportioned variation. These results are displayed visually in the line graph provided in Figure 5.

![Figure 5. Line graph - Effects of Time and Feedback Group on CMap scores.](image-url)
The change in Concept Map scores by Feedback Group is indicated in Figure 6.

The changes indicated that both groups experienced primarily growth in scores from the first map to the second. However, there were more scores that regressed in the Peer group, and there were more scores that increased by greater than 25 points in the Expert Map group.

![Figure 6. Change in Concept Map scores after feedback sessions.](image)

**Research Question Two Results**

The 12-item questionnaire was given after the completion of the feedback session and the second Concept Map. The questionnaire consisted of 10 statements that used a 5-point Likert scale for responses, ranging from “Strongly agree” to “Strongly disagree.” It also consisted of two open-ended questions (see Appendix A for Student Questionnaire). The questionnaire was designed to elicit feedback around two themes – student perceptions regarding the usefulness of Concept Maps and student perceptions regarding the value of the feedback they received in the process. Responses to the 10 items were
combined to create a total score (“Utility of Activity”). A comparison of the utility score across the two groups yielded non significant findings ($t = .385, p = .704$).

**Likert scale responses.** Overall, the responses to the first 10 Likert scale questions indicated generally positive student perceptions of both Concept Map creation and feedback. All questions received a greater number of positive responses than negative responses. More specifically, 7 of the 10 questions had a median answer of four (Agree). Questions 7, 9, and 10 had medians of 3.5, 3, and 3, respectively. For these three questions, there remained a greater frequency of positive responses (4s and 5s on the Likert scale) than negative (1s and 2s on the Likert scale). Question 7 had 50% positive responses and 14% negative; Question 9 had 42% positive responses and 14% negative; and Question 10 had 39% positive responses and 32% negative.

To summarize results regarding the mapping process, most students found Concept Maps beneficial for seeing big ideas (75% Strongly agreed or Agreed, 3% Disagreed), for seeing different ways of understanding concepts (71% Strongly agreed or Agreed, 3% Disagreed), for organizing concepts into a visual picture to aid understanding (71% Strongly agreed or Agreed, 3% Disagreed), and for recognizing concepts that they did not yet understand (53% Strongly agreed or Agreed, 14% Disagreed).

The responses regarding feedback had slightly more variation. Students generally agreed that the feedback discussions helped them to see new relationships (50% Strongly agreed or Agreed, 3% Disagreed), helped them to see how concepts were related to one another (60% Strongly agreed or Agreed, 21% Disagreed), and helped confirm their existing ideas about how concepts were related to one another (79% Strongly agreed or Agreed, 3% Disagreed). Student responses were split regarding the statement that the
feedback helped them understand the content more than simply reading the textbook (39% Strongly agreed or Agreed, 32% Disagreed, 29% Neither Agreed or Disagreed).

**Open-ended responses.** Questions 11 and 12 on the questionnaire were intended to elicit student information about their feedback group discussions. Students were asked to record two or more questions that their group talked about during the discussion time regarding concepts or maps, and they were asked to describe what, if any, changes they would have made to their initial Concept Maps after they had discussed these maps together. These questions were analyzed for themes, and then responses were organized in a table by Feedback Group.

Question 11 was answered by all students. It asked students to record two or more items that they talked about during their group discussion. After analysis, three themes emerged: Elements, Relationships, and Big Ideas. “Elements” was a theme that reflected the more technical nature of Concept Maps and the mapping process. Answers that fell in this theme group mentioned a focus on definitions and word meanings (“What does ___ mean?”), placement issues regarding a word, which were often linked to definition questions (“Where would ___ go?”), and Concept Map structural considerations (“How do you do ___? Is there a right or wrong way?”). “Relationships” was a theme that encompassed those answers that involved conceptual relationships and connections. Many of these responses involved the word, “Why…?” A typical response in this category might have involved a question about why certain concepts were related. Finally, the “Big Ideas” theme reflected answers that were bigger than the exercise itself. These responses reflected discussion involving higher order thinking and answers suggested synthesis or analysis of ideas. A sample response for this theme was a group
discussion that centered on “the vital importance of bias and wording effects in statistics” (actual student response). The results are shown in Table 5.

**Table 5**

*Frequency of Question 11 Responses by Theme and by Feedback Group*

<table>
<thead>
<tr>
<th>Group</th>
<th>Elements</th>
<th>Relationships</th>
<th>Big Ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer</td>
<td>15</td>
<td>7</td>
<td>2</td>
</tr>
<tr>
<td>Expert Map</td>
<td>8</td>
<td>12</td>
<td>5</td>
</tr>
</tbody>
</table>

Question 12 was answered by all students. It asked students to describe changes they would have made to their initial Concept Maps after their discussion group session. After analysis, four themes emerged: Structure, Relationships, Presentation, and Big Ideas. “Structure” was a theme that reflected a student’s desire to change basic elements of the Concept Map. Responses that described moving a specific concept from one place to another or adjusting the structure of the map by adding branches or using columns or categories aligned with this theme. “Relationships” was a theme that described a student response about changing the relationships that were illustrated on the map. Responses covered by this theme involve improving links between concepts, creating more crosslinks, and showing more connections between ideas. “Presentation” was a theme that encompassed answers regarding making the map neater or improving the overall organization. “Big Ideas” was a theme that reflected changes that moved beyond map characteristics and involved analytical changes. The frequencies were recorded and then organized by Feedback group. The results are shown in Table 6.
Finally, the data from these two open-ended questions were coded with values ranging from one to three for Question 11 and one to four for Question 12. For Question 11, the themes were coded as follows: 1 - Elements, 2 - Relationships, 3 - Big Ideas. For Question 12, the themes were coded as follows: 1 - Presentation, 2 - Structure, 3 - Relationships, 4 - Big Ideas. The responses were coded so that they could be merged with other data and possibly reveal additional patterns.

Because these responses and subsequent codes were linked with students, the codes could also be used to further evaluate the Concept Map scores. It might be, for instance, that the lowest scorers had similar plans for revising their Concept Maps or that the highest scorers might have discussed similar ideas in their feedback groups. With this in mind, the researcher disaggregated the Concept Map scores by codes for Question 11 and then by codes for Question 12. The data were analyzed in this way to see if student perceptions about Concept Map creation and feedback would provide a more thorough understanding of the Concept Map structure scores or any further insights about the quantitative results. However, when analyzed in this way, there were no new results or patterns. These coded results, though, were also merged with structural classifications in a way that did yield some insights as described in the next section.

