The Impact of Formative Assessment Cycles on Students' Attitudes and Achievement in a Large-Enrollment Undergraduate Introductory Statistics Course

KimberLeigh Felix Hadfield
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd
Part of the Curriculum and Instruction Commons

Recommended Citation
Hadfield, KimberLeigh Felix, "The Impact of Formative Assessment Cycles on Students' Attitudes and Achievement in a Large-Enrollment Undergraduate Introductory Statistics Course" (2023). All Graduate Theses and Dissertations, Spring 1920 to Summer 2023. 8856.
https://digitalcommons.usu.edu/etd/8856

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations, Spring 1920 to Summer 2023 by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
THE IMPACT OF FORMATIVE ASSESSMENT CYCLES ON STUDENTS’ ATTITUDES AND ACHIEVEMENT IN A LARGE-ENROLLMENT UNDERGRADUATE INTRODUCTORY STATISTICS COURSE

by

KimberLeigh Felix Hadfield

A dissertation submitted in partial fulfilment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Education

Approved:

Katherine Vela, Ph.D. Patricia S. Moyer-Packenham, Ph.D.
Major Professor Committee Member

Jessica F. Shumway, Ph.D. Tyson S. Barrett, Ph.D.
Committee Member Committee Member

Kady Schneiter, Ph.D. D. Richard Cutler, Ph.D.
Committee Member Vice Provost for Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2023
ABSTRACT

The Impact of Formative Assessment Cycles on Students’ Attitudes and Achievement in a Large-Enrollment Undergraduate Introductory Statistics Course

by

KimberLeigh Felix Hadfield, Doctor of Philosophy
Utah State University, 2023

Major Professor: Katherine Vela, Ph.D.
Department: School of Teacher Education and Leadership

Although there is much research on the importance of formative assessments and feedback in education literature, implementing these formative assessment practices is largely missing in large-enrollment undergraduate introductory statistics courses. Additionally, large-enrollment introductory statistics courses tend to use few high-stakes tests rather than frequent formative assessment with reassessment opportunities. Therefore, this study aimed to investigate the impact of student attitudes toward statistics and student achievement after engaging in large-enrollment introductory statistics course curriculum using continuous formative assessments with feedback and reassessment opportunities. Using regression discontinuity to investigate course achievement in semesters with and without formative assessment cycles (FACs) in large-enrollment introductory statistics, meaningful differences in course achievement were evident, suggesting co-requisite courses with FACs could help students successfully navigate this
quantitative requirement. Using the SATS-36 pre- and post-survey scores from students in semesters of large-enrollment introductory statistics courses implementing FACs, students’ attitudes in affect, cognitive competence, and difficulty showed 4-5 times greater improvement than past studies on attitudes, with past studies showing 1-2 times greater decreases in effort, interest, and value than in this study. Students’ final grades significantly moderated these changes in attitudes from pre- to post-survey despite the students filling out the surveys 4 weeks before the end of the course. Students with higher final course grades showed greater improvements in attitudes that increased on average. For those attitudes which decreased on average, students with higher course achievement showed little to no declines. These findings suggest that large-enrollment introductory statistics courses implementing FACs can improve student achievement which moderates attitudes, improving students’ enjoyment of the course, beliefs of their computational abilities, and feelings that the course is less complicated or difficult than at the beginning of the course.

These findings have the potential to contribute towards more intentional policy development for statistics programs, better course designs, and additional pathways for student success in these courses. As students experience increased opportunities in introductory statistics through self-assessment from formative feedback and repeating assessment opportunities, FACs could improve students’ attitudes towards statistics and increase student achievement, preparing a new generation of statistically literate citizens for a data-driven world.
The Impact of Formative Assessment Cycles on Students’ Attitudes and Achievement in a Large-Enrollment Undergraduate Introductory Statistics Course

KimberLeigh Felix Hadfield

This study aimed to investigate the impact of student attitudes toward statistics and student achievement after engaging in large-enrollment introductory statistics course curriculum using continuous formative assessments with feedback and reassessment opportunities. This framework, called Formative Assessment Cycles (FACs) was implanted, providing students formative assessments both in and out of the classroom, with feedback and reassessment. A quasi-experimental, quantitative research design allowed for the investigation of course achievement from pre-FACs to FACs semesters using regression discontinuity methodology. Changes in attitudes from pre- to post-survey in semesters using a curriculum with FACs were analyzed by multilevel regression techniques. Course achievement improved in the co-requisite introductory statistics course using FACs for those who have less mathematical knowledge, suggesting the need for co-requisite courses and formative feedback and reassessment to provide students successful pathways to achieve their quantitative literacy requirement. Additionally, students with higher course achievement had significantly better attitudes towards statistics than their peers with lower final course grades. These students experienced more appreciation for the course and the science of statistics in their field of study, improved feelings of competence to do statistical calculations, and believed the
course was less confusing and easier than they first believed at pre-survey. These attitudes exhibited in this study were higher than previous studies on students’ attitudes toward statistics, suggesting that students who have opportunities to learn from their mistakes enjoy their introductory statistics class better and feel empowered by their newfound understanding.
ACKNOWLEDGMENTS

First, I must thank my committee chair, Dr. Vela, for the amazing input, feedback, assistance, and time spent reading my edits, who took me on less than a year ago, and who has shouldered the weight of the most important part of this process, thank you. Dr. MacDonald, thank you for teaching me to write a published literature review and a dissertation proposal, and helping me find my conceptual framework for my study. Dr. Moss, thank you for teaching and leading me to my first publication. Dr. Moyer-Packenham, you pushed me through this program with expectations of excellence. I loved learning from you. Thank you for molding me into a research writer. Dr. Shumway, thank you for starting me out on the right foot with my very first doctoral class. Your patient feedback to this novice writer after 25 years since getting my master’s degree was the gas I needed to get going. Dr. Schneiter, how lucky I am to be your colleague. I am so grateful for your constant encouragement. Dr. Schwartz, your kindness and patience helped me through my coding issues and statistical understanding. Thank you. And, Dr. Barrett, thank you for saying the words “regression discontinuity” which had been brewing in my mind since my first year of graduate studies. Thank you so much, Dr. Barrett, for your compassion and encouragement throughout this process. You all have left an indelible mark on my future scholarship and service on student committees.

Thank you to my cohort who have academically travelled the past five years with me and who have become dear friends. Amy, Carrie, Danielle, Joey, Kent, Will, Allison, and finally, my writing buddy and my shoulder to cry on, Lise. Thank you. Thank you. Thank you.
I am so grateful for my colleagues in the Mathematics and Statistics Department, who all were so supportive of me taking on this goal to complete my education. Your patience, your help, and your encouragement have been priceless. Adele and Jenny, thank you two, most of all. You two carried me when I thought I was drowning. I couldn’t have survived without you both. Hannah, the conversation we had where you helped me develop the Formative Assessment Cycle was monumental. Thank you.

I have so many friends, true friends, who kept inviting me to parties and brunches, knowing full well I would say no. You knew I would come through this—but thank you for sticking around despite my absence from your lives. Aimee—thank you for being my friend/therapist and treadmill buddy. Shelley—thank you for your cheerleading and love.

I have missed spending time with my family, immediate and extended, through the past five years. I haven’t been able to be there for my extended family—all the birthdays and celebrations, pinochle nights, dinners, and get-togethers. Yet, you never complained. None of you. Ever. Rather, you kept me going. Sydney, thanks for all the reels and texts to keep me laughing through this. You’ve truly been such a supportive niece and friend.

Mom and dad, seriously, how can I ever thank you for all your support? Almost 50 years of nothing but unconditional love, patience, and understanding. You two are truly the best, and everything I have accomplished in my life is because I had you two. You supported me, held me up, let me vent, and kept pushing me forward. You taught me to be a fierce, strong, independent but loving woman from day one. I love you so much. Thank you, times infinity.
Now, to my sweet boys. Emerson, I started this journey when you were just 9 years old. You have seen me cry a lot—because hard things are hard and exhausting. But I hope you have witnessed grit and perseverance from my example. Tennyson, who reminds me that “we aren’t average, we are excellent,” meaning we sacrifice to achieve our dreams, thank you for your example of pushing through your own trials in life, and finding that consistent effort leads to results. I am so proud of you, my sons. You two are my world. You both keep me laughing. You are the best.

Bretton, thank you for your partnership and love. Who knew that my best friend and crush in eighth grade would still be my best friend and crush 37 years later? Thank you for letting me pursue this crazy adventure. I could not have done this without your patience and support. You never once doubted my abilities, even though I constantly doubted myself. You told me what I was doing would make a difference for students. (Fingers crossed.) Thank you for everything. We have the best little family, that is the reality of my happiest dreams. Thank you, Mr. Dr. Hadfield, Esq. (Inside joke.)

I love you all. Thank you.

Finally, I dedicate this effort to my brother, Kaden Lloyd Felix, who passed away February 19, 2021, in the middle of this journey, and Becky Y. Hadfield, my angelic mother-in-law, who I know would be proud. Thank you, angels.

KimberLeigh Felix Hadfield
# CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>PUBLIC ABSTRACT</td>
<td>v</td>
</tr>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>xiii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xv</td>
</tr>
<tr>
<td>CHAPTER I: INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>Background of the Problem</td>
<td>3</td>
</tr>
<tr>
<td>Purpose of the Study</td>
<td>6</td>
</tr>
<tr>
<td>Significance of the Study</td>
<td>7</td>
</tr>
<tr>
<td>Research Questions</td>
<td>8</td>
</tr>
<tr>
<td>Summary of Research Design</td>
<td>8</td>
</tr>
<tr>
<td>Summary</td>
<td>9</td>
</tr>
<tr>
<td>Definition of Terms</td>
<td>9</td>
</tr>
<tr>
<td>CHAPTER II: LITERATURE REVIEW</td>
<td>12</td>
</tr>
<tr>
<td>Introduction</td>
<td>12</td>
</tr>
<tr>
<td>Formative Assessment and Student Attitudes Toward Statistics and Student Achievement: A Conceptual Framework</td>
<td>21</td>
</tr>
<tr>
<td>Formative Assessment Cycle for Large-Enrollment Courses</td>
<td>29</td>
</tr>
<tr>
<td>Three Elements of the Formative Assessment Cycle and Student Attitudes and Achievement</td>
<td>31</td>
</tr>
<tr>
<td>Conclusion</td>
<td>41</td>
</tr>
<tr>
<td>CHAPTER III: METHODOLOGY</td>
<td>44</td>
</tr>
<tr>
<td>Overview</td>
<td>44</td>
</tr>
<tr>
<td>Research Questions and Hypotheses</td>
<td>44</td>
</tr>
<tr>
<td>Research Design</td>
<td>45</td>
</tr>
<tr>
<td>Participants and Setting</td>
<td>48</td>
</tr>
<tr>
<td>Existing Data Set</td>
<td>50</td>
</tr>
<tr>
<td>Instruments and Data Sources</td>
<td>51</td>
</tr>
<tr>
<td>Data Analysis</td>
<td>58</td>
</tr>
<tr>
<td>Section</td>
<td>Page</td>
</tr>
<tr>
<td>------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>Summary</td>
<td>69</td>
</tr>
<tr>
<td>CHAPTER IV: RESULTS</td>
<td>71</td>
</tr>
<tr>
<td>Analysis for Research Question 1</td>
<td>72</td>
</tr>
<tr>
<td>Analysis for Research Question 2</td>
<td>90</td>
</tr>
<tr>
<td>Summary</td>
<td>115</td>
</tr>
<tr>
<td>CHAPTER V: DISCUSSION</td>
<td>116</td>
</tr>
<tr>
<td>Overview of the Study</td>
<td>116</td>
</tr>
<tr>
<td>Discussion of the Analysis of Research Question 1</td>
<td>117</td>
</tr>
<tr>
<td>Discussion of the Analysis of Research Question 2</td>
<td>123</td>
</tr>
<tr>
<td>Limitations</td>
<td>138</td>
</tr>
<tr>
<td>Generalizability</td>
<td>141</td>
</tr>
<tr>
<td>Implementing the Formative Assessment Cycle in Large-Enrollment Introductory Statistics Courses</td>
<td>142</td>
</tr>
<tr>
<td>Future Research</td>
<td>145</td>
</tr>
<tr>
<td>Conclusions</td>
<td>146</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>150</td>
</tr>
<tr>
<td>APPENDICES</td>
<td>161</td>
</tr>
<tr>
<td>Appendix A: Data Sources, Deidentified for Proposed Study</td>
<td>162</td>
</tr>
<tr>
<td>Appendix B: Permission to Use SATS-36</td>
<td>163</td>
</tr>
<tr>
<td>Appendix C: SATS-36 Pre-Survey</td>
<td>164</td>
</tr>
<tr>
<td>Appendix D: SATS-36 Post-Survey</td>
<td>169</td>
</tr>
<tr>
<td>Appendix E: Additional Questions Asked on SATS-36 Pre- and Post-Surveys</td>
<td>174</td>
</tr>
<tr>
<td>Appendix F: Introductory Statistics Course Objectives</td>
<td>176</td>
</tr>
<tr>
<td>Appendix G: Learning Mastery Gradebook Examples: Teacher and Student View</td>
<td>184</td>
</tr>
<tr>
<td>Appendix H: Units Covered in the Stat 1040/1045 Courses</td>
<td>186</td>
</tr>
<tr>
<td>Appendix I: Numerical and Graphical Summaries with Crossovers Removed</td>
<td>188</td>
</tr>
<tr>
<td>Appendix J: Numerical and Graphical Summaries with Extreme Cases Removed</td>
<td>191</td>
</tr>
<tr>
<td>Appendix K: Numerical and Graphical Summaries with Crossovers and Extreme Cases Removed</td>
<td>194</td>
</tr>
<tr>
<td>Appendix L: Simple Slopes Analysis</td>
<td>197</td>
</tr>
<tr>
<td>Appendix M: LOESS Curve on Model 1</td>
<td>198</td>
</tr>
</tbody>
</table>
Appendix N: Side by Side Plots of Final Grade Percentage at the Cutoff, Pre-FACs to FACs Semester .......................................................... 199
Appendix O: Descriptive and Summary Statistics by Prerequisite .......... 200
Appendix P: Descriptive and Summary Statistics of Students who Took the ALEKS Math Placement Exam ........................................ 201
Appendix Q: Additional Descriptive and Summary Statistics ................. 203
Appendix R: Summary Statistics of All Students by Recitation TA ......... 204
Appendix S: Histograms of Attitude Components .................................. 206
Appendix T: Person-Profile Violin Plots by Course for Attitude Components ......................................................................................... 212
Appendix U: Estimated Marginal Means for Female Students’ Changes in Attitude Components from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95% ........................................ 217
Appendix V: Estimated Marginal Mean Plots by Sex on Attitude Components Change from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95% ........................................ 218
Appendix W: MLM Models for ALEKS Subset by Attitude Component .... 220
Appendix X: Permission to Reprint from Statistics Education Research Journal ......................................................................................... 222