Table 6

*Frequency of Question 12 Responses by Theme and by Feedback Group*

<table>
<thead>
<tr>
<th>Group</th>
<th>Presentation</th>
<th>Structure</th>
<th>Relationships</th>
<th>Big Ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Expert Map</td>
<td>5</td>
<td>5</td>
<td>3</td>
<td>1</td>
</tr>
</tbody>
</table>
Research Question Three Results

Research Question Three explored the differences between groups when structural classification and student choice of starting terms was considered.

Structural classification. All Concept Maps were analyzed based on overall structure and given one of two structural categorizations – Discrete or Integrated. The map structures were designated “Discrete” if they consisted of discrete branches from the starting concept, if the branches were connected by few or no crosslinks, and if there were few or no crosslinks between concepts within a branch (#Crosslinks ≤ 3). Concept Map structures were classified as “Integrated” if they consisted of discrete branches off of the starting concept but the branches were connected by crosslinks and there were crosslinks between concepts within a branch (#Crosslinks > 3). These two structural classifications arose after consideration of the work by Kinchin et al. (2000), which classified structures as Chain, Spoke, or Network. These three classifications increased in complexity and involved considerations similar to those described for Discrete and Integrated. The use of the Chain structure was eliminated as it was overly simplistic for the Concept Maps from this study. The structural classification frequencies before and after feedback sessions appear in Tables 7 and 8.
Table 7

*Structural Classifications of Concept Map 1 by Feedback Group*

<table>
<thead>
<tr>
<th>Structure</th>
<th>Feedback Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peer</td>
</tr>
<tr>
<td>Discrete</td>
<td>8</td>
</tr>
<tr>
<td>Integrated</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 8

*Structural Classifications of Concept Map 2 by Feedback Group*

<table>
<thead>
<tr>
<th>Structure</th>
<th>Feedback Group</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peer</td>
</tr>
<tr>
<td>Discrete</td>
<td>7</td>
</tr>
<tr>
<td>Integrated</td>
<td>7</td>
</tr>
</tbody>
</table>

The results indicated a difference in both groups. When Expert Maps were used in feedback groups, there was a greater percentage of final Concept Maps that displayed Integrated structures – 79% of the second Concept Maps made with Expert Map feedback indicated Integrated structures, while only 50% of the second Concept Maps made without Expert Map feedback indicated Integrated structures. When considering percent change from Concept Map 1 to Concept Map 2, it was clear that change happened for those in the Expert Map group. In the Peer group, Integrated maps increased by 17% after feedback, while in the Expert Map group, the number of Integrated maps increased by 83%. The proportions of Discrete and Integrated classifications were analyzed using
separate 2-sided exact McNemar tests. The tests indicated that for both Peer and Expert Map groups, the proportions of Integrated maps before and after feedback were not significantly different at the $p \leq .05$ level (Peer $p = 1.0$; Expert $p = .065$).

These structural classifications then were merged with the questionnaire responses. While the questionnaire data initially seemed to indicate overall consensus on all questions, an analysis by structural classifications resulted in different patterns for three questions. Patterns arose with Questions 6 and 7, and the coded responses to Question 12. The results are shown in Figures 7 through 9.

Disaggregating the results to Question 6 (Figure 7) by structural classification indicated that students with Discrete map structures exclusively agreed or strongly agreed that the feedback sessions helped them with conceptual understanding (100%). Those students who made Integrated structural maps from the start were more mixed in their responses (42% Agreed, 33% Disagreed, and 25% Neither agreed nor disagreed).

![Classification of Concept Map 1]

*Figure 7: Responses to Question 6 by Classification (D = Discrete; I = Integrated). Question 6 stated, “The small group discussions helped you to better understand statistical concepts.” (Likert scale responses: 1 – Strongly disagree to 5 – Strongly agree).*
Considering the results to Question 7 (Figure 8) in the same way, students who had maps with Discrete structures overwhelmingly agreed or strongly agreed (82%) that the feedback sessions helped them to see new relationships. Students with Integrated maps responded more neutrally (25% Agreed, 58% Neither agreed nor disagreed, 17% Disagree).

![Classification Concept Map 1](image)

*Figure 8. Responses to Question 7 by Classification (D = Discrete; I = Integrated). Question 7 stated, "The small group discussions helped you to see new relationships between statistical concepts." (Likert scale responses: 1 – Strongly disagree to 5 – Strongly agree).*

Finally, disaggregating the results to Question 12 (Figure 9) by structural classification indicated a greater number of students with Discrete maps would have focused on changes to structure and relationships (85%). Additionally, of the students focused on revising presentation elements, the majority were those who had created Integrated maps (67%).
Figure 9. Coded responses to Question 12 by Classification (D = Discrete; I = Integrated). Question 12 asked, “If you were asked to revise your first Concept Map after talking about the maps together in your discussion groups and seeing other maps, what changes would you make?” (1 – Presentation, 2 – Structure, 3 – Relationships, 4 – Big Ideas).

**Choice of starting concepts.** Concept Maps also were analyzed based on the concepts that students chose to connect to the starting concept. While not an idea represented in the literature regarding structural analyses of maps, this researcher suggests it can fit in a structural analysis because a student’s first concept choices very strongly drives structural considerations (Novak, 2010). For this reason, it was included here. When designing the selection of ten concepts to be used in each of the maps, the researcher had created concept banks that consisted of five general terms (terms such as “bias” and “inference”) and five terms that were specific and could serve as examples (terms such as “wording effects” and “random digit dialing”). The organization of Concept Maps typically radiates away from the center concept in a move from general to specific (Novak, 2010). A map that continued from the given concept using general words first would be considered to have a stronger conceptual presentation than a map.
that began with an example, then used a general concept at the end of a branch, a placement usually reserved for the most specific terms (Novak, 2010).

The researcher analyzed those concepts that directly attached to the center concept and recorded how often each of the general and specific words was chosen. Responses were recorded by feedback group. When scores were analyzed overall without group consideration, there were a total of 81 starting concepts. Of these, 54 (67%) were general concepts and 27 were specific (33%). This indicated that although students in both groups made appropriate starting choices some of the time, there was a large use of specific terms to start.

Concept Map 2 was created after the Peer and Expert Map feedback discussions. The same counting of general and specific concepts on the maps made after feedback revealed that of the 71 total starting concepts used, 64 of them (90%) were general concepts and seven (10%) were specific (see Table 9).

<p>| Table 9 |
|---|---|
| <strong>Choice of Starting Concepts for Concept Maps 1 and 2 - Overall</strong> |</p>
<table>
<thead>
<tr>
<th>Map</th>
<th>Starting Concept</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>General</td>
</tr>
<tr>
<td>Concept Map 1</td>
<td>54</td>
</tr>
<tr>
<td>Concept Map 2</td>
<td>64</td>
</tr>
</tbody>
</table>

Evaluation of these results using a 2-sided, exact McNemar test indicated a statistically significant difference in the proportion of students using General starting terms pre- and post-feedback, \( p = .016 \).
When results were split by feedback group, there were differences between groups when the second Concept Maps were considered. A greater proportion of Concept Maps used the most general starting terms at the start – 97% of the second Concept Maps made with Expert Map feedback began with the general concepts, while only 83% of the second Concept Maps made without Expert Map feedback began in this way (see Table 10). The Peer feedback group results and the Expert Map group results were both analyzed using 2-sided exact McNemar tests. Results indicated that for both groups the proportions of students choosing General terms pre- and post-feedback were statistically significantly different (Peer $p = .002$; Expert Map $p = .008$).

Table 10

*Choice of Starting Concepts for Concept Maps 1 and 2 by Feedback Group*

<table>
<thead>
<tr>
<th>Group</th>
<th>Starting Concept</th>
<th>General</th>
<th>Specific</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer</td>
<td>CMap 1</td>
<td>27</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>CMap 2</td>
<td>29</td>
<td>6</td>
</tr>
<tr>
<td>Expert</td>
<td>CMap 1</td>
<td>27</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>CMap 2</td>
<td>35</td>
<td>1</td>
</tr>
</tbody>
</table>

**Summary of Results**

The researcher sought to examine the results from the Concept Map study. These results were in the form of quantitative scores given on two separate Concept Maps and on results from a short questionnaire given to students. This questionnaire provided both
quantitative data in the form of Likert scale responses using a 5-point scale and qualitative data from two open-ended questions. Additional analyses focused on structural classification of Concept Maps and on student choice of starting concept.

The quantitative scores from the Concept Maps were analyzed using a mixed repeated measures ANOVA to determine whether a statistically significant difference existed between the Concept Map scores of those who received Peer Feedback and those who received Expert Feedback.

The mixed repeated measures ANOVA provided results for a Within Subjects Effect (Time), a Between Subjects effect (Feedback Group), and an interaction effect for Time*Feedback Group. Of these, only the main effect of Time was significant ($F(1, 26) = 4.92, p = .035, \eta^2_p = .159$).

The responses for the first 10 questions on the Questionnaire were analyzed using frequencies for each Likert scale response. These were disaggregated by group and examined for patterns. There was consensus on most answers that students positively perceived Concept Maps and feedback. For all questions, the positive responses outnumbered the negative.

The two open-ended questions from the Questionnaire were analyzed and themes emerged. Three themes were similar for both – Elements (Structure), Relationships, and Big Ideas - and an additional theme – Presentation - was added for Question 12. The themes were then given point values of one to three for Question 11, and one to four for Question 12.

The data were then merged with the quantitative data from the Concept Map scoring and analyzed. Combining data in this way was meant to further inform the results.
that were obtained through scoring the structure of Concept Maps. Results of this merger were inconclusive and revealed no new patterns.

Finally, two additional analyses focused on structural classification of Concept Maps and on student choice of starting concept. All Concept Maps were examined based on overall structure and given one of two structural categorizations – Discrete or Integrated. After feedback, both groups experienced some changes in structural classification. These differences were analyzed using separate 2-sided exact McNemar tests. The tests indicated that for both Peer and Expert groups, the proportions of Discrete maps before and after feedback were not significantly different at the \( p \leq .05 \) level (Peer \( p = 1.0; \) Expert Map \( p = .065 \)).

Student choice of starting concepts was also examined. For both Concept Maps, the starting concepts were recorded, categorized as either general or specific, and frequencies recorded. On the initial Concept Maps, of the total concepts used to start maps, 67% were general terms and 33% were specific terms. This indicates that although students in both groups made appropriate starting choices some of the time, there was a large use of specific terms to start. After feedback discussions, students chose their starting terms differently, with 90% of the selected terms now being general and only 10% specific. Evaluation of these results using a 2-sided, exact McNemar test indicated a statistically significant difference in the proportion of students using General starting terms pre- and post-feedback, \( p = .016 \).

When results were split by feedback group, differences between groups were evident with the second Concept Maps. A greater percentage of the Concept Maps made after Expert Map feedback (97%) used the most general starting terms at the start than
after Peer feedback (83%), but both groups experienced statistically meaningful change. The Peer feedback group results and the Expert Map group results were both analyzed using 2-sided exact McNemar tests. Results indicated that for both groups the proportions of students choosing General terms pre- and post-feedback were statistically significantly different (Peer $p = .002$; Expert Map $p = .008$).

These results are addressed in Chapter Five. The findings are discussed in terms of practical significance and considered within the context of findings within the literature. Chapter Five also includes a discussion of the limitations of the study and suggestions for future research.
Chapter 5

Discussion of Results and Conclusions

Chapter Overview

This chapter begins by providing a discussion of results for each of the three research questions explored by this study. The discussion for each question includes an analysis of results and connections of the results with findings from other studies in the literature. After this analysis by research question, limitations of the study and possible improvements are discussed. The chapter concludes with discussions on the implications of findings for theory and practice, as well as some thoughts regarding future research directions.

A convergent mixed methods design was used for this exploratory study. This design was chosen as it best fit the research questions due to its use of both quantitative, qualitative, and merged results. The intent was to gather as much information as possible to more fully understand the results. Concept Maps by their very nature and structure do not lend themselves to easy quantitative measure, so this approach provided different insights into the data that might lead to greater understanding.

Research Question One Discussion

For Research Question One, quantitative data were gathered and analyzed using inferential statistics. Concept Map structures were scored using a version of a well-established scoring method created by Novak (1984) (see Figure 3), and then scores were analyzed using a mixed repeated measures ANOVA. There were two independent variables, each with two levels – Time (1, 2) and Feedback Group (Peer, Expert Map) – and the dependent variable was Concept Map score. The sample size consisted of 28
students. Descriptive and inferential statistics were run, and results were analyzed at a significance level of .05.

The data analysis for the first research question indicated that the interaction effect of Time*Feedback Group and the Between Groups main effect of Feedback Group were non significant at the $p = .05$ level. However, there was a statistically significant result for the Within Groups main effect of Time ($F(1, 26) = 4.92, p = .035, \eta^2_p = .159$). This finding indicated that making the maps across time accounted for about 16% of the variation in scores. The partial eta squared value represents the amount of “Effect + Error” variance that can be attributed to the Effect, and so in this instance, a partial eta squared value of .159 indicates that about 16 percent of the variance can be attributed to Time. Because the two Concept Maps were made close together in time, it seems unlikely that this improvement would be attributable to things outside of the Concept Making activity, such as additional learning. This suggests that peer feedback, which both groups participated in between the Time 1 and Time 2 measures, had a large effect on the Concept Map scores.