CURRICULUM VITAE ..................................................................................................................... 223
LIST OF TABLES

Table 1  Research Questions, Data Source and Instruments, Participants, and Analysis ........................................................................................................... 46

Table 2  Pre-FACs and FACs Cohorts for Proposed Research Question 1 ........... 47

Table 3  Introductory Statistics Courses ................................................................... 48

Table 4  Number of Participants by Research Question ........................................... 50

Table 5  SATS-36 Attitude Components, Definitions, Examples, and Number of Items Per Component with Cronbach Alpha Ranges per Component ... 52

Table 6  Descriptive and Summary Statistics of Participants for Research Question 1 ................................................................................................... 74

Table 7  Regression Estimates for Baseline Models at the Cutoff............................ 78

Table 8  Models Fit to EQ(full) ................................................................................ 80

Table 9  Three-Way Interaction Models ................................................................... 81

Table 10 Standardized Mean Differences and Pairwise t Tests for Change in Final Grade Before and After the Cutoff (Less than 30 – At least 30) for all Four Models ................................................................................................. 85

Table 11 Standardized Mean Differences (Effect Sizes) and Pairwise t Tests for the Difference in Final Grade Change from Pre-FACs to FACs (Pre-FACs – FACs) for all Four Models ......................................................... 88

Table 12 Students Participation in the SATS-36 Survey: Total and After Removing Student Responses Who Withdrew or Failed............................ 92

Table 13 Descriptive and Summary Statistics of the Participation in the Surveys.... 93

Table 14 Cronbach’s Alphas for the SATS-36 Attitude Survey by Attitude Component ........................................................................................................ 94

Table 15 Summary of Dependent Variables ................................................................ 102
Table 16  Baseline Models for Multilevel Models for Attitude Components .......... 103
Table 17  Final Multilevel Models for Attitude Components................................. 106
Table 18  Standardized Mean Differences (Effect Sizes) and Pairwise $t$ tests for Change in Pre- to Post-Survey Attitudes by Final Grade .................................. 110
Table I.1  Descriptive and Summary Statistics of Participants for Research Question 1, Crossovers Removed ................................................................. 188
Table J.1  Descriptive and Summary Statistics of Participants for Research Question 1, Extreme Cases Removed ............................................................. 191
Table K.1  Descriptive and Summary Statistics of Participants for Research Question 1, Crossovers and Extreme Cases Removed ........................................ 194
Table L.1  Simple Slopes Analysis ........................................................................... 197
Table O.1  Descriptive and Summary Statistics by Prerequisite ............................... 200
Table P.1  Descriptive and Summary Statistics of Students who Took the ALEKS Math Placement Exam ............................................................................. 202
Table Q.1  Descriptive and Summary Statistics of Participants by Major ................. 203
Table R.1  Student Participation by Recitation Teacher (TA)...................................... 204
Table R.2  Participants in the Analysis by Recitation TA ........................................... 205
Table W.1  MLM Models for ALEKS Subset by Attitude Component ..................... 221
| Figure 1 | Situating Formative Assessment in Theoretical Perspectives in Higher Education | 20 |
| Figure 2 | Conceptual Framework for Formative Assessment Cycles in Large-Enrollment Introductory Statistics Courses | 22 |
| Figure 3 | Assessment Continuum | 24 |
| Figure 4 | Simulated Data Representing Observed Data Points Along a Running Variable Below and Above Some Binding Cutoff | 61 |
| Figure 5 | Graphical Depiction of the Nested Nature of the Data for Research Question 1 | 62 |
| Figure 6 | Nesting Structure for Analyzing Attitudes as a Repeated Measure for Research Question 2 | 67 |
| Figure 7 | Histograms of ALEKS Placement Scores by Course | 75 |
| Figure 8 | Histogram of the Final Grade Percentages by Course | 76 |
| Figure 9 | Estimated Means of Final Grade Percentage by ALEKS scores Around the Cutoff, Pre-FACs and FACs | 82 |
| Figure 10 | Side by Side Plots of Final Grade Percentage at the Cutoff, Pre-FACs to FACs Semester for Model 4 | 87 |
| Figure 11 | Estimated Means of Final Grade Percentage by ALEKS Scores Around the Cutoff for Model 4 | 89 |
| Figure 12 | Affect Person-Profile Plot by TA | 96 |
| Figure 13 | Cognitive Competence Person-Profile Plot by TA | 97 |
| Figure 14 | Difficulty Person-Profile Plot by TA | 98 |
| Figure 15 | Value Person-Profile Plot by TA | 99 |
| Figure 16 | Interest Person-Profile Plot by TA | 100 |
Figure 17  Effort Person-Profile Plot by TA ............................................................. 101

Figure 18  Person-Profile Violin Plot by Course for Cognitive Competence ........... 104

Figure 19  Estimated Marginal Means for Males’ Changes in Attitude Components from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95% ................................................................................. 107

Figure G.1  Teacher View of the LMG for the Course ............................................... 184

Figure G.2  Test Student View of the LMG ................................................................ 185

Figure I.1  Histogram of ALEKS Scores in Pre-FACs and FACs Semesters, Crossovers Removed ................................................................. 189

Figure I.2  Histogram of Final Grades in Pre-FACs and FACs Semesters, Crossovers Removed ................................................................. 190

Figure J.1  Histogram of ALEKS Scores in Pre-FACs and FACs Semesters, Extreme Cases Removed ................................................................. 192

Figure J.2  Histogram of Final Grades in Pre-FACs and FACs Semesters, Crossovers Removed ................................................................. 193

Figure K.1  Histogram of ALEKS Scores in Pre-FACs and FACs Semesters, Crossovers and Extreme Cases Removed ................................................................. 195

Figure K.2  Histogram of Final Grades in Pre-FACs and FACs Semesters for Students with Crossovers and Extreme Cases Removed ......................... 196

Figure M.1  LOESS Curve on Model 1 ....................................................................... 198

Figure N.1  Side by Side Plots of Final Grade Percentage at the Cutoff, Pre-FACs to FACs Semester ................................................................. 199

Figure S.1  Histograms of Students’ Pre-and Post-Survey Affect Scores .................... 206

Figure S.2  Histograms of Students’ Pre-and Post-Survey Cognitive Competence Scores ................................................................. 207

Figure S.3  Histograms of Students’ Pre-and Post-Survey Difficulty Scores .......... 208
Figure S.4  Histograms of Students’ Pre-and Post-Survey Effort Scores ................. 209
Figure S.5  Histograms of Students’ Pre-and Post-Survey Interest Scores .............. 210
Figure S.6  Histograms of Students’ Pre-and Post-Survey Value Scores ................. 211
Figure T.1  Person-Profile Violin Plot by Course for Affect.................................. 212
Figure T.2  Person-Profile Violin Plot by Course for Difficulty ............................ 213
Figure T.3  Person-Profile Violin Plot by Course for Effort ................................... 214
Figure T.4  Person-Profile Violin Plot by Course for Interest ............................... 215
Figure T.5  Person-Profile Violin Plot by Course for Value ................................. 216
Figure U.1  Estimated Marginal Means for Female Students’ Changes in Attitude Components from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95% .......................................................... 217
Figure V.1  Estimated Marginal Mean Plots by Sex on Attitude Components Change from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95% ........................................................................... 219
CHAPTER I
INTRODUCTION

“...given the present strength of the evidence for the effectiveness of formative assessment, or assessment for learning, it is somehow surprising that the implementation of better classroom practices has not been more evident.”
(Hopfenbeck, 2018, p. 548, paraphrasing D. Wiliam, 2018)

The American Statistical Association (ASA) and the Mathematical Association of America (MAA) delineate several recommendations for improving assessments in undergraduate introductory statistics courses. One recommendation by the ASA in the Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report (American Statistical Association Revision Committee, 2016) states that assessments should be used both formatively to improve learning and summatively to evaluate learning continually. Formative assessments are those that inform the teacher as to their teaching and instructional practices and inform the student of where they are in their understanding by using feedback from the assessment process (Black & Wiliam, 1998; Cowie & Bell, 1999; Ghaicha, 2016; Harlen, 2012; Shute, 2008). In contrast, evidence obtained from summative assessments provides judgments about student understanding and reports on the achievement with no cycle of feeding back (Harlen, 2012). Both the ASA and the MAA have published recommendations and guidelines to incorporate formative and summative assessments in the undergraduate introductory statistics curriculum.

The GAISE College Report (ASA, 2016) stresses the need for introductory

---

1 Portions and sections of Chapters I, II, and V have been submitted and published in the Statistics Education Research Journal, 2023, vol. 22 (see Appendix X for permission to reprint in part or whole).
statistics courses to provide feedback to students regarding their learning by utilizing frequent formative assessments. In particular, the *MAA Instructional Practices Guide* (Abell et al., 2018) provides vignettes with examples of different formative assessment strategies for instructors and course designers to implement in order to improve curriculum and instruction in undergraduate quantitative courses. This guide also emphasizes the need for evidence-based assessment practices in large-enrollment courses to enhance various cognitive and performance-based student outcomes (Abell et al., 2018). Additionally, several principles for assessments outlined in the *MAA Instructional Practices Guide* echo the *GAISE College Report*. These reports recommend that courses integrate assessments, not as stand-alone events, but rather as a “continuous cycle” of assessment throughout the course (Abell et al., 2018; ASA Committee, 2016). The *GAISE College Report* and *Instructional Practices Guide* call for updating assessments in quantitative courses to motivate student learning through frequent formative assessments with feedback rather than the heavy focus of summative examinations.

Despite the calls for improved assessment practices from the ASA and MAA, transforming assessment practices has been slow to be implemented in large-enrollment undergraduate introductory statistics courses due to the volume of students and the time-intensive nature of the assessment process. For instance, a large study of university courses found that students noticed a stark difference in the types of assessments assigned in large- and small-enrollment classes (Cash et al., 2017). Unsurprisingly, the findings revealed that assessments used in large-enrollment courses were less frequent and did not often vary in their design; rather, they utilized summative assessments (Cash et al., 2017).
In fact, summative high-stakes assessments accounted for more than 95% of the assessments in large-enrollment classes (Cash et al., 2017). In addition, the findings suggested the students’ assessments comprised of one to two midterm exams and one final exam in these large courses, a glaring departure from the MAA and ASA recommendations to utilize a continuous cycle of formative assessment with ongoing feedback (Abell et al., 2018; ASA Committee, 2016). Given this, the focus of this study was to investigate the impact of formative assessment practices in large-enrollment introductory statistics courses to improve student achievement and student attitudes towards statistics, providing students a pathway to complete their quantitative literacy requirement successfully.

**Background of the Problem**

Introductory statistics is one of the most critical quantitative courses in a student’s university experience. Moreover, the 21st century requires students to critically navigate statistical information in a data-driven world (Rumsey, 2002; Tishkovskaya & Lancaster, 2012). Additionally, the introductory statistics course is quickly becoming the course that satisfies the quantitative literacy requirement in higher education, replacing college algebra (Hoang et al., 2017). The increased enrollment of undergraduate students in introductory statistics courses brings diverse student interests, majors, and mathematical background knowledge (ASA Revision Committee, 2016; Blair et al., 2018). These students often do not major in science, technology, engineering, or mathematics (STEM) fields, and the introductory statistics course becomes their only quantitative literacy
course in their program of study. Due to this growth and diversity of students, Fong et al. (2015) found that students enrolled in introductory mathematics courses lack the fundamental mathematical knowledge needed for mathematical thinking. This deficiency in mathematical knowledge results in a financial burden for many students as they attempt multiple semesters to successfully pass their introductory statistics class. These multiple attempts create a “bottleneck” by slowing the students’ path toward further studies (Complete College America, 2012; Peck, 2019). To further exacerbate these problems, introductory statistics courses tend to be large-enrollment courses in large universities, disallowing students individualized academic attention (Blair et al., 2018; Cash et al., 2017). Additionally, research on class size suggests that as a developmental mathematics class size increases, the probability of students successfully passing decreases (Fong et al., 2015). Instructors of large-enrollment introductory statistics courses find it nearly impossible to provide individualized student attention and feedback on students’ assessments with the added challenge of attending to the students’ different mathematical preparedness for statistics (Cash et al., 2017). As universities address the increased enrollment and demand for introductory statistics courses, large-enrollment courses become the answer.