The observed power of this result was .57, well below the .80 level that is generally accepted for offering a strong probability of correctly representing the null hypothesis (Gall et al., 2007). The low power was a result, in part, of the very small sample size. An a priori power analysis using G*Power 3 for two groups and four cells at an alpha level of 0.05, with a desired medium effect size and a desired power of 0.80, indicated that the required sample size was 82. This sample ($N = 28$) was well below the suggested level. A larger sample was desired, but the choice was made to trade power for a small sample to start for two reasons. First, the choice of a smaller sample fit the
exploratory nature of the study. Before bringing the study idea to a much larger sample, the researcher desired more information about possible effects and possible areas of greater focus if the results indicated significant effects. Second, the study was constrained by the size of the available statistics class. Securing class time in statistics classes was difficult given the quantity of content that must be covered and the resulting preciousness of class time, but the alternative of attempting this study in a laboratory setting did not align with certain elements of the research questions. Because of the study’s low power to detect differences, it could be that the analysis missed findings of significance, which might lead the researcher to make a Type II error. However, of more concern than this possibility was the inability of the scoring system that was used to fully capture the structural changes that were occurring in student maps.

The scoring system very likely contributed to findings that yielded difficult to interpret information. While the method chosen was selected for its concentration solely on structural elements, this stringent focus led to very low scores for maps that when considered holistically showed structural improvements. For example, no links or crosslinks were able to be considered “valid” when scoring if they lacked a linking word. In viewing student maps, this lack of consideration for structural improvement due to a missing connector word quite likely led to a lack of variability in the scores. Even a slight change in the weightings of the elements – a strategy commonly used throughout the literature – resulted in very different effects. This difficulty in assigning one numerical score to Concept Maps that accurately captures all relevant elements aligns with research that has found scoring in general to be a challenge (Watson et al., 2016). It also specifically resonates with the study done by Watson, Pelkey, Noyes, and Rodgers (2016)
in which researchers scored the same maps using different systems and weightings and came up with different results for each.

A similar limitation occurred in this study when two similar scoring systems were used that weighted links and crosslinks differently given the presence or absence of labels, for example, which resulted in two different outcomes. In one of these results, large group differences were present, but in another, these differences were nearly erased, resulting in groups that seemed fairly homogenous going into the treatment. The two scoring methods also indicated different effects of Feedback Group on Concept Map score. The first method (which gave differing point values to links with and without labels but counted them both in some way) seems to indicate that while the treatment affected both groups, it affected the Expert Map group in a much greater manner. The second (which gave points only to those links which were labeled) shows effect slopes that seem to indicate that the treatment had similar levels of effect on both groups, but that the groups began in different places based on initial scores. These results are seen in Figure 10.

Figure 10. Sample line graphs of re-scored effects. These indicate the effects of Feedback Group and Time on Concept Map scores when the data was re-scored using two systems that weighted Concept Map characteristics differently.

The difficulties in scoring Concept Maps encountered during this study have the effect of placing this study and its results squarely within the findings represented in the
literature. Despite many studies stretching from the 1980s, no one “best” scoring method has emerged. In fact, the opposite has occurred and researchers continue to develop many new methods (Nesbit & Adesope, 2006; Watson et al., 2016). Numerous differing systems have been created in an attempt to more accurately quantify the results of maps in some way so as to make them analyzable, and as a result, those maps that are easiest to score have had growing representation in the literature (Hay, 2007; Kinchin et al., 2000; Miller & Cañas, 2008; Rye & Rubba, 2002). However, creating a mapping system that is easy to score has led, at times, to maps that are overstructured and leave little room for student conceptual representation. Some of these maps are fill-in-the-blanks maps with a few concepts missing, which can be scored quickly and with high accuracy, but that have limited student input regarding structure or conceptual organization. These versions of Concept Maps as discussed at length in the literature review (Chapter 2) do not rely on student prior knowledge structures or on student derived understandings of conceptual relationships, and they often result in lower level answers that are guesses as to what relationship the teacher is asking about. This presents a complicated trade-off between error and certainty that researchers have clearly had to struggle with when exploring Concept Maps. It also explains the divergence of the literature which splits between student created maps and pre-constructed fill in maps.

The significant main effect of time aligns with some findings in the literature. The main effect of time might illustrate in part the idea expressed by Nesbit and Adesope (2006) that some improvement is likely to be seen as students develop their skills in creating Concept Maps. Greater fluency with the learning tool also may explain some of the improvements seen in the structural changes from Discrete to Integrated and also
possibly some of the changes shown in student choice of starting concepts. However, for the present study, the skill development effect is likely to be limited because the two Concept Maps studied were created so close together in time. If maps were analyzed across the time of a few weeks or a semester, for example, more importance might need to be placed on the effect of mastery of the learning tool.

This study’s results differ from some findings in the literature in that typical findings regarding Concept Maps and learning have tended to have large effect sizes on learning, especially when measured using objective measures. Findings across the research by Nesbit and Adesope (2006) might be used to explain some of the large effect sizes associated with Concept Maps as learning tools and one possible reason for no large effect in this study. In collecting and analyzing research for their meta-analyses, the researchers noted that many Concept Map studies may have shown such large effects because the comparisons that researchers had chosen were very different from each other (Nesbit & Adesope, 2006). For instance, engagement effects for Concept Maps were very high when the studies concerning engagement compared Concept Map making to something much less engaging such as lecture. Another example might be those studies that compared Concept Mapping as a learning tool (a tool that requires creative approaches and thinking) with standard note-taking (a tool that requires little independent decision-making or thinking). Nesbit and Adesope (2006) found this type of Concept Map comparison using very different things in much of the literature surrounding Concept Maps and indicated these large differences in the natures of what were being compared may have accounted for some of the large effect sizes. In this study, structural
elements of maps are compared against structural elements of maps, and the amount of change expected over this short duration might be very small.

However, the lack of significance does not necessarily mean that there was no effect. The limited power of the study allows for a greater chance of a Type II error being made. Additionally, the scoring was not meant to be the only or final means of analysis of change. In line with studies such as that done by Hay, Wells, and Kinchin (2008) and in line with mixed method design (Creswell & Plano Clark, 2018), the research questions for this study were meant to be explored using both qualitative and quantitative findings with the hope that they would complement each other and provide a fuller picture of effects.