Students’ challenges in navigating large-enrollment statistics courses impact their experiences, affect whether they value statistics, and influence their attitudes toward statistics throughout their adult lives (Ramirez et al., 2012; Tichkovskaya & Lancaster, 2012). Evidence suggests that students who are not STEM majors experience greater statistics anxieties and tend to avoid courses in statistics, which negatively affects their
attitudes toward statistics and their achievement in the course (Chew & Dillon, 2014; Chiesi & Primi, 2010; Lavidas et al., 2020; Onwuegbuzie & Wilson, 2003; Williams, 2015). These negative attitudes toward statistics stick with students long after they experience their introductory course, affecting their motivation and appreciation for statistical literacy (Ramirez et al., 2012; Xu & Schau, 2019). Additionally, negative attitudes are associated with decreased achievement (Chiesi & Primi, 2010; Emmioğlu & Capa-Aydin, 2012). These negative attitudes, linked to testing and assessment performance in students’ introductory statistics courses, can be attributed to their lack of mathematical understanding and preparation from prior mathematics courses (Chiesi & Primi, 2010; Harlow et al., 2002; Malik, 2015; Onwuegbuzie & Wilson, 2003). With the large-enrollment course designs, high degrees of negative attitudes, and use of high-stakes examinations with minimal, if any, feedback provided, many students struggle to be successful in these statistics courses (Cash et al., 2017; Onwuegbuzie & Wilson, 2003; Peck, 2019). Where large-enrollment classes answer the need to support increased enrollment and the aforementioned bottleneck by many non-STEM students, many of these statistics courses aggravate students’ negative attitudes toward statistics, affecting their achievement.

Conversely, teachers using innovative assessments in higher education courses can better influence their students’ attitudes, persistence, and achievement in introductory statistics (Abell et al., 2018; ASA Revision Committee, 2016). Specifically, students’ positive attitudes toward statistics are associated with students’ self-reported degrees of motivation and higher achievement scores (Chiesi & Primi, 2010; Ramirez et al., 2012).
Furthermore, when classrooms use formative assessment with feedback focused on learning rather than performance, students experience a change in their mindset and relationship with their learning (Boaler & Confer, 2017). Finally, recent research asserts that a “comprehensive approach to designing a successful statistics pathway” is needed to provide support structures in introductory statistics for underprepared students to “effectively complete their college-level statistics course” (Peck, 2019, p. 35). Heeding this call, the research purports that introductory statistics courses utilizing formative assessments with feedback can improve the important student outcomes of attitudes toward statistics and achievement.

**Purpose of the Study**

Although there is much research on the importance of formative assessments and feedback in education literature, implementing these formative assessment practices is absent in most large-enrollment undergraduate introductory statistics courses. Large-enrollment introductory statistics courses tend to use few high-stakes tests rather than the recommended frequent formative assessments with feedback. The many challenges of large-enrollment courses, including less individualized attention and feedback, burden students financially and academically as repeated attempts at the course may be required for successful completion. As enrollments in introductory statistics continue to climb, pathways for successful completion of those courses must be prioritized (Peck, 2019). Therefore, the purpose of the study was to investigate the impact of an embedded cycle of formative assessment with feedback and reassessment opportunities in the curriculum of
large-enrollment introductory statistics courses on student attitudes toward statistics and student achievement scores.

**Significance of the Study**

The effects of this study have far-reaching implications for mathematics and statistics departments, instructors, and students. This study provided important information as a feasibility study for a future large-scale endeavor. Regression discontinuity served as a viable way to determine the impact of formative assessment cycles (FACs) on course achievement. Additionally, this study found that course achievement moderated student attitudes toward statistics. These findings can significantly contribute to mathematics and statistics departments’ introductory statistics curriculum design. Departments must allocate resources to assist instructors and curriculum creators of these large-enrollment courses to transform their assessments with smaller, more frequent formative assessment practices, with feedback and reassessment opportunities. Creating corequisite courses can benefit students of different mathematical backgrounds. Moreover, with future research, introductory statistics courses will readily employ FACs in the curriculum. The findings of this study suggested that FACs assisted students with successful completion of their quantitative literacy requirement, especially students with lower math placement scores. Additionally, this study suggested that students’ attitudes toward statistics improved with greater student achievement, empowering non-STEM majors to find STEM fields more accessible. Thus, these findings have the potential to contribute toward more intentional policy development for
statistics programs, better course designs in statistics, and additional pathways for students to succeed who are often disempowered in STEM fields.

**Research Questions**

This study sought to answer the following two research questions regarding the effects of FACs on student attitudes toward statistics and statistics achievement.

**Research Question 1**

How do formative assessment cycles (FACs) affect student achievement in large-enrollment introductory statistics courses for different mathematically prepared students?

**Research Question 2**

After allowing for student-to-student variability, which student attitude components change after a semester of a large-enrollment introductory statistics course with FACs? Also, how do demographic factors impact attitude, and do these effects change over time?

**Summary of Research Design**

To quantitatively analyze the impact of FACs on students’ attitudes and achievement scores, this dissertation research used a quasi-experimental research design (Cresswell & Cresswell, 2018). To investigate these research questions, I utilized surveys, exam scores, course grades, and demographic information from existing student data to investigate changes in students’ attitudes and student achievement in large-
enrollment introductory statistics courses. A quasi-experimental study was most appropriate because it was impossible to randomize the participants into treatment and control groups (Scher et al., 2015). Through these approaches, this research design allowed for effective quantitative analysis of the research questions.

**Summary**

Because more and more students are choosing introductory statistics to fulfill a quantitative literacy requirement, student enrollments in introductory statistics courses are ever-increasing. Large class sizes are the norm for navigating both the student’s educational trajectory and for instructors to teach statistical thinking. As such, the need for students to successfully complete their introductory statistics course to mitigate the bottleneck of their educational progress is evermore paramount. In particular, assessment practices that benefit students, such as formative assessment and feedback, are lacking in large-enrollment courses, affecting student attitudes towards statistics and their achievement. Thus, this study investigated FACs and their impact on student attitudes and achievement, using a quasi-experimental research design to analyze the research questions. A list of key terms and definitions completes this chapter.

**Definition of Terms**

Terms that are key to this study are defined below.

**Assessment for learning**: “Any assessment for which the first priority in its design and practice is to serve the purpose of promoting students’ learning” (Black et al.,
Assessment of learning: Assessment is carried out only for the purposes of grading and reporting (Assessment Reform Group [ARG], 2002).

Corequisite course: A corequisite course concomitantly teaches the prerequisite knowledge needed for the current course to eliminate the need for an extra semester for course preparation (Complete College America, 2012).

Feedback: “Information communicated to the learner that is intended to modify their thinking or behavior to improve learning” (Shute, 2008, p. 154).

Formative assessment: Provides evidence about student achievement, which is obtained and utilized by the teacher and/or the student to make decisions about the next steps in the learning process, intending to improve and elucidate student understanding (Black & Wiliam, 2009).

Formative assessment cycles (FACs): A cyclical formative assessment process using formative assessments with immediate feedback for students’ self-assessment to learn from their mistakes, with reassessment opportunities offered for the student to show an increased understanding.

Introductory statistics (Stat 1040): Quantitative literacy course requirement for most non-scientific majors. The course covers descriptive and inferential statistical methods emphasizing conceptual understanding and statistical thinking. This is a 3-credit hour, one-semester course (https://catalog.usu.edu/).

Introductory statistics with algebra (Stat 1045): Co-requisite course integrating elements of algebra with Introductory Statistics Stat 1040 for students whose math
placement score is below what is required for Stat 1040. This is a 5-credit hour, one-semester course (https://catalog.usu.edu/).

**Large-enrollment:** The number of students in a class or course is greater than what one instructor can teach on one’s own, requiring multiple smaller recitation or lab sections to facilitate both student and teaching needs (Hornsby & Osman, 2014).

**Summative assessment:** The evidence obtained from summative assessments are judgments about the achievement in order to report on the achievement with no cycle of feeding back to instruction (Harlen, 2012). Most summative assessments are considered Assessments of Learning, but when coupled with corrective feedback can be used by students as Assessments for Learning (Harlen, 2012).

**Reassessment:** When learning tasks are assigned to a learning criterion, multiple attempts at learning from those criteria are allowed to gain an understanding about one’s progress through a curriculum (Abell et al., 2018).
CHAPTER II
LITERATURE REVIEW

“Assessment is operationally defined as part of the educational process where instructors appraise student achievements by collecting, measuring, analyzing, synthesizing and interpreting relevant information...under controlled conditions in relation to curricula objectives set for their levels” (Ghaicha, 2016, p. 211).

Introduction

Recommendations for improving assessments are detailed in reports from both the ASA and the MAA. For instance, the *GAISE College Report* (ASA Committee, 2016) explicitly addressed how instructors should employ assessments in introductory statistics courses. Specifically, to “[u]se assessments to improve and evaluate student learning” (p. 3), the *GAISE* authors stressed that (a) students should receive timely feedback throughout the course, (b) assessments should align with learning outcomes, and (c) instructors should maximize the use of varying types of formative assessments in addition to summative examinations. *Summative assessments* are designed to judge achievement with no cycle of feedback to instruction (Harlan, 2012). Summative assessments are usually considered high-stakes exams because they contribute to a large portion of the overall grade, such as midterm and final exams. In contrast, *formative assessments* are lower-stakes assessments that provide evidence about achievement through feedback utilized by the teacher or the student to inform the learning process and subsequent actions (Black & Wiliam, 2009; Harlan, 2012). These reports from the ASA and MAA described guidelines and recommendations for undergraduate quantitative courses for
formative assessment practices.

The MAA *Instructional Practices Guide* (Abell et al., 2018) provided an assessment framework, gleaning from research about the benefits of evidence-based assessment practices, ASA recommendations, and the National Council of Teachers of Mathematics (NCTM) standards (ASA Committee, 2016; Gold et al., 1999; NCTM, 2000; Steen, 2006). The assessment framework is based on the following six principles.

1. Assessment is not a single event but a continuous cycle.
2. Assessment must be an open process.
3. Assessment must promote valid inferences.
4. Assessment that matters should always employ multiple measures of performance.
5. Assessment should measure what is worth learning, not just what is easy to measure.
6. Assessment should support every student’s opportunity to learn important mathematics (Steen, 1999, p. 5-6).

Several of MAA’s principles for assessments overlapped with the *GAISE College Report’s* (ASA Committee, 2016) recommendations, specifically that assessments are not singular events but should be considered cycles, where feedback cycles back to the teacher and student throughout the course (Steen, 1999). Additionally, the MAA *Instructional Practices Guide* stressed that teachers and students must use learning goals and course objectives in the feedback process (Abell et al., 2018). The MAA *Instructional Practices Guide* devoted a section to assessment in large-enrollment courses, suggesting online response systems, online homework systems, and the use of technology to support instructors in providing timely feedback to students (Abell et al.,
Research must address how instructors can take these recommendations to create meaningful assessments in large-enrollment courses.

With the focus on contributing to the field of statistics education and heeding the call to create successful student pathways to complete the introductory statistics course, the purpose of this literature review is to examine formative assessment concepts and their relationships in the literature to student attitudes and achievement. Before presenting the conceptual framework, this introduction continues with a broad synthesis of the research on student attitudes toward statistics and student achievement in statistics. I introduce the important variable of mathematical preparedness, providing frameworks in the research where this variable was significantly associated with student attitudes toward statistics and student achievement. Then, I provide the theoretical context for the conceptual framework, in which the lens of andragogy informs formative assessment theory in higher education. The second section of the chapter presents the conceptual framework for formative assessment cycles (FACs). I clarify the importance of the three main elements of the FAC in formative assessment literature: formative assessments as Assessments for Learning, feedback and self-assessment, and reassessment. The third section of the chapter summarizes the statistics education research on utilizing technology and computer-based assessments to import the FAC into large-enrollment introductory statistics courses. Finally, the fourth section expounds on the relations between the three formative assessment elements and students’ attitudes and achievement from the literature. In conclusion, I discuss how the conceptual framework informs the proposed research as a curricular intervention to impact student attitudes toward statistics
and achievement in the large-enrollment introductory statistics course.

**The Association Between Student Attitudes Toward Statistics and Student Achievement**

Research has confirmed the existence of associations between student attitudes toward statistics and their achievement in undergraduate introductory statistics courses. Several studies have demonstrated a positive association between students’ attitudes and achievement in introductory statistics (Chiesi & Primi, 2010; Emmioğlu & Capa-Aydin, 2012; Evans, 2007; Maure & Marimon, 2014; Ramirez et al., 2012; Schau, 2003). Additionally, the research suggested that high levels of statistics anxiety were associated with negative attitudes toward statistics (Chiesi & Primi, 2010, Kesici et al., 2011; Malik, 2015; Slootmaeckers et al., 2014; Williams, 2015). These high levels of statistics anxiety were also associated with decreased student achievement in introductory statistics courses (Chew & Dillon, 2014; Chiesi & Primi, 2010). Specifically, students’ self-concept and perceived worth of statistics predicted their final exam scores in online and face-to-face large student sections of introductory statistics courses (Zimmerman & Austin, 2020). Moreover, high levels of anxiety evidenced an inverse relationship with attitudes, which negatively affected achievement (Chiesi & Primi, 2010). These associations demonstrate an important connection between students’ attitudes and their achievement in introductory statistics.

The authors of the studies on attitudes and achievement provided suggestions and recommendations for further research to explore ways to improve student attitudes toward statistics due to the connections between student attitudes and achievement.
Chiesi and Primi (2010) encouraged implementing interventions like formative assessments to lower anxiety and increase positive attitudes in introductory statistics students with various mathematical preparedness. Malik (2015), in a phenomenological study of undergraduate introductory statistics, interviewed students and found that students had high levels of statistics anxiety concerning their being assessed in statistics. These students overwhelmingly stated that testing situations caused the most significant anxiety response. As per the ASA and MAA recommendations, using formative assessments with feedback in a lower-stakes environment is a possible pedagogical intervention that could decrease anxiety for students, improve attitudes, and lead to greater student achievement (Abell et al., 2018; ASA Revision Committee, 2016; Ramirez et al., 2012). Given these findings, it is apparent that students’ attitudes can be a gateway or a gatekeeper to achievement in statistics or other quantitative courses, especially in less mathematically prepared students.

Using the Survey of Attitudes Toward Statistics (SATS), Schau (2003) found students’ attitudes toward statistics negatively changed over the course; however, positive attitudes correlated with increased achievement. Thus, Schau believed that changes in attitudes impacted students’ achievement in introductory statistics. This association was further confirmed by an analysis completed by Ramirez et al. (2012). The authors found 17 studies that evaluated the relationships between SATS attitudes and achievement, where 15 had significant positive associations. A meta-analysis of the research on student attitudes using the SATS instrument and student achievement conducted by Emmioğlu and Capa-Aydin (2012) looked at effect sizes and components of the SATS instrument
across different countries. Across all countries, attitudes toward statistics and students’ achievement were positively associated; however, larger effect sizes were evidenced in studies on United States’ students regarding the two components of the SATS: *affect* and *cognitive competence*. Thus, research has demonstrated the effectiveness of measuring student attitudes toward statistics using the SATS instrument and its association with achievement.

**The Associations Between Mathematics Preparedness and Student Attitudes Toward Statistics and Student Achievement**

Measuring mathematical preparedness for introductory statistics differs in the research, such as a mathematics placement exam, a survey of prior mathematics courses, or a mathematics assessment given prior to class to measure a student’s mathematical knowledge. All research presented in this section utilized correlational studies to measure a student’s mathematical understanding in these various ways before the introductory course and evaluated the students’ preparedness with course attitudes and achievement.

Several studies developed frameworks relating students’ mathematical preparedness prior to introductory statistics and student attitudes as measured by the SATS instrument and student achievement. Schau (2003) identified that students’ prior achievement in mathematics influenced students’ course achievement in introductory statistics: her model related prior mathematics achievement with pre-course attitudes and, subsequently, post-course attitudes. In addition to Schau, Ramirez et al. (2012) developed a model relating students’ prior achievement, attitude, and course outcomes from a meta-analysis. They identified ten studies that examined the relationship between past
mathematical experience and statistics attitudes. All ten studies showed significant positive associations between prior mathematical achievement and course attitudes. Those studies which utilized the SATS instruments to measure attitudes showed higher positive associations with the attitude components of *affect*, *cognitive competence*, and *difficulty* when the students had more prior experience with mathematics. Additionally, Ramirez et al. found six studies that examined the relationship between prior mathematics experience and course achievement, and all six showed a significant positive correlation. These studies confirm that prior student experience with mathematics is an important variable that impacts both student achievement and student attitudes toward statistics.

Several studies specifically measured students’ pre-course mathematical abilities to determine the effect on their attitudes toward introductory statistics and achievement. In a large-scale study in Italy, Chiesi and Primi (2010) utilized a pre- and posttest design measuring introductory statistics students’ (80% female) anxiety, attitudes using the SATS instrument, and achievement. A unique aspect of this study was the varied experiences and degrees of mathematical preparedness and statistics experience these students had before their enrollment. Findings indicated that students who were less mathematically prepared had less confidence, more negative attitudes, and thought statistics was relatively more complicated than students who were more mathematically prepared. At the end of the course, both the more and less mathematically prepared students’ attitudes improved, but the less prepared students had greater self-confidence and felt that statistics had value (Chiesi & Primi, 2010). Using structural equation
modeling, Harlow et al. (2002) found a significant link between the students’ pre-course mathematical ability and their post-course achievement: 38% of the variation in post-course achievement was due to the students’ mathematical preparedness. Additionally, students' pre-course attitudes significantly predicted their mathematics pre-course ability (Harlow, 2002). Taken together, these studies provide important evidence that current research must measure the students’ prior mathematical knowledge as it relates to students’ attitudes and achievement in introductory statistics.

Theoretical Perspectives to Examine Formative Assessment in Higher Education

Andragogy is a theoretical perspective explaining how adult learners engage in their learning environments (Knowles, 1978). Viewing learning theories through the lens of andragogy provides unique insight and understanding as to how adult learners learn and process information. It posits five central tenets.

1. Adult learners are motivated and desire to learn.
2. Adult learners want to apply information to life situations directly.
3. Adults’ life experiences provide a valuable resource for their learning.
4. Adult learners are self-directed.
5. As age increases, differences across individuals are vast and contextualized (Knowles, 1978; Lindeman, 1926; Merriam, 2001).

For nearly a century, these tenets have critically informed adult education worldwide (Merriam, 2001). As stated by Knowles (1978), “adult education is an attempt to discover a new method and create a new incentive for learning” (p. 11). Additionally, andragogy is learner-centered, with educators given the charge to “involve learners in as many aspects
of their education as possible,” such that the educational climate fosters positive learning (Houle, 1996, p. 30). Using these tenets, higher education can more effectively value and address adult learners’ differences and unique abilities that they bring to their educational environments. Viewing learning theories through the lens of andragogy provides context for understanding undergraduate students and their experience in introductory statistics.

These learning theories include social-cultural theories to situate formative assessment in undergraduate education (see Figure 1). By taking up an andragogical lens, we can draw from these tenets of andragogy to leverage adult learners’ self-efficacy and self-regulation through assessment design in large-enrollment introductory statistics classes.

**Figure 1**

_Situating Formative Assessment in Theoretical Perspectives in Higher Education_
To consider relations between assessments and adults’ learning, self-efficacy is used to explain motivation with both constructs. Andragogy posits that adult learners are internally motivated and more self-directed in their learning (Knowles, 1978; Merriam, 2001; Sosibo, 2019). Self-efficacy theory regards the beliefs in one’s actions that can produce or thwart desired or damaging results (Bandura, 2001). Thus, without these beliefs and motivation, adult learners would have “little incentive to act or to persevere in the face of difficulties” (Bandura, 2001, p. 10). Additionally, self-regulation is the theory of using those beliefs and motivation to actively engage in learning. The actions of adult learners are rooted in their core beliefs that their efforts will produce the desired outcomes in their education (Bandura, 2001). Thus, if adult learners believe in their abilities to achieve their goals, their actions will stem from those beliefs (Mangels et al., 2006). Together, these theoretical perspectives lay the foundation for the conceptual framework of formative assessment cycles (FACs) and the outcomes of student attitudes toward statistics and student achievement.

**Formative Assessment and Student Attitudes Toward Statistics and Student Achievement: A Conceptual Framework**

The review of the assessment literature resulted in a three-pronged approach to formative assessment in large-enrollment introductory statistics courses: frequent lower-stakes assessments (i.e., Assessments for Learning), feedback, and reassessment opportunities. Additionally, the research on assessment in undergraduate quantitative literacy courses revealed empirical studies implementing at least one of the three-prongs
of formative assessment and measuring student attitudes and/or student achievement. Thus, the conceptual framework for the proposed study displays the three-prongs as a formative assessment cycle (FAC) embedded in the introductory statistics curriculum to improve both attitudes and achievement. The dotted blue arrows in Figure 2 indicate those relationships representing the research questions this study investigated.

Figure 2

Conceptual Framework for Formative Assessment Cycles in Large-Enrollment Introductory Statistics Courses

The formative assessment cycle employs the use of frequent low-stakes assessments, feedback, and reassessment. These three elements, embedded into a large-enrollment course, create the continuous, open assessments suggested by the MAA recommendations (Abell et al., 2018; Steen, 1999). First, by employing frequent, low-stakes assessments rather than a few high-stakes assessments, students are given more assessment opportunities instead of the few midterm exams which account for a large
portion of their grade. The MAA described assessment as more than a few high-stakes
tests, but rather, as a “wider set of measures,” in which varied assessments measure
students’ progress on learning outcomes (Abell et al., 2018, p. 50). Second, instructors
must connect timely feedback from the formative assessments to learning goals. The
learning goals allow students to self-assess their level of understanding and create the
next actions for improving their knowledge. Third, when students take advantage of
reassessment opportunities of the learning goals not yet mastered, students see their
learning and understanding improve. The literature explicated formative assessment as
Assessments for Learning, where students employed self-assessment by using feedback
to improve their understanding and used reassessment to reevaluate their knowledge.

Formative Assessments as Assessments for Learning

Assessment for Learning, defined by Black et al. (2004), is “any assessment for
which the first priority in its design and practice is to serve the purpose of promoting
students’ learning” (p. 10). The Assessment Research Group (2002) defines Assessment
of Learning as summative assessments that are only used for the purposes of grading and
reporting. Harlen (2012) created a framework for assessments which places assessments
on a continuum, ranging from formative to summative assessments and broadly as
Assessment for Learning to Assessment of Learning (see Figure 3).

According to Harlen (2012), assessments can be categorized as informal
formative, formal formative, informal summative, or formal summative. Formative
assessments are those that inform the teacher as to their teaching and instructional
practices and, additionally, inform the student of where they are in their understanding
Note. Adapted from Harlen (2012, 1998). Assessment for Learning overlaps Assessment of Learning when informal summative assessments use feedback that cycles back to the learner to inform them of their understanding.

using feedback from the assessment process (Black & Wiliam, 1998; Cowie & Bell, 1999; Ghaicha, 2016; Harlen, 2012; Shute, 2008). This definition of formative assessment has been used extensively over the past twenty years and is currently used by both ASA and MAA in their recommendations for assessment practices. Formative assessments inform Assessment for Learning, but informal summative assessments can be used formatively if they incorporate feedback for students and teachers (Harlen, 2012). Figure 3 shows this overlap of Assessment for Learning and Assessment of Learning depending on the instructor’s use of informal summative assessments.

Assessments can range from in-class questions (informal formative) to final
examinations (formal summative). Formal formative assessments are graded with feedback cycling back to the student and instructor, such as homework and quizzes. Quizzes and midterms can be informal summative if feedback does not inform the teaching process but is also used formatively to inform the student of their learning (Davies & Marriott, 2010; Harlen, 2012). Thus, these classifications of assessments can be seen as a continuum from “Assessment for Learning” to “Assessment of Learning” (Harlen, 2012) in Figure 3. Additionally, Stiggins (2002) and Black and Wiliam (2009) stressed the importance of moving more assessments from the Assessments of Learning category to those in the Assessments for Learning category. Together, these definitions situate Assessment for Learning and encompass the types of assessments utilized in a FAC.

Learning outcomes are a necessary component of Assessments for Learning. Black and Wiliam (2009) posited that to implement Assessment for Learning, the instructor must employ three critical processes: (1) clarify learning outcomes, (2) create tasks consistent with those outcomes to provide evidence of student learning, and (3) provide feedback (p. 8). Choosing the tasks for formative assessments must be “justified in terms of the learning aims that they serve” (Black & Wiliam, 1998, p. 143). Thus, learning outcomes must be clearly defined and communicated for formative feedback to be utilized (Stiggins, 2002). Stiggins suggested that instructors must know and communicate those learning outcomes in the syllabus before students engage in course material over those objectives. This communication allowed adult learners to recognize the learning goals or outcomes expected in the course curriculum (Yorke, 2003). Ghaicha
(2016) also expressed the importance of learning objectives in assessments. Specifically, Ghaicha defined assessments as an integral part of the learning process where instructors evaluate student achievements by “collecting, measuring, analyzing, synthesizing, and interpreting relevant information…under controlled conditions in relation to curricula objectives set for their levels” (p. 211). Indeed, specifying the course's learning outcomes assists the instructor in creating tasks relevant to the course material.

The tasks in formative assessments provide evidence of the learning outcomes and allow the adult learner to act upon the feedback from the assessments (Davies & Marriott, 2010). Thus, formative assessments allow students to progress through the course’s learning objectives by receiving feedback and taking ownership of their learning (Black & Wiliam, 2009). Assessment for Learning encourages adult learners by engaging with their learning directly through assessments that reflect the course’s content. As self-directed learners, the students act upon formative feedback to increase their understanding.

Feedback and Self-Assessment

Formative assessments comprise low to no-stakes assessments with feedback. Simple yet effective feedback is a necessary condition for an assessment to be formative. Black and Wiliam (1998) made a case for formative assessment in classroom practice through a meta-analysis of over nine years of empirical research. In a meta-analysis by Hattie and Timperley (2007), feedback was most effective when it related to learning goals. Hattie and Timperley described feedback as “one of the most powerful influences on learning and achievement” (p.81) by answering students’ questions about their
progress and direction in their learning. Steen (1999) maintained that assessment must be an open process through learning goals to inform both the teacher and the student of the student’s progress. Thus, the onus is on the adult learner to use that feedback to improve their understanding (Sosibo, 2019). Feedback that aligns with learning outcomes can help identify gaps in student knowledge and assist adult learners in seeing where they can improve.