**Research Question Two Discussion**

The questionnaire results were designed to provide additional information that would help to answer the second research question, which asked if there might be differences by Feedback Group regarding perceptions of usefulness of Concept Maps and feedback. The 12-item questionnaire was given after the completion of the feedback session and the second Concept Map. The questionnaire consisted of 10 statements accompanied by a 5-point Likert scale for responses, which ranged from “Strongly agree” to “Strongly disagree.” It also consisted of two open-ended questions (see Appendix A for Student Questionnaire). The questionnaire was designed to elicit feedback around two themes – student perceptions regarding the usefulness of Concept Maps and student perceptions regarding the value of the feedback they received in the process.

Overall, the responses to the first 10 Likert scale questions indicated generally positive student perceptions of both Concept Map creation and feedback. All questions
received a greater number of positive responses – Strongly agree or Agree – than negative responses – Strongly disagree or Disagree. There was consensus that these maps functioned as helpful scaffolds for students’ learning of statistics concepts. Additionally, students seem to have appreciated the collaborative discussions, finding the group work beneficial to their understanding. This finding aligns with studies that have indicated that students tend to do well collaboratively when working together with Concept Maps, especially when provided with discourse scaffolds (Nesbit & Adesope, 2006). These supports when used with maps were associated with more focused student discussions and better outcomes when students were tasked with refining and creating conceptual relationships together. Although the general analysis of the questionnaire items seemed to indicate consensus, the additional analysis using structural classifications revealed that some differences in opinion coincided with structural group. These will be considered in the Research Question Three discussion.

The analysis of the open-ended questions, Questions 11 and 12, resulted in more detailed information about student plans for revision and their discussions during feedback. Question 11 asked students to record two or more items that they talked about during their group discussion. Students answered this question fully, with all students providing two clear and discrete items that were discussed. After analysis, three themes emerged: Elements, Relationships, and Big Ideas.

The answers to Question 11, presented in Table 5, are compelling. In the groups led and directed by peers, the discussions regarding Concept Maps most often centered on low level technical issues, such as definitions or placement of certain words. This description of the discussion topics was voiced 15 times in the Peer Group responses and
only eight times in the Expert Map Group responses. The Peer Group members described some discussions centering on relationships between concepts (seven responses), and a very few that took the discussions to levels of analysis with big idea questions (two responses).

Alternately, the groups led and directed by students and which had the presence of an Expert Map to show one more formalized way of presenting and organizing concepts had more reports of discussions that were on a higher level of relationships (12 responses versus seven in the peer groups) and big ideas (five responses versus two in the peer groups). This difference might provide one explanation for subsequent low scores and the lowering of some higher scores on a second map. Students in groups that were focused on lower level features would not have received the benefits of group discussions on broader relationships and consequently might have struggled to see interrelationships on their second map, even though the content was similar.

Question 12 provided more detailed results as well. This question asked students to describe changes they would have made to their initial Concept Maps after their discussion group session. After analysis, four themes emerged: Presentation, Structure, Relationships, and Big Ideas. The frequency of responses is provided in Table 6. The responses to this question were similar by feedback group, suggesting that students in both feedback groups would focus on making similar changes to their Concept Maps if given the opportunity and that these changes would involve better structures and relationships. One interesting difference, however, concerned the focus of changes to the presentation of the map. Whereas the presentation responses in the Peer Feedback groups focused on the word or idea of “neatness” almost exclusively, those in the Expert Map
groups focused on organizational tidying – as described by phrases such as “make it more organized.” This might suggest a greater desire by one group to refine the overall organization or the conceptual relationships versus simply presenting the map more neatly, or it might simply be a slightly different way of articulating the same idea.

Finally, the data from these two open-ended questions were coded with values ranging from one to three for Question 11 and one to four for Question 12. These values then were plotted against Concept Map 1 scores. The data were merged in this way to see if student perceptions about Concept Map creation and feedback would provide a more thorough understanding of the Concept Map structure scores or any further insights about the quantitative results. For Question 11, the themes were coded as follows: 1-Elements, 2-Relationships, 3-Big Ideas. When plotted against Concept Map 1 scores and grouped by Feedback Group, these responses showed no discernible pattern and as a result they yielded no further information about the scores or feedback groups.

For Question 12, there was a pattern of interest when plotted against Concept Map 1 scores and grouped by Feedback Group. This pattern regards the responses on presentation. While the student answers used language to suggest neatness and “cleaning up” of the presentation, the merging of the data revealed that an envisioning of cleaning up the way the Concept Maps looked was held primarily by those who scored the highest on their first maps. Those with lower scores tended to see structural or relationship concerns as areas of next focus. These results are possibly beneficial in that they might indicate that most students, after mapping and feedback, would hope to improve their maps in ways that indicate more complete conceptual relationships. This shows a possible new awareness of the ways in which their initial maps fell short. It also might
indicate that after feedback of either type, students were able to self-assess in a fairly accurate manner regarding the structure and adequacy of their Concept Maps and regarding the changes that need to be made to strengthen it. These findings might also illustrate that those students with high quality maps from the start were generally conscientious about the characteristics and presentation of their work. The merged data appears in Figure 11.

![Figure 11](image)

*Figure 11. Merged data – Question 12 results with Concept Map 1 scores.*

**Research Question Three Discussion**

The classification of Concept Maps based on structure and the merging of this classification with other results provided an additional way to analyze the existing data. As described in the previous chapter, all Concept Maps were analyzed based on overall structure and given one of two structural categorizations – Discrete or Integrated. The map structures were designated “Discrete” if they consisted of discrete branches from the starting concept, if the branches were connected by few or no crosslinks, and if there
were few or no crosslinks between concepts within a branch ($\#\text{Crosslinks} \leq 3$). Concept Map structures were classified as “Integrated” if they consisted of discrete branches from the starting concept but the branches were connected by crosslinks and there were crosslinks between concepts within a branch ($\#\text{Crosslinks} > 3$).