Adult learners are motivated to learn and are self-directed. Thus, assessments that employ self-assessment from the feedback allow the student to learn from the feedback. Students who engage in self-assessment become more motivated to act on feedback to improve their learning (Ghaicha, 2016). “Indeed, unless linked to an effective process of reflection, assessment can easily become what many faculty fear: a waste of time and effort” (Steen, 1999, p. 5). Black and Wiliam (2009) drew on a partnership between teachers and students in their formative assessment framework, stating that students must be resources for themselves by taking ownership of their learning. Black and Wiliam (1998) cautioned that formative assessment cannot be productive without students being able and willing to self-assess to further their learning goals. Students’ self-assessment is the “essential component of formative assessment” (p. 143). Additionally, Stiggins (2002) underscored critical features of formative assessment by explaining that when students employ continuous self-assessment aligned with learning outcomes, they can better reflect on their knowledge developed over time. By incorporating regular self-assessment and feedback as part of the assessment cycle, students watch themselves grow as learners (Black & Wiliam, 2009; Stiggins, 2002; Wride, 2017). One of the outcomes of
Assessment for Learning, which Boaler and Confer (2017) found in their work, is that students changed perceptions of who they were as learners, thinkers, and problem-solvers after engaging in formative assessment procedures. Thus, Assessment for Learning can change the landscape of mathematics and statistics education by improving quantitative learners’ self-efficacy—their hope and belief in successfully meeting their educational pursuits (Boaler & Confer, 2017). Taken together, the body of research provides evidence that feedback with self-assessment assists adult learners with the desire and motivation to learn from their mistakes.

**Reassessment**

Reassessment allows for multiple attempts at learning tasks to progress through a curriculum. Having utilized feedback by self-assessment, reassessment allows students an additional opportunity to show evidence of learning. Because one purpose of formative assessment regards the learning process, reassessment can foster self-motivation, goal-orientation, and positive motivational beliefs of persistence and confidence; specifically, the ability to persevere despite an initial poor performance (Duckworth et al., 2007; Dweck, 2008; Yin et al., 2008). Steen (1999) asserted, “assessment that matters should always employ multiple measures of performance” (p. 4). One way of creating multiple measures of performance is by setting up assessments with opportunities for numerous attempts, retakes, or reassessments, giving students additional chances to demonstrate learning from the formative feedback (Abell et al., 2018; Dweck, 2008; Grant & Dweck, 2003). The opportunity to learn from their mistakes gives adult learners ownership of their learning and promotes self-efficacy and self-regulation.
Cognitive psychology has consistently found that errors on tests can spark significant learning and retention, but only if the feedback is immediate, not delayed (Brame & Biel, 2015; Hays et al., 2013). Brame and Biel (2015) recommended that low- and no-stakes testing environments offer the most benefit of “test-enhanced learning” (p. 9). Test-enhanced learning is when testing becomes a learning opportunity for students. Additionally, Brame and Biel (2015) found that incorporating frequent testing opportunities such as reassessments or retakes increased student learning in undergraduate science courses. The increased frequency of formative feedback allowed students to view formative assessments as “learning events,” and thus, students began to evaluate their errors based on learning outcomes for the course, creating the potential for greater recall on the reassessment (p. 10). This research offered evidence that Assessment for Learning provides many benefits to the learner when reassessment is allowed, such as retention of material, self-efficacy, and self-regulation. Together, these elements form a formative assessment cycle. How these elements of the FAC can be employed in a large-enrollment course is needed for instructors to import FACs in the curriculum successfully.

**Formative Assessment Cycle for Large-Enrollment Courses**

Although emphasized as important for student learning, formative assessment cycles are not yet fully incorporated in large-enrollment courses. Cash et al. (2017) discovered that the majority of large-enrollment courses use high-stakes summative assessments. To rectify this, the MAA suggested that large-enrollment courses utilize
technology to create continuous assessments with feedback and reassessments to achieve course objectives and learning goals (Abell et al., 2018). Computerized testing and automatic feedback are vehicles for timely, concise feedback and consistent grading in large-enrollment courses (Hattie & Timperley, 2007; Shute, 2008; Stirling, 2010). In fact, the use of technology in mathematics and statistics courses has been studied for over 35 years. An overwhelming body of evidence supports its use, such that both the MAA and ASA recommend technology be used throughout the introductory courses (Abell et al., 2018; ASA Revision Committee, 2016). The benefits of technology were further illuminated in a study by the American Mathematical Society (2009). They stated that students and instructors reported positive experiences using technology for online homework and quizzes. Having surveyed over 1,200 mathematics and statistics departments in universities in the U.S., they found three main advantages to online homework and learning systems: immediate feedback, multiple attempts allowed for incorrect problems, and less grading. Years of research on computerized assessments with automated feedback implicates that using technology is not just beneficial for students but recommended for introductory quantitative courses, as it allows for reassessment without an additional burden on the instructor and offers individualized attention to the student in large-enrollment courses.

Implementing formative assessments with feedback and self-assessment with reassessment opportunities creates a continuous cycle rather than isolated assessment events in large-enrollment courses. Furthermore, the use of technology can aid instructors and course designers in providing timely and consistent feedback that students and
instructors can use to determine the achievement of course objectives and learning goals. Finally, given the opportunity to reassess their understanding, adult learners in large-enrollment courses can better retain their learning and improve their self-efficacy and self-regulation, promoting greater achievement.

**Three Elements of the Formative Assessment Cycle and Student Attitudes and Achievement**

The three elements of formative assessment cycles are frequent formative assessments, feedback, and the opportunity to reassess. Large-enrollment courses have lagged in implementing formative assessments (Cash et al., 2017), yet formative assessments are recommended in the curriculum to help students improve both their attitudes and achievement (Abell et al., 2018; ASA Committee, 2016). Implementing FACs in large-enrollment introductory statistics courses could be a catalyst to successful student experiences, especially for students who are non-STEM majors. The literature on these three elements of the FAC elucidates the benefits to both students’ achievement and attitudes; however, the entirety of the cycle has yet to be incorporated as a framework for the large-enrollment introductory statistics course curriculum.

**Frequent, Low-Stakes Assessments and Attitudes and Achievement**

Evidence suggests that formative assessments as frequent, smaller assessments assist students’ attitudes and achievement in large-enrollment undergraduate courses by increasing motivation. Broadbent et al. (2018) suggested breaking up larger summative assessments into smaller, lower-stakes assessments to assist large-enrollment courses by
increasing the number of formative assessments. Although this was tested in a large-enrollment introductory psychology course and not a quantitative one, this research found that students reported increased motivation, improved ability to self-assess, and greater learning in the course (Broadbent et al., 2018). Breaking up larger summative assessments into more frequent formative assessments heeds the call to move more assessments from Assessments of Learning to Assessments for Learning.

Formative assessments can be employed both in-class and out of class. For example, in-class assessment strategies can utilize an online student response system (OSRS) or polling systems. Several studies suggested these in-class, no-stakes assessments effectively elicited student motivation and activated learning (Freeman et al., 2014; Gundlach et al., 2015, Muir et al., 2020). What was apparent was that through using these in-class low and no-stakes formative assessments, the curriculum embedded active-learning approaches, which positively affected many aspects (i.e., student achievement, attitudes, motivation, engagement, and perceived achievement) of the educational climate in large-enrollment courses (Freeman et al., 2014; Gundlach et al., 2015, Muir et al., 2020). For example, in a meta-analysis of 225 studies in STEM courses, online student response systems promoted active learning in large-enrollment course lectures. In these activated lectures, exam scores increased by 6% over traditional lectures (Freeman et al., 2014). These studies suggest that informal formative assessments positively impact important student outcomes in undergraduate STEM courses.

Using online formative activities as formative assessments also improved the
student outcomes of attitudes and achievement in large-enrollment introductory statistics courses. Gundlach et al. (2015) used both an OSRS and online homework in a large face-to-face section of introductory statistics. The authors measured student attitudes using the SATS (Schau, 2003) instrument and found that student attitudes improved in both the affect and cognitive competence subscales. In addition, these students’ summative exam scores were higher than in the flipped and online introductory statistics sections (Gundlach et al., 2015). In another large-enrollment introductory statistics study for non-statistics majors by Hodgson and Pang (2012), online formative activities (OFAs) improved self-regulation. More than 60% of students reported increased motivation, and over 70% stated that the OFAs helped their learning and understanding of statistics (Hodgson & Pang, 2012). These studies provide evidence that student attitudes toward statistics and student achievement are associated with and impacted by online formative activities.

Other quantitative introductory courses have also seen increased course performance using computer-assisted assessments such as online homework. For example, in a large study of both freshmen mathematics and statistics students, computer-based formative assessments helped identify underperforming students and improved final exam scores in learning mathematics and statistics (Tempelaar et al., 2014). Another large-scale study evaluated the use of web-based homework for a calculus course. The grades of the freshman students who utilized the web-based homework improved on average by two letter grades over those who did not (Hirsch & Weibel, 2003). Other studies with web-based homework showed advantages to online homework systems.
These web-based programs administer and provide automatically generated feedback to both the student and the instructor. Physiology students reported that using OFAs helped them prepare for and improve their scores on summative exams by using feedback from the OFAs to identify the gaps in their knowledge (de Kleijn et al., 2013). In a study of introductory statistics, the online homework system helped students increase exam performance significantly. The authors attributed this to the immediate feedback from the online homework (Balta & Güvercin, 2016). Thus, there is evidence that online formative activities can improve summative exam scores and final grades in large-enrollment undergraduate courses.

Another example of these smaller, lower-stakes assessments in large-enrollment quantitative literacy courses is the Just in Time Teaching (JiTT) model developed in 1996 to aid undergraduate science and mathematics students in using out-of-class time more effectively (Novak et al., 1999). With web-enhanced learning, the use of JiTT quizzes before class to inform teachers of student understanding relative to learning outcomes has been improved and made simpler (Abell et al., 2018). Natarajan and Bennett (2014) used modified JiTT quizzes in their study of calculus courses. The students made significant academic gains in calculus topics when formative quizzes on review material before class were implemented. This study provided evidence that altering the just-in-time teaching assessment protocol still improved student learning outcomes (Natarajan & Bennett, 2014). The implementation of JiTT has also been successful in introductory statistics courses at major universities. Testing JiTT’s effectiveness by analyzing pre- and posttest scores resulted in higher average posttest scores than in semesters where JiTT was not
implemented (McGee et al., 2016). Implementing these formative assessments before class time can improve student participation and ownership of class material (McGee et al., 2016; Natarajan & Bennett, 2014). These studies provide evidence that the JiTT model, which utilizes web-based formative assessments, improved several important student outcomes, from student attitudes to student achievement.

Therefore, the body of research suggests that computer-assisted formative assessments as frequent, smaller assessments improve students’ attitudes and achievement in large-enrollment undergraduate courses. Online or computer-assisted formative assessments, before, during, or after class, benefited students’ attitudes and academic achievement. Several studies have attributed the academic improvement to the automatic feedback that the online or computer-assisted formative assessments provided, allowing students to utilize self-assessment for greater understanding and performance on summative assessments.

**Formative Feedback and Attitudes and Achievement**

Students in large-enrollment courses have reported that communication with the professor lacks one-on-one, personable interaction (Cash et al., 2017), making feedback that students receive from these classes even more critical for student learning. Bayerlein (2014) investigated undergraduate students’ perceptions of feedback: both the timeliness of feedback and constructiveness. Constructiveness is concerned with the use of automatically generated feedback versus handwritten feedback. Interestingly, undergraduate students found the automatically generated feedback to be substantially
more constructive than the manually written feedback (Bayerlein, 2014). Simple, automatic feedback with non-judgmental wording aligned with learning outcomes is all that is needed to create constructive feedback (Stiggins, 2002). Complex feedback is unnecessary for students to gain information about correctness and learning goals; instead, productive, concise feedback improves learning and student outcomes (Abell et al., 2018; Shute, 2008). Simply stated, “provide feedback that moves learning forward” (Black & Wiliam, 2009, p. 8). These studies illustrate the benefit of automated feedback in providing succinct information regarding the students' understanding of the learning goals for the course.

Although complex feedback is unnecessary, the importance of feedback cannot be undersold as it is the component of formative assessment that is, in fact, “formative” (Black & Wiliam, 2008; Hattie & Timperley, 2007; Shute, 2008). It is often studied apart from formative assessment as students report motivation and competence with self-assessment and find that the formative feedback assists them in improving their performance on assessments (Abell et al., 2018; ASA Revision Committee, 2016; Shute, 2008). Two meta-analyses conducted on formative feedback and achievement found that providing feedback positively affected student achievement (Hattie & Timperley, 2007; Wisniewski et al., 2020). Specifically, computer-generated and corrective feedback benefited student learning ($d = 0.79$; Hattie & Timperley, 2007). Wisniewski et al. (2020) found that feedback is most effective when it includes several factors: timeliness, strategy, and self-regulation ($d = 0.48$). Additionally, feedback is associated with increased performance and learning when it maintains timeliness, concise elaboration on
the learning outcomes, and computer delivery (Shute, 2008). As the adult learner is self-directed, these meta-analyses provide important evidence that concise, automated feedback is sufficient to elicit self-regulation and affect student achievement.

Feedback can also improve self-efficacy in adult learners. Several studies have reported that the automated feedback in the formative assessments contributed to students’ attitudes regarding their learning experience (Balta & Güvercin, 2016; Beemer et al., 2018; Hodgson & Pang, 2012; Krause et al., 2009; Posner, 2011). For example, Krause et al. (2009) found that the online learning environment provided university students with perceived competence in statistics. Additionally, the feedback helped students’ achievement with little prior statistics knowledge or experience, which is particularly notable as students experience anxiety toward statistics and lack self-efficacy regarding their quantitative abilities in introductory statistics courses (Onwuegbuzie & Wilson, 2003; Williams, 2015). When Broadbent et al. (2018) studied a large-enrollment undergraduate course \((N \sim 1,500)\) that used formative assessments, they found over 83% of the students agreed that the online formative feedback on assessments motivated them to learn, improved their understanding, and increased their learning. Moreover, because the instructors improved upon the feedback for the following semesters, students’ average achievement increased by over 10% in the subsequent semesters as the feedback improved (Broadbent et al., 2018). Thus, feedback can provide essential information to the adult learner in large-enrollment courses, resulting in improved attitudes and achievement.

Large-enrollment introductory statistics courses employing formative assessments
with feedback have reported improved learning outcomes as well. Massing et al. (2018) studied computer-assisted assessments with automated feedback in a large-enrollment statistics course. They found three significant increases in academic outcomes due to using computer assessments with automatic feedback: student effort, student success in achieving learning outcomes, and final grades (Massing et al., 2018). Additionally, in another study on computer-assisted formative assessments in large-enrollment introductory statistics courses, the authors attributed the learning gains students made to the formative feedback that the online assessments promptly provided (Balta & Güvercin, 2016). Together, these studies support formative assessment with feedback as instructional interventions to improve student outcomes in large-enrollment introductory statistics courses. However, to make the cycle complete, allowing adult learners to retest their knowledge and learn from their mistakes is needed.

**Reassessment and Student Attitudes and Achievement**

An often-overlooked aspect of the formative assessment cycle is reassessment. With formative feedback being a necessary condition of formative assessment, allowing students to learn from their mistakes is a natural next step to creating a formative assessment cycle (FAC). Unfortunately, there is less literature studying reassessment opportunities in large-section introductory quantitative courses. However, as andragogy posits that adult learners are motivated and desire to learn through self-assessment, reassessment is a valuable element in the FAC for creating successful student pathways in the introductory statistics curriculum. For instance, Grant and Dweck (2003) studied
pre-med majors in a large chemistry class. They found that the ability for students to recover from a poor initial attempt through using achievement goals was the key to making retakes work as students coped better, increased their motivation, and performed better on future exams (Grant & Dweck, 2003). Reassessment allows adult learners to persist in their educational goals by learning from mistakes using the feedback related to learning outcomes for the course.

A study of undergraduate mathematics students specifically investigated reassessment opportunities through mastery-based testing (Collins et al., 2019). The assessments allowed multiple attempts with credit only for mastery. Over 80% of the students felt that the assessments and reassessment opportunities helped them understand the material, prepared them for problem-solving, and reflected their knowledge. Students reported feeling less pressure during examinations due to the reassessment opportunities and believed their attitudes improved because the reassessments added additional opportunities for success in the course (Collins et al., 2019). In addition, the reassessments also showed that students felt motivation to utilize the feedback, which stated simply whether the student “mastered,” was “progressing,” or was “insufficient” in that concept for the next assessment attempt. Because the learning outcomes were directly related to assessment material, the study suggested that students revisited course material, developing further understanding (Collins et al., 2019). Thus, the reassessment opportunities can provide students with the desire to self-assess their learning from the assessment feedback to progress in their mathematics course.

In a large study of undergraduate mathematics students ($N \sim 1,200$) using web-
based homework with the opportunity for students to revise and resubmit answers, Hirsch and Weibel (2003) found a high correlation ($r = .944$) between attempts made at homework problems and the percentage of problems solved in a calculus course. This study suggested that students persisted with the web-based homework system: they kept working on a problem until they achieved the correct solution, despite their prior mathematical ability for the course. Thus, students’ persistence was more a function of “effort rather than ability” (Hirsch & Weibel, 2003, p. 14). And in another study of mathematics undergraduates, Lenz (2010) found students’ homework scores improved due to the use of web-based homework, which utilized both feedback and the opportunity to resubmit homework multiple times. These studies suggest that students of varying mathematical backgrounds, when given an opportunity to rework missed homework problems, show persistence and greater achievement by working through their mistakes.

Two studies that evaluated reassessment specifically in introductory statistics found that attitudes toward statistics improved when students were provided opportunities to resubmit work. Posner (2011) found that those students who chose to resubmit work increased their proficiency in introductory statistics concepts, comparable to those who achieved proficiency with only one submission. Hodson and Pang (2012) noted that computer-assisted assessments are robust for creating favorable attitudes when online formative activities used feedback and multiple attempts. This study found that students were highly satisfied with the online approach and noted that their abilities to self-assess increased as they actively worked on finding solutions. Thus, providing opportunities for students to resubmit work allows students to gain a deeper understanding of the content.
and improve their attitudes and success in the course.

**Conclusion**

With the current focus on improving student achievement and creating successful experiences for all students in large-enrollment introductory statistics courses, this chapter offered a conceptual framework for a formative assessment cycle. The FAC comprises three elements: frequent low-stakes assessments aligned with learning outcomes, computer-generated feedback, and reassessment. Research has suggested that each of these elements of the formative assessment cycle is integral for assessments to be formative and improve student learning through adult learners’ self-efficacy and self-regulation. When an undergraduate quantitative course curriculum incorporates at least one of the three elements of the formative assessment cycle and measures student attitudes and achievement outcomes, the research indicates that students make positive gains in their understanding with an increased motivation to learn. Thus, the literature review has provided evidence that the three elements of formative assessment cycles are associated with positive attitudes and increased achievement. However, there is a lack of studies in statistics education research utilizing all three elements of the formative assessment cycle together to provide a comprehensive approach to introductory statistics courses to impact students’ attitudes toward statistics and student achievement.

Findings from the research on student attitudes and achievement emphasized that both attitudes and achievement are essential outcomes in introductory statistics courses (Emmioğlu & Capa-Aydin, 2012; Ramirez et al., 2012; Xu & Schau, 2019). Student
attitudes toward statistics were most often measured by the Survey of Attitudes Towards Statistics (SATS; Schau, 2003) in the research. The quantitative analyses of these studies provide a backbone for analyzing students' attitudes using the SATS instrument.

The focus of this study was to implement formative assessment cycles (FACs) in the introductory statistics curriculum to benefit both student attitudes and student achievement for students of different mathematical backgrounds. The body of assessment research provided evidence for formative assessments to be viewed as evaluative and transformative for student learning when accompanied by immediate feedback and the opportunity to reassess. The review also found that using technology aids instructors of large-enrollment courses to import formative assessment cycles into the curriculum. Individualized student attention can be achieved through computerized testing with automatic feedback. Where feedback is effective when it is concise, automated, and linked to the course's learning outcomes, students know whether they are achieving the course objectives. Additionally, reassessing students is done simply and efficiently through computerized test banks over course objectives. The reduction of time spent grading is also a benefit to instructors of these large-enrollment courses.

The formative assessment process has changed students’ beliefs about testing situations and their abilities by directing the learner’s goals from a focus on performance to a focus on learning (Grant & Dweck, 2003). Feedback accomplished this shift in goal orientation by helping students view learning as a skill advanced by practice, effort, and mistakes (Grant & Dweck, 2003; Shute, 2008). Building confidence, students have shifted responsibility for their learning to themselves, leading to life-long learning and
success in future classes (Ghaicha, 2016; Hassi & Laursen, 2015; Shute, 2008; Wride, 2017). The benefits of formative assessment on student learning and students’ self-efficacy are evident. Still, many large-enrollment courses have yet to adopt formative assessment practices to improve upon the important outcomes of attitudes and achievement. Taken together, embedding formative assessment cycles (FACs) in the introductory statistics curriculum could transform the experience of the introductory statistics student, encourage improved student attitudes toward statistics, and inspire greater academic achievement.
CHAPTER III
METHODOLOGY

Overview

Formative assessment, with feedback and multiple attempts for learning, embedded in an introductory statistics large-enrollment course can be a comprehensive pathway for students to achieve their quantitative literacy requirement in higher education. FACs provide introductory statistics students with frequent assessment opportunities, computer-generated feedback on their assessments, and multiple assessment attempts. This study implemented a quasi-experimental design to quantitatively evaluate the impact of FACs on both student achievement and student attitudes toward statistics.

This chapter describes the methodology for which the research questions were analyzed. The research design and the description of the participants and setting follow, after which the existing data set and instruments are discussed. The chapter concludes with the methods of data analyses for each research question.

Research Questions and Hypotheses

The purpose of this study was to quantitatively analyze the impact of FACs on student achievement and student attitudes toward statistics in large-enrollment introductory statistics courses. To meet the study’s purpose, the following research questions were offered.
Research Question 1

How do formative assessment cycles (FACs) affect student achievement in large-enrollment introductory statistics courses for different mathematically prepared students?

The hypotheses for Research Question 1 were as follows.

$H_0$: The difference in average achievement (average test scores) between students below an ALEKS score of 30 and those above a score of 30 before using FACs in large-enrollment introductory statistics courses is unchanged after FACs.

$H_A$: The difference in average achievement (average test scores) between students in large-enrollment introductory statistics courses below an ALEKS score of 30 and those above a score of 30 before using FACs in introductory statistics is reduced after FACs are implemented. In other words, FACs reduce the achievement gap between students below a placement score of 30 and those above.

Research Question 2

After allowing for student-to-student variability, which student attitude components change after a semester of a large-enrollment introductory statistics course with FACs? Also, how do demographic factors impact attitude, and do these effects change over time?

The hypotheses for Research Question 2 were as follows.

$H_0$: Student attitudes remain unchanged after a semester of introductory statistics with FACs.

$H_A$: Student attitudes change after a semester of introductory statistics with FACs.

Research Design

To examine these phenomena, I utilized a quasi-experimental research design with existing and deidentified data based on cohorts described in Table 1 to analyze the
research questions. A quasi-experimental design is utilized when a randomized, controlled trial is not feasible (Cresswell & Cresswell, 2018). A quasi-experimental design is most appropriate for this study because it is impossible to randomize the participants into treatment and control groups (Scher et al., 2015).

Table 1

*Research Questions, Data Source and Instruments, Participants, and Analysis*

<table>
<thead>
<tr>
<th>Research question</th>
<th>Instruments or data sources</th>
<th>Participants</th>
<th>Proposed data analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ 1 How do formative assessment cycles affect student achievement in large-enrollment introductory statistics courses for different mathematically prepared students?</td>
<td>Final exam scores, Course percentage grade ALEKS placement scores</td>
<td>Pre-FACs cohorts (Fall 2017 and Spring 2018) and FACs-cohort (Fall 2019)</td>
<td>Regression discontinuity; Slope plots; Descriptive statistics</td>
</tr>
<tr>
<td>RQ 2 After allowing for student-to-student variability, which student attitude components change after a semester of a large-enrollment introductory statistics course with FACs?</td>
<td>SATS-36 attitudes instrument (Schau, 2003) pre- and post-surveys ALEKS placement scores</td>
<td>FACs Cohorts: Fa 2021, Sp 2022</td>
<td>Descriptive statistics; Multilevel Modelling (MLM)</td>
</tr>
<tr>
<td></td>
<td>Demographic Data obtained in the pre-survey</td>
<td></td>
<td>Effect sizes of significant main effects or covariates in the model</td>
</tr>
</tbody>
</table>

This research took place in two phases. The first phase is regarded as Research Question 1. This quantitative, quasi-experimental design used existing and deidentified data from students in Fall 2017 through Fall 2019. Variables collected regarded students’
mathematics placement exam scores and achievement scores, described in Table 2. A quantitative, quasi-experimental design was most appropriate for this study to utilize the regression discontinuity method to effectively analyze Research Question 1 (Cunningham, 2021).

**Table 2**

*Pre-FACs and FACs Cohorts for Proposed Research Question 1*

<table>
<thead>
<tr>
<th>Semester</th>
<th>Course</th>
<th>Instructor</th>
<th>Exam type</th>
<th>Semester</th>
<th>Course</th>
<th>Instructor</th>
<th>Exam type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp 2018</td>
<td>1040-001</td>
<td>T1</td>
<td>Two identical paper-pencil midterms, final exam</td>
<td>Fa 2019</td>
<td>1040-001</td>
<td>T2</td>
<td>Six (T1) or seven (T2) computer exams with q's from identical question banks, two attempts allowed, final exam</td>
</tr>
<tr>
<td>Fa 2017</td>
<td>1040-001</td>
<td>T2</td>
<td>Two identical paper-pencil midterms, final exam</td>
<td></td>
<td>1045-001</td>
<td>T1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* T1 and T2 are two different lecturers.

The second phase of this quantitative, quasi-experimental research design was a pre- and post-survey methodology (Creswell & Creswell, 2018) to explore students’ attitudes toward statistics in semesters of large-enrollment introductory statistics courses implementing FACs. A quantitative design using a pre- and post-survey method was most appropriate to explore Research Question 2 using multilevel modeling (hierarchical regression) techniques as the students were nested in recitation sections and further nested in lecture sections within semesters (Hox et al., 2018). The survey data was obtained from existing and deidentified student survey data from the Fall 2021 and Spring 2022 semesters of the large-enrollment introductory statistics sections.
Participants and Setting

This study was set in an R1 university located in the western portion of the United States. The university population is primarily Caucasian, with an average undergraduate age of 22, and approximately 55% of the students are female. For the purposes of this study, the participants were undergraduate students enrolled in one of two introductory statistics courses to satisfy their quantitative literacy general education requirement, see Table 3.