When the classifications were organized by Feedback Group, there were clear differences between Feedback Groups in the numbers of Discrete and Integrated maps for each, with a larger number of Integrated maps present in the Expert Map feedback group. When considering percent change in Integrated classifications from Concept Map 1 to Concept Map 2, greater change happened for those in the Expert Map group. In the Peer group, the number of Integrated maps increased by 17% after feedback, while in the Expert Map group, the number of Integrated maps increased by 83%. The proportions of Discrete and Integrated classifications were analyzed using separate 2-sided exact McNemar tests. The tests indicated that for both Peer and Expert groups, the proportions of Integrated maps before and after feedback were not significantly different at the $p \leq .05$ level (Peer $p = 1.0$; Expert Map $p = .065$). However, the $p$-value for the Expert Map group is close to $p = .05$, and although it is greater than .05, the value is still very small ($p = .065$). Given that this widely accepted $p$-value is recommended as a guideline but not as a stringent rule, the results from the Expert Map group might be considered by some to be statistically significant (Field, 2013). Additionally, the McNemar test value of $p = 1.0$ for the Peer group data indicates that the pre-post feedback proportions for this group are statistically the same. These results indicate that the proportions of Integrated structures changed in significant ways after feedback for one group and not for the other.
The strength of this finding lies in two places. First, the results are not dependent upon an easily manipulated Concept Map score. Instead, a fixed classification is given and this classification is used in the analysis. Second, the results appear clear—no significant change in one group in the proportions and significant change in the other group. These findings give greater clarity to the potential presence of a meaningful difference between groups.

This finding seems to clearly indicate that Expert Maps during feedback might strongly influence the integrated structure of subsequent maps. This is a result that might be of practical significance as it would be easy for teachers to provide Expert Maps for student use during feedback. Students could also use them to check for misunderstandings and to reinforce correct conceptual relationships while working independently or in groups. These maps also eliminate the need to score a map or perfect it. Instead, they offer one full conceptual representation that students can use to refine their own understandings.

It might also be that feedback regardless of type proves to be valuable in the refinement of student conceptual representations. This certainly seems to be the case when considering the results from analyzing student starting concepts, as students overall, regardless of feedback group, chose more appropriate starting concepts with their second maps. Analysis of starting concepts before and after feedback indicated students improved in their starting choices. Before feedback, students chose overly specific terms to start 33% of the time, while after feedback this number dropped to 10%. Analysis of the overall class results using a 2-sided McNemar test indicated that the pre-post feedback proportions of general and specific terms were statistically significantly
different \( (p = .016) \). Further analysis revealed that the improvements were statistically significant in both groups (Peer \( p = .002 \); Expert Map \( p = .008 \)).

Finally, the classification of maps as either Discrete or Integrated provided a clear way to disaggregate the responses to the questionnaire data. This enabled the researcher to see if differences existed within groups of respondents. In fact, while the data looked homogeneous when simply recording frequencies, when reviewed after separating respondents into Discrete or Integrated groups based on their maps, the data provided more detailed and group specific information, most especially for questions 6, 7, and 12.

Disaggregating the results to Question 6 by structural classification indicated that students with Discrete map structures exclusively “agreed or strongly agreed” that the feedback sessions helped them with conceptual understanding (100%). Those students who made Integrated structural maps from the start were more mixed in their responses (42% Agree, 33% Disagree, and 25% Neither). This finding is important as it suggests that one group might benefit more from the use of Concept Maps and feedback. Those students who begin with poorly structured Concept Maps might benefit most from this learning approach. Concept Maps and feedback might be important learning supports and might benefit those who need the conceptual scaffolding the most. Those students who begin with Integrated maps might already have conceptual relationships well organized and understood, and consequently might find the use of these somewhat helpful, or possibly not helpful. This finding possibly aligns with the research that indicates that certain groups might experience greater benefits from using Concept Mapping than others, especially groups that begin with poorer prior knowledge structures (Nesbit & Adesope, 2006). Alternately, it might also be the case that the students in the Expert Map
groups were not cuing in on the Expert Map as a source of feedback, and so answered this question using just their experiences with peer discussions.

The results for Question 7 were clearer regarding interpretation, but possibly less surprising. These results indicated that students who had maps with Discrete structures overwhelmingly Agreed or Strongly Agreed (82%) that the feedback sessions helped them to see new relationships. Students with Integrated maps responded more neutrally (25% Agree, 58% Neither, 17% Disagree). This finding makes practical sense in that it is likely that students who can form an integrated conceptual picture from the start might already have a relatively sophisticated sense of how to relate concepts and might not learn enough new relationships from the use of maps and feedback to answer this statement in the affirmative. Those students who cannot yet see a group of concepts in relationship to one another might have the most to gain. They have the potential to see many new relationships given that they have seen so few from the start.

Finally, the results to Question 12 indicated that students after feedback might have a fairly accurate sense of how to revise and improve their Concept Map structures. The results to Question 12 by structural classification indicated a greater number of students with Discrete maps would have focused on changes to structure and relationships (85%). A map would have been considered Discrete if it showed little to no interrelationship of ideas and an overly simplistic structure. Students with Discrete structures would need to revise both structures and conceptual relationships before considering presentation items or big ideas. It seems as if students accurately recognized their shortcomings and knew how to plan for adjustments and improvements. If this is
supported through further research, this might be a powerful tool for those interested in more self-directed student learning and an important practical finding.

**Limitations**

This study was intended to be exploratory, but even with this in mind, the intention was to construct a carefully considered study. Thought was given to controlling error and reducing the introduction of new confounds. Despite these efforts, there are some important limitations.

The convergent mixed methods design was chosen for its ability to provide a fuller understanding of a question through its inclusion of a variety of data. This design worked well for the intents and questions of this study. However, the sample for this design was a concern. The sample size could be increased to provide greater power for detecting significant results and greater ability to generalize. While the very small sample size of $N = 28$ was fitting for the type of questions that were studied, it resulted in very low power to detect statistical results and very low generalizability.

Another design limitation was the absence of a control group. This absence is a common weakness in the literature and was noted by Nesbit & Adesope (2006) as a reason for the lack of inclusion of many studies in their meta-analysis. In this study, it was a design choice that was made due to the limited number of available students and the lack of research done in this specific area, as well as due to the exploratory nature of the questions. It is possible that future studies would now have more focused questions and could use control groups to explore more specific differences.

In addition, a convenience sample was used, which also affects generalizability. However, this type of sample is a reality of educational studies because a randomized
design regarding sample participants is not attainable or realistic for classrooms of students. The use of a random sample would have been possible if the study was conducted in a laboratory setting, but this was not an appropriate option given the nature of the research questions being explored here.

The use of the chosen scoring system became a methodological issue. The scores were widely inaccurate at providing a numerical interpretation that accurately portrayed the maps. While it may have been adequate in previous studies using different groups, in this study it very likely exacerbated the differences in results. This limitation affects the findings from the first research question, and while a difference seemed to be present in every variation of the scoring system used, this finding would need to be replicated using a more reliable measure.