Table 3

*Introductory Statistics Courses*

<table>
<thead>
<tr>
<th>Course number</th>
<th>Semester credit</th>
<th>Minimum ALEKS placement score</th>
<th>Lecture hours per week (max. Size of class)</th>
<th>Recitation hours per week (max. Size of class)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stat 1045</td>
<td>5</td>
<td>14</td>
<td>4 (350)</td>
<td>1.67 (30)</td>
</tr>
<tr>
<td>Stat 1040</td>
<td>3</td>
<td>30</td>
<td>2.5 (170)</td>
<td>1.67 (30)</td>
</tr>
</tbody>
</table>

Non-STEM majors primarily take the introductory statistics courses (Stat 1040 and Stat 1045). The two courses both emphasize conceptual understanding and statistical reasoning and cover the same statistical concepts: “types of studies, summarizing data, probability, [and] hypothesis testing” (http://catalog.usu.edu/). Additionally, these courses require students to complete the same final examination. However, the courses differ regarding the ALEKS mathematics placement exam score required for registration. Introduction to Statistics (Stat 1040) requires a score of 30 on the ALEKS mathematics placement exam. Introduction to Statistics with Elements of Algebra (Stat 1045) requires
a 14 on the ALEKS mathematics placement exam. Stat 1045 is considered a co-requisite course as it covers the algebra skills needed for the statistical topics, allowing less mathematically prepared students entry into the course. Due to the co-requisite nature of Stat 1045, it is a 5-credit semester-long course to include foundational algebra topics, while Stat 1040 is a 3-credit course. Both courses meet in a large-student lecture format twice a week: Stat 1040 has 2.5 student contact hours per week, and Stat 1045 meets for 4 student contact hours per week. Also, students in both courses register for a recitation section with enrollments of 20-30 students that meet twice a week for 50 minutes. Recitation leaders, who are students employed by the Mathematics and Statistics Department, lead the recitation sections. The recitation leaders are usually in their senior year of undergraduate study in a mathematics or statistics major or are mathematics or statistics graduate students. The recitation leaders work closely with the course instructor to provide similar instruction across the recitations, consistent with the content for the week outlined in the syllabus. The Stat 1040 and Stat 1045 weekly statistical content is identical.

It should be noted that the instructors (T1 and T2 in Appendix A) of the Stat 1045 and Stat 1040 courses for this study have a combined 55 years of teaching experience in teaching introductory statistics. These teachers are also award-winning instructors who students rate above the institutional and departmental averages in the university teaching evaluations each semester.

The USU registration website provided the number of participants enrolled in the Stat 1040 and 1045 courses used in this study, detailed by semester in Appendix A. The
aggregated numbers of participants are displayed in Table 4 by research question.

### Table 4

**Number of Participants by Research Question**

<table>
<thead>
<tr>
<th>Research question</th>
<th>Course</th>
<th>Participants</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research Question 1</td>
<td>Stat 1045</td>
<td>$n = 527$</td>
</tr>
<tr>
<td></td>
<td>Stat 1040</td>
<td>$n = 905$</td>
</tr>
<tr>
<td>Research Question 2</td>
<td>Stat 1045</td>
<td>$n = 194$</td>
</tr>
<tr>
<td></td>
<td>Stat 1040</td>
<td>$n = 347$</td>
</tr>
</tbody>
</table>

### Existing Data Set

I obtained Institutional Review Board (IRB) permission to receive the existing and deidentified data set from the semesters listed in Appendix A. The Center for Student Analytics and Register’s Office prepared and deidentified the data prior to my obtaining the data. Furthermore, my permission to access any of the student information from the Canvas course management system from the course sections listed in Appendix A was removed.

The existing data set was obtained from both large-enrollment introductory course sections of Stat 1045 and Stat 1040 detailed in Appendix A. For the semesters described in Table 2, from Fall 2017 through Fall 2019, these data were used to investigate Research Question 1. Due to the pandemic affecting Spring 2020, Fall 2020, and Spring 2021 semesters, course data from these semesters were not used for the study. The Fall 2021 semester was entirely webcast, with the Fall 2021 and Spring 2022 semesters
having the homework and final exam in a different format than Fall 2019. Thus, the Fall 2021 and Spring 2022 semesters were not included in the analysis of Research Question 1. To analyze Research Question 2, the SATS-36 (Schau, 2003) pre- and post-survey data were obtained from the large-enrollment sections of Stat 1045 and Stat 1040 in Fall 2021 and Spring 2022 semesters which have been embedded with FACs in the curriculum. The Registrar’s Office provided ALEKS placement score data, matching the students to their ALEKS scores before I accessed the deidentified data.

**Instruments and Data Sources**

Appendix A, “Data Sources, Deidentified for Proposed Study,” displays the data sources per course and semester that the Center for Student Analytics obtained and deidentified for this study. The data sources and instruments are explained as follows.

**Survey of Attitudes Toward Statistics: SATS-36**

In the Fall 2021 and Spring 2022 semesters, all students were assigned the SATS-36 (Schau, 2003) pre- and post-surveys. Candace Schau (Schau, C., personal communication, October 27, 2020) granted permission to use the SATS pre- and post-surveys through email correspondence (see Appendix B). These surveys are designed to measure students’ attitudes towards statistics both at the beginning and end of the semester. The pre- and post-surveys are found in Appendices C and D, respectively.

The pre-survey was assigned for students to complete during the first two weeks of class. Rather than the end of the semester to assign the post-survey, the post-survey was assigned over weeks 12 and 13 of a 15-week course. The post-surveys were assigned
in the Fall 2021 and Spring 2022 semesters during weeks 12 and 13 to not coincide with a unit test or final examination preparation.

The 36-item SATS pre- and post-surveys measure six attitude components: affect, cognitive competence, value, difficulty, interest, and effort. See the definitions of these components in Table 5 and see Appendix E and F for the full surveys. The 36-item surveys use a seven-point Likert scale (1 = “strongly disagree,” 2 = “agree,” 3 = “somewhat agree,” 4 = “neither agree nor disagree,” 5 = “somewhat disagree,” 6 = “disagree,” and 7 = “strongly agree”) (Schau, 2003).

Table 5
SATS-36 Attitude Components, Definitions, Examples, and Number of Items Per Component with Cronbach Alpha Ranges per Component

<table>
<thead>
<tr>
<th>Component</th>
<th>Definition</th>
<th>Example items</th>
<th>No.</th>
<th>Cronbach’s alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>affect</td>
<td>students’ feelings concerning statistics</td>
<td>“I am scared by statistics.”&lt;sup&gt;a&lt;/sup&gt; “I will like statistics.”</td>
<td>6</td>
<td>.80-.89</td>
</tr>
<tr>
<td>cognitive competence</td>
<td>students’ attitudes about their intellectual knowledge and skills when applied to statistics</td>
<td>“I can learn statistics.” “I will make a lot of math errors in statistics.”&lt;sup&gt;a&lt;/sup&gt;</td>
<td>6</td>
<td>.77-.88</td>
</tr>
<tr>
<td>value</td>
<td>students’ attitudes about the usefulness, relevance, and worth of statistics in life</td>
<td>“I use statistics in my everyday life.” “Statistics is not useful to the typical professional.”&lt;sup&gt;a&lt;/sup&gt;</td>
<td>9</td>
<td>.74-.90</td>
</tr>
<tr>
<td>difficulty</td>
<td>students’ attitudes about the difficulty of statistics as a subject</td>
<td>“Most people have to learn a new way of thinking to do statistics.”&lt;sup&gt;a&lt;/sup&gt; “Statistics formulas are easy to understand.”</td>
<td>7</td>
<td>.64-.81</td>
</tr>
<tr>
<td>interest</td>
<td>students’ level of interest in statistics</td>
<td>“I am interested in using statistics.”</td>
<td>4</td>
<td>.85&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>effort</td>
<td>amount of work the student expends to learn statistics</td>
<td>“I plan to work hard in my statistics course.”</td>
<td>4</td>
<td>.79&lt;sup&gt;c&lt;/sup&gt;</td>
</tr>
</tbody>
</table>

Note. The table is taken from [https://www.evaluationandstatistics.com/](https://www.evaluationandstatistics.com/).

<sup>a</sup> Negatively worded items.

<sup>b,c</sup> Interest and Effort components are pooled and reported from Nolan et al. (2012).
Additionally, the SATS-36 pre- and post-surveys ask global attitude items, and the post-survey also contains a global effort item. Additional items ask for relevant demographic and academic background information. Appendices C and D contain these specific questions for the pre- and post-survey, respectively.

**Validity of the SATS-36 Instrument**

Nolan et al. (2012) assessed the SATS (Schau, 2003) survey’s structural validity using confirmatory factor analysis, resulting in six factors congruent to the survey's six attitude components. Cronbach’s alpha coefficients measured the internal consistency of the SATS instrument's pre- and post-surveys (pooled administration of the surveys). Regarding the reliability of the SATS-36, the Cronbach alpha coefficients ranged between .79 and .90 for the six components, indicating internal consistency (Nolan et al., 2012). Templaar et al. (2014) found predictive validity with the SATS-36 instrument.

The cognitive competence component accounted for as much as 14% of the variability in student academic achievement. Generally, cognitive competence accounted for 2-14% of the variability in student academic outcomes (Nolan et al., 2012).

**Additional Questions to Assess Students' Experience of FACs**

Appendix E displays the additional questions I asked students in both the pre- and post-surveys to rate their expected and end-of-course experience with FACs.

**Achievement Scores**

Achievement scores for the pre-FACs Fall 2017 and Spring 2018 semesters of
Stat 1040 and Stat 1045 include two midterm test scores, the final examination score, and the final course percentage for each student. For the FAC semester, Fall 2019, achievement scores include all unit test scores for all attempts, the final examination score, and the final course percentage for each student.

**Midterm and Unit Test Scores**

Students took two midterm exams in Stat 1040 and Stat 1045 during class on paper in the pre-FACs semesters of Fall 2017 and Spring 2018. During the Fall 2018 and Spring 2019 semesters, test banks were created over the course learning objectives. The test banks were used to create the midterms for Fall 2018 and Spring 2019 semesters in both Stat 1040 and Stat 1045 courses. Thus, the midterms for Fall 2018 and Spring 2019 semesters were now computer-generated tests from the test banks. The students also took these exams at the university’s testing center. By the Fall 2019 semester, the test banks were fully created, aligned to learning objectives in the LMS, and the FACs were embedded in the courses as unit tests, rather than two midterms, and given in the testing center over a testing window of four days with retakes allowed.

In March 2020, during the Spring 2020 semester, students were sent home to study remotely due to the beginning of the pandemic. Fall 2020 and Spring 2021 semesters were also disrupted by physical distancing requirements and a hybrid teaching model: half the class attended while the other half joined via Zoom. For the Fall 2021 semester, all large-enrollment courses were required to teach via webcast. During the Spring 2021 semester, the introductory statistics courses were as similar as possible to a pre-pandemic teaching and learning environment; however, the final examination had
changed from pencil and paper to computer-graded, and the homework assignments were improved. Additionally, student attendance was low due to the high numbers of Covid-19 illnesses, suspected illness, or contact with someone with the illness. Thus, as regression discontinuity requires consistency across the pre-FACs and FACs semesters, the semesters of student achievement scores collected for analysis for Research Question 1 come from the Fall 2017, Spring 2018, and Fall 2019 semesters of Stat 1040 and Stat 1045 students.

After the year of creation of the computer-generated examinations from test banks on learning objectives (see Appendix F for the list of learning objectives for the introductory statistics course), FACs were completely implemented in the large-enrollment sections of Stat 1040 and Stat 1045 beginning Fall 2019 semester. Thus, instead of two midterms, students were assessed by smaller and more frequent unit tests using the FAC framework. Tests were computer-generated by randomly selecting questions from test banks on course objectives, automatic feedback was given immediately on each test question regarding correctness and the tested learning objective, and each unit test allowed two attempts. Both examination attempts were completed at the university’s testing center over a 4-day testing window. Upon completing the test, the university’s learning management system (LMS) automatically graded the exam and provided immediate feedback. The student could review their completed test to see what they did and did not get correct before leaving the testing center. Because the questions come from test banks linked to learning objectives, the LMS provides a Learning Mastery Gradebook (LMG) for students and instructors to view the learning objectives and
whether they were “mastered.” Appendix G displays pictures of the instructor’s view of the LMG (Figure G1) and a test student’s view of the LMG in their gradebook (Figure G2). The LMG requires at least an 80% level of correctness on each learning objective for mastery. Students can take the unit test again within the testing window; however, the test would not be identical to the first attempt because the computer randomly selects questions from the test banks. The student’s gradebook reflects the best of the two examination attempts. Thus, for the Fall 2019 FACs semester, the course data includes the achievement scores of all unit tests for all attempts.

The Final Examination

The final examination is a formal summative assessment given to students during the final examination week after each semester in all the introductory statistics courses included in this study. Since the Fall 2016 semester, the final examination has been a departmental examination. Thus, students in Stat 1040 and Stat 1045 all receive the same final examination in their respective semesters. The final exam questions change from semester to semester but cover the same topics. The examination is comprehensive, covering all seven units of the course, detailed in Appendix H, with approximately 33% of the final material from Units 1-3, 33% covering Units 4-6, and 33% covering Unit 7.

A team of instructors assigned to teach the course write the final examination each semester. The team also checks the final for accuracy and coverage. Students took the final examination on paper until the Spring 2020 semester when the pandemic closed campus. Beginning the Spring 2020 semester, the LMS administered, proctored, and graded the final examination, changing the final examination format to a digital version
of the paper and pencil version. Prior to Spring 2020, the instructor and the recitation leaders graded the final exam, with each page specifically graded by only one recitation leader to minimize grading variability.

**Final Course Percentage**

In addition to the midterm scores (pre-FACs), unit test scores (FACs), and final examination scores, the data set included the overall final course percentage for each student.