Another issue arose from the set time given for Concept Map construction regarding early finishers. When a student finished early and considered the representation complete, there may have felt pressure to add more detail if other students nearby were working furiously. Alternately, if a student was working and noticed that others had already finished, this may have provided a cue to stop working even though more would have been added. A greater awareness of the implications of the social nature of learning might have resulted in more protections against lack of independence of results. This limitation is mentioned as there may have been a violation of the independence of scores on the first Concept Map. While difficult to determine given the wide range of results from different weights of the scoring elements, there at times seemed to be a non-random pattern to results by group, with more low Concept Map 1 scores falling into the peer group even before the groups were announced. Despite the fact that randomization
procedures were used to assign students to groups and that students had no idea of groups until after these initial maps were made, there seemed to be differences by group score. The researcher considered three explanations that may have contributed to this result. First, the sample size ($N = 28$) was small, and while a small size has been used often in the literature for this type of study, the small size in conjunction with a very small classroom might have outweighed the effects of the randomization attempts. Second, learning in a classroom is invariably a social event. Students knew one another, worked at long open tables where they could see the work of one another both down the rows and in rows in front of them, were familiar with the room, and knew the expectations of the teacher, to name just a few, and these factors together might have affected the randomization effects. If students took cues from one another about when to consider their maps done, for example, or how many labels were “enough,” then the independence of observations condition would have been violated and the parametric model would no longer be appropriate.

The instrument used to gather student perceptual data was also a limitation as it was researcher created and specific to the study. While the researcher had hoped to use an existing measure, there was not an instrument in use that fit the questions the researcher was trying to answer. The concerns in using a researcher created questionnaire lie in the reliability and generalizability of the results. As this was an exploratory study, these concerns were considered but the potential benefits that accompanied answers to the specific questions asked outweighed these concerns.

Time and experimenter bias were additional limitations. A longer period of student use of Concept Maps before the actual study took place would have been more
ideal so that error and effects associated with the learning of a new tool could have been minimized. This limitation sprung from the limited nature of available class time, and may have resulted in true effects being minimized or missed due to the additional noise present in the results. Experimenter bias may also have occurred. The researcher attempted to provide Expert Maps to groups in a casual manner, but the fact that these maps were coming from the researcher may have influenced the perceived valued.

These limitations were all considered before the study and choices were made to work within these parameters given the realities present, but they limit the scope of what can be said about the results.

There are many improvements that might be made, many of which have already been discussed. In line with the limitations discussed here, possible additional improvements include the use of samples from differently focused classes that would include more diversity of learners. The chosen classroom consisted of a majority of students with declared interests in science. Of possibly greater interest would be how this learning tool affected students who were not science- or math-minded, but who were enrolled in statistics classes. Another group of interest might be those who struggle with learning or those who have had rote learning experiences and are struggling with conceptual thinking.

To address the prior concern regarding shared results and lack of independence of observations, the researcher considered whether a second task might be offered to allow students to feel finished and then move on at their own pace, but the presence of a second task might also encourage some students to rush, with the knowledge that they had two things to do. Another option might be the presence of dividers, but this consideration, in
the end, needs more thought and would depend in part on the group(s) in future studies. Additionally, a greater physical separation of students attained through meeting in an overly large room may alleviate this issue, spreading students out enough so that this is no longer such a great concern.

Finally, a stronger and more direct qualitative component might also enrich the findings in a way that provides clearer illumination of results. The two open-ended questions did provide details that enriched the results from the quantitative analyses. However, components such as those that tracked student perceptions or choices across time, discussions with students about the mapping process, and reflections as students revised their maps might be beneficial additions to future research.

**Implications for Theory**

Because of the limited and exploratory nature of this study, the implications for theory are modest. In general, however, the findings seem to add some support for the benefits of feedback in the classroom and more specifically when using Concept Map construction. Student map structure improved over time, and while it cannot be said that feedback alone caused this improvement, it did not appear to negatively affect performance. The structural classifications clearly seemed to change based on feedback, but this finding would need to be explored further for support. Additionally, student perceptions about the feedback groups were overall very positive, which adds another layer of support for the combination of peer/student feedback and Concept Maps.

Theoretically, Concept Maps draw support from constructivist epistemology, specifically social constructivism, and from Ausubel’s learning psychology, and this study’s results align with these ideas even if the results are not strong enough to further
them substantially. Constructivism’s foundational idea is that knowledge is constructed by learners who take in new information and modify it in an interplay with their existing prior knowledge. It is an active process, and requires learners to engage with learning, not simply observe. The positive perceptions recorded by students in the questionnaire seem to support that they viewed Concept Maps as learning tools that helped them modify their understandings, whether this was in the form of seeing new relationships between concepts or benefitting from constructing and discussing an overall big conceptual picture. This integration of meaning in prior structures is an important part of social constructivism, and the groupwork feedback with peers supported the idea that meaning and understanding is constructed together with others in social and cultural contexts and then internalized (Chaiklin, 2003). It is important to realize that the results from this study regarded Concept Map structure only, not the associated learning of students, so it is crucial not to over-reach regarding results. It is also important that results be interpreted conservatively, and not generalized beyond the study group.

**Implications for Practice**

Student engagement was high during the study, and the use of this learning tool might be a beneficial way to engage students while building on student prior learning. This engagement might also be a pathway to meaningful learning, given that part of what is required according to Ausubel (1966b) is the choice by students to actively engage. The results suggest that this learning tool has the potential to help change the way students structure their conceptual understandings, possibly moving them from Discrete to Integrated representations, but more research needs to be done to support this. Results
also indicate that those students with the most discrete initial representations perceived that they benefitted from making the maps and from feedback.

While this greater conceptual understanding might not be rewarded within some areas of the current educational system, it is the goal and desired end for statistics learning, at least as articulated by the GAISE (2016) guidelines. Additionally, this study was meant to explore Concept Maps as learning tools, and when considered as tools and not the ends of learning, they can supplement any type of statistics class orientation and might serve as a helpful bridge to conceptual understanding.

The use of Expert Maps and feedback also provides an easy to apply and low- to no-cost option for scaffolding student conceptual representations. Allowing students to create maps first alone challenges them to productively struggle with what they know, and the follow-up using feedback groups and Expert Maps gives learners ways to access knowledge at whatever levels they need. This supports the use of a learner’s ideal range for learning, what Vygotsky (1978) called the “zone of proximal development.” Scaffolding in this way allows students to integrate conceptual structures in meaningful ways. Both Expert Maps and Concept Maps become learning scaffolds that allow learners to operate within their zones of proximal development.

Finally, meaningful learning involves learners choosing to learn in a way that brings new information into meaningful relationship with what the learners already know (Ausubel, 1966a). Learners have to actively bring new ideas into interplay with pre-existing knowledge, and this integrates new knowledge into their existing conceptual understanding. The study actively required students to portray conceptual relationships, discuss other representations, and then refine their understandings. Simply asking
students to create Concept Maps with a given set of concepts requires this struggle, refinement, and integration. Concept Maps also allowed students to express their own understandings in a meaningful, creative, and collaborative manner.

**Conclusions and Further Research**

This study was meant to be an exploratory study. Given the uncertain results from the quantitative analysis of Concept Maps, it might be worthwhile for future research to explore feedback and Concept Map creation in a more qualitative fashion. While the non significant findings might seem to indicate that this combination of feedback and mapping is not effective for changing conceptual representations, an extensive body of research has indicated that effects using Concept Maps are widely varied and very sensitive to the task, to the measurement of outcomes, and to the scaffolds that are in place, so there may not be enough information yet to make a conclusion (Nesbit & Adesope, 2006). In terms of future research, it might be meaningful to explore the ideas from this study but with classes of students who are taking statistics and who are not math or science majors. Could this approach with feedback and Expert Maps help with the conceptualizations of big ideas in math classes? And, how should feedback group work be structured to encourage a movement from Discrete structures to Integrated? It also might be interesting to explore how the use of Expert Maps during feedback might affect the maps of students who struggle with conceptual understanding due to early years of rote learning. Finally, given the results of the quantification of Concept Maps, great scrutiny should accompany any result that relies solely on this evaluation method. In future studies, more attention and great care should be given to different ways of capturing exactly what is being studied regarding Concept Maps.
References


Cobb, G. W. (1992). Teaching statistics. In L. Steen (Ed.), *Heeding the call for change* (MAA Notes No. 22) (pp. 3-43). Retrieved from https://pdfs.semanticscholar.org/d8a8/7ea99e0f8babd4fe09134e64c3a751677f06.pdf#page=14


Kinchin, I. M., Hay, D. B., & Adams, A. (2000). How a qualitative approach to concept map analysis can be used to aid learning by illustrating patterns of conceptual

http://dx.doi.org/10.1080/001318800363908


http://www.jstor.org/stable/4124424


https://www.researchgate.net/publication/228761562


http://dx.doi.org/10.1080/00220670009598724


Appendix A

Student Questionnaire

<table>
<thead>
<tr>
<th>Please circle the number that best indicates your response.</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
<tr>
<td>Strongly agree!</td>
</tr>
</tbody>
</table>

1. Making Concept Maps helped you to see different ways of understanding the same concepts. 5 4 3 2 1
2. Making Concept Maps helped you to see big ideas in statistics. 5 4 3 2 1
3. Seeing statistics concepts organized together in a visual picture helped you understand the concepts better. 5 4 3 2 1
4. Making Concept Maps helped you to better understand statistics concepts. 5 4 3 2 1
5. Making Concept Maps helped you to recognize concepts that you don’t yet understand. 5 4 3 2 1
6. The small group discussions helped you to better understand how statistics concepts were related to one another. 5 4 3 2 1
7. The small group discussions helped you to see new relationships between statistics concepts. 5 4 3 2 1
8. The small group discussions helped you to confirm your existing ideas about how concepts were related to one another. 5 4 3 2 1
9. The small group discussions helped you to correct any misconceptions or misunderstandings you had. 5 4 3 2 1
10. The small group discussions helped you to learn more than what you learned from just reading. 5 4 3 2 1

11. Write two (or more) questions about the concepts or maps that your group talked about during your group discussion. (You can continue onto the back if you need more space.)

12. If you were asked to revise your first Concept Map after talking about the maps together in your discussion groups and seeing other maps, what changes would you make?

A little about you:
1. Age ______
2. University Grade Level (circle one) Freshman Sophomore Junior Senior
3. Intended University Major
4. Gender Identification (circle one) F M Other
5. Ethnicity (indicate with a check mark)
   ___ Hispanic or Latino or Spanish Origin
   ___ American Indian or Alaskan Native
   ___ Asian
   ___ Native Hawaiian or other Pacific Islander
   ___ Black or African American
   ___ White
   ___ Two or more races
Appendix B

Expert Concept Map

Create a Concept Map below using the given focus question and concepts.

Focus question:
How does the way you choose people for your sample (sample design) relate to the inferences you make from your data?

Concepts:

<table>
<thead>
<tr>
<th>random digit dialing</th>
<th>do not call list</th>
</tr>
</thead>
<tbody>
<tr>
<td>convenience sample</td>
<td>under-coverage</td>
</tr>
<tr>
<td>random sample</td>
<td>chance</td>
</tr>
<tr>
<td>voluntary response sample</td>
<td>inferences</td>
</tr>
<tr>
<td>bias</td>
<td>non-random sample</td>
</tr>
</tbody>
</table>

sample design

Bias

Random sample

Undercoverage

Convenience sample

Voluntary response sample
Appendix C

Concept Map 1

Name_____________________

Create a Concept Map below using the given focus question and concepts.

Focus question:
How does the way you choose people for your sample (sample design) relate to the inferences you make from your data?

Concepts:

<table>
<thead>
<tr>
<th>random digit dialing</th>
<th>do not call list</th>
</tr>
</thead>
<tbody>
<tr>
<td>convenience sample</td>
<td>under-coverage</td>
</tr>
<tr>
<td>random sample</td>
<td>chance</td>
</tr>
<tr>
<td>voluntary response sample</td>
<td>inferences</td>
</tr>
<tr>
<td>bias</td>
<td>non-random sample</td>
</tr>
</tbody>
</table>

sample design
Appendix D

Discourse Supports for Group Work

Question stems for group discussions

I see you connected _______ and _________. Can you explain how they are related?

Why did you connect _______ with _________ instead of _______?

Can you explain why you put _________ in that spot on your map?

I don’t understand where _________ might connect. Where did you put it and why?

I am having a hard time finding a linking word for _______ and ________. How do you see them as being related?

Is there another place that I could put _________? Why would that make sense?

I’m not sure about crosslinks for ______. Can you think of a way ______ might be connected with another concept?

I connected ________ and ________ because...

I like how you connected ______ and ______ because...

I am least certain about my connections for ______ and ______ because...

I really understand the connection between ______ and ______ because...
Create a Concept Map below using the given focus question and concepts.

Focus question:
How does the group you choose for your sample relate to bias?

Concepts:

<table>
<thead>
<tr>
<th>web survey</th>
<th>under-coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>wording effects</td>
<td>random sample</td>
</tr>
<tr>
<td>bias</td>
<td>mall interviews</td>
</tr>
<tr>
<td>non-random sample</td>
<td>sample surveys</td>
</tr>
<tr>
<td>population</td>
<td>non-response</td>
</tr>
</tbody>
</table>

Name__________________________

sample