**ALEKS Math Placement Examination Score**

Beginning Fall 2017 semester, the USU Mathematics and Statistics Department used the ALEKS math placement examination (http://www.aleks.com) to serve as an indicator of mathematical readiness for quantitative coursework at the university. The scores on the ALEKS test are integer values between 0 and 100. The points come from correctly responding to prompts that assess prealgebra to trigonometric content. For example, if a student scores 100 on the ALEKS test, their performance is judged to be a perfect understanding of the content. A score of 30 means that the student understands 30% of the content. Students’ ALEKS placement scores are one way that determines placement into either Stat 1040 or Stat 1045. As a reference, students must score a minimum of 14 on the ALEKS test to take Stat 1045, the co-requisite course that teaches the exact curriculum as the Stat 1040 course but with elements of algebra. A minimum of 30 is required for enrollment into introductory statistics without the algebra co-requisite (Stat 1040), and a minimum of 46 to take pre-calculus (https://www.usu.edu/mathprep/)
The dataset included the ALEKS math placement score for students who used the ALEKS examination pre-requisite for registration into Stat 1040 and 1045.

**Data Analysis**

Multilevel modeling (MLM), sometimes referred to as Hierarchical Linear Models or Linear Mixed Effects Models, was used to determine significant “relationships between variables that are measured at a number of different hierarchical levels” (Hox et al., 2018, p. 3). This allows the use of regression for situations where there are clusters (e.g., students within a class), which otherwise violates assumptions of linear regression. The intraclass correlation (ICC) quantifies the within-level variability, with higher levels indicating higher dependency within clusters. More than one level of clustering is also straightforward to manage in this framework. For instance, analysis of students nested within recitation sections and classes is possible with the incorporation of additional random effects in MLM and can be investigated.

Each research question employs MLM analyses but with different goals. All data analysis was conducted with R 4.2.2 (R Core Team, 2022), and MLM was implemented with the lme4 package (Bates et al., 2015). Unless otherwise stated, a significance level of .05 was utilized.

**Analyses for Research Question 1**

The data set captured the following variables on each student for each semester listed in Appendix A to analyze Research Question 1. For the pre-FACs and FACs semesters, the following qualitative and quantitative variables were obtained for each
student. Variables for descriptive statistics (DS) and exclusionary factors (EV) are noted below. For the MLM analysis of Research Question 1, independent variables (IV), possible nesting variables (NV), and dependent variables (DV) are also noted. The qualitative variables were:

- sex (DS)
- class rank: freshman, sophomore, junior, senior, graduate (DS)
- attempt at the class (first attempt, second, third, etc.) (EV)
- whether the student completed or withdrew from the course (EV)
- pre-requisite utilized: ACT score, previous math class, ALEKS (EV)
- semester: fall or spring (NV)
- year of course (NV)
- type of introductory statistics taken: Stat 1040, Stat 1045 (IV)
- recitation section (TA) (NV)
- instructor (coded as T1 or T2) (NV)

The quantitative variables were:

- age (DS)
- ALEKS math placement exam score (IV)
- final exam percentage (DV)
- course final grade percentage (DV)
- each unit or midterm test percentage (DV)
- the reassessed unit test scores (if applicable, DV)

To answer Research Question 1, regression discontinuity methodology was used. Regression discontinuity investigates causation using a quasi-experimental design by providing unbiased regression estimates on a treatment effect without randomization of the subjects to treatment (Thistlewaite & Campbell, 1960). This is accomplished by using an exogenous cutoff score from a continuous variable measured prior to the treatment. Thus, using the mathematics placement exam score, which measures the mathematical preparedness of a student before enrolling in introductory statistics is, as Boylan et al. (1999) describe, an appropriate way to create groups in a regression discontinuity design.
Moreover, by using the cutoff to assign students to the two groups who are otherwise similar with respect to all variables, controlling for covariates is not needed in order to obtain the regression estimates on the treatment effect (Lesik, 2006). Even more importantly, regression discontinuity is unique in that its methodology provides causal inferences by using a sample of participants on either side of the cutoff (Cunningham, 2021). The key to providing causal inferences in the regression discontinuity methodology is its ability to eliminate selection bias (Cunningham, 2021). Using the cutoff as the determining factor to place students into treatment and control groups, the methodology requires that these groups are otherwise similar. Due to the pandemic affecting the teaching modality and the way the final exam was administered, only the Fall 2019 semester is the FACs semester for which all other variables are as similar as possible to the pre-FACs semesters of Fall 2017 and Spring 2018.

Simulated data is shown in Figure 4 to provide an example of regression discontinuity using the ALEKS placement cutoff score of 30 and the achievement score as the dependent variable. Students scoring below 30 are placed into the corequisite introductory statistics course, Stat 1045, and students who score at least 30 are placed into Stat 1040.

Because students below a mathematics placement score of 30 placed into Stat 1045 and those who score above 30 placed into Stat 1040, I compared the achievement scores of these students around the cutoff \( (D = 30) \) to find a baseline difference between the achievement of each group. The university began using the ALEKS math placement exam beginning Fall 2017. In Fall 2017 through Spring 2018, both Stat 1040 and 1045
received the same two midterms and final examinations in their respective semesters. The difference between the slopes of the achievement scores around the cutoff provided a baseline estimate of the achievement gap between the two courses.

Statistical inference was used to detect changes at the discontinuity via MLM due to the nested nature of the data. Figure 5 shows the nested structure of the data. As shown, students are nested within the recitation section in the class and the semester. Random intercepts by class and semester helped to control for section-to-section variability, accounting for the teaching effect of the teaching assistant.
Using random intercepts to control for teacher-to-teacher variability of the recitation section and centering the ALEKS scores at the cutoff of 30, the baseline model was as follows.

\[ \text{achievement}_{Di} \sim (\text{aleks}_i - 30) + (1|\text{section}) \text{ where } D_i = \begin{cases} 1 & \text{if ALEKS } \geq 30 \\ 0 & \text{if ALEKS } < 30 \end{cases} \]  

(1)

Thus, combining the two models in Equation 1 for the slopes before and after the cutoff yielded Equation 2. This equation was used for both baseline and FACs comparisons separately.

\[ Y_i = \beta_0 + \beta_1 A_i + \beta_2 D_i + \beta_3 (A_i \times D_i) + u_j + e_i \]

(2)

where \( e_i \sim N(0, \sigma^2) \) and \( u_j \sim N(\mu_j, \sigma_j^2) \)

The analyses were investigated at the baseline and again for the FACs semester (Fall 2019). The coefficients of interest were \( \beta_2 \) (whether there is a difference in achievement
right below and right above the cutoff) and $\beta_3$ (whether the slope of ALEKS differs for those below and above the cutoff). First, I investigated whether there is a gap in the achievement scores. Second, I determined the significance of the relationship of the interaction between the student being above or below the cutoff and the math placement score. This first analysis helped determine the overall patterns for baseline and FACs separately.

There was a second analysis to determine whether the performance gap depended on FACs depicted in Equation 3.

$$Y_i = \beta_0 + \beta_1 A_i + \beta_2 D_i + \beta_3 (A_i \times D_i) + \beta_4 (T_i) + \beta_5 (D_i \times T_i) + u_j + e_i \quad (3)$$

where $e_i \sim N(0, \sigma^2)$ and $u_j \sim N(\mu_j, \sigma_j^2)$ and $T_i = \begin{cases} 0 & \text{if baseline semester} \\ 1 & \text{if FACs semester} \end{cases}$

The coefficient of interest was $\beta_5$ (whether there is a difference in achievement right below and right above the cutoff and if that difference depends on whether they had FACs or baseline). Non-significant results for this estimate did not indicate that the difference in performance depended on FACs. Notably, this model specification assumes that the slopes did not significantly differ in the first analyses (the coefficient labeled $\beta_3$ in that model). Then a three-way interaction was analyzed between ALEKS scores, the cutoff indicator ($D$) and $T$ (the indicator for FACs or baseline) within this model to allow the slopes to vary by both indicators. This three-way interaction allowed for interaction plots and simple slopes analyses to be conducted, and regression estimates calculated.

Assumptions of MLM were visually assessed with residual diagnostics with no evident violations.
**Threats to Validity**

Three main threats to validity that could affect the regression estimates in regression discontinuity analysis in this quasi-experimental design (Lesik, 2006): (a) scores from students who take introductory statistics more than once; (b) scores from students who do not take the ALEKS placement exam but place into the class via their ACT or prior math course grade; and (c) scores from students who took Stat 1045 even though they placed in Stat 1040 based on their ALEKS score.

**Taking the course multiple times.** The deidentified data provided a unique ID number for each student. Thus, students who drop or retake the course showed up in the data set multiple times and then were dropped from the analysis. Thus, after dropping these students, there were no students in the pre-FACs semesters who appeared in the FACs semester; the students in these groups were mutually exclusive. Naturally, care was taken that all students in the analysis received the same treatment amount to have unbiased estimates of the treatment effect. I utilized the method from Lesik (2006) to verify that the student received one and only one full semester of introductory statistics. That is, students who took the final exam and received a D- or better grade received the full semester, and those who withdrew or received an F grade did not and were dropped from the analysis.

**Not taking the placement exam.** A necessary condition of regression discontinuity is the cutoff, or the math placement exam score. Because the university allows students to place into math classes using methods other than ALEKS scores, only students with ALEKS scores were included in the regression discontinuity analysis.
**Crossovers.** Regression discontinuity relies on “perfect compliance. . . based solely on the student's placement test score” (Lesik, 2006, p. 11). Students who “crossover” and choose to take Stat 1045 even though they placed in Stat 1040, if random, can be dropped from the analysis to ensure compliance. Thus, care was taken to ensure that students in the analysis were not crossovers. Additionally, the estimates and standard errors were computed with and without crossovers to verify that the crossovers were not affecting the estimates.

**Limitations**

As a quasi-experimental study, there are limitations to the results. Generalizability beyond this university is cautioned, as other settings, students, course curriculum, teaching practices, and assessments can impact the achievement outcomes (Cresswell & Cresswell, 2018). Second, the students used in the analysis were from semesters before the pandemic. The pandemic and subsequent stresses placed on students could impact students in various ways, mentally and physically, barring students from a semester similar to a pre-pandemic experience.

**Analysis for Research Question 2**

The purpose of Research Question 2 was to investigate changes in student attitudes toward statistics over a semester of Introductory Statistics with FACs. To analyze this research question, variables obtained on students in the Fall 2021 and Spring 2022 semesters include those used for descriptive statistics, covariates, nesting variables and independent and dependent variables. For the MLM analysis, independent variables
(IV), nesting variables (NV), and dependent variables (DV) are noted. They are as follows.

- introductory statistics course taken: Stat 1040, Stat 1045 (qualitative, NV)
- ALEKS math placement exam score (quantitative, IV)
- SATS-36 survey scores (quantitative, DV)
- semester: fall or spring (qualitative, NV/IV)
- recitation section (qualitative, TA) (NV)
- Other survey information and demographics listed in Appendix E (quantitative and qualitative, IV)

After controlling for student-to-student variability, again using MLM, statistical inference detected whether there was improvement in each of the six attitude components from pre- to post-survey. This approach allows the investigation of any longitudinal changes similar to repeated measures analysis of variance (RM ANOVA) or multivariate analysis of variance (MANOVA) but offers additional advantages. Specific to this research, MLM allowed for the incorporation of incomplete data while accounting for the dependence among observations (Hox et al., 2018). Additionally, both categorical and continuous covariates were explored to control for potential confounding or moderating effects. As before, further nesting of students within the recitation section and lecture is possible with the incorporation of additional random intercepts in MLM and were investigated in the analysis of Research Question 2. See Figure 6 for the nesting structure of the data for this research question.

The following explains the models and equations used for the analysis of Research Question 2.
Figure 6

Nesting Structure for Analyzing Attitudes as a Repeated Measure for Research Question

1. After allowing for student-to-student variability using random intercepts for the nesting, how do student attitude components (dependent variables, DV) change after a semester (repeated measure at two time points) of large-enrollment introductory statistics courses with FACs?

2. Can further nesting allow for recitation section or recitation leader (TA) variability (random effects for section and TA)?

3. Also, how do demographic factors impact attitude, and do these factors affect
change over time? (The models below depend on whether significance of nesting students by recitation leader is determined in Model 1. (For ease, the nesting is not included in the following models, and error terms are defined in EQ1.)

- **Model2**: attitude ~ time + covariate + ... + (1|id)

- EQ2: \( Y_i = \beta_{00} + \beta_{1i} T_{ti} + \sum_j \beta_{j} \times \text{covariate}_i + \ldots + u_{0i} + e_{it} \)

- **Model3**: attitude ~ time*covariate + (1|id)

- EQ3: \( Y_i = \beta_{00} + \beta_{1i} T_{ti} + \sum_j \beta_{j} \times \text{covariate}_i + \sum_j \beta_{j} T_{ti} \times \text{covariate}_i + u_{0i} + e_{it} \)

Potential confounding and moderating effects of covariates were investigated by incorporating fixed effects and associated likelihood ratio tests (LRT) between nested MLM models optimized with maximum likelihood (ML). Final MLM models were reoptimized with restricted maximum likelihood (REML), parameter estimates tabulated, and estimated marginal means visualized for attitude components in which a statistically significant change was found. Post hoc pairwise \( t \)-tests were computed, and effect sizes were calculated as standardized mean differences (a Cohen’s \( d \)-like measure), in which point differences are divided by the pooled standard deviation. The pooled standard deviation was estimated by pooling the variance components of the best fitting MLM.

Assumptions of MLM were visually assessed with residual diagnostics, and no evident violations were found. All data analyses were conducted with R 4.2.2 (R Core Team, 2022), and MLM was implemented with the lme4 package (Bates et al., 2015). Unless otherwise stated, a significance level of .05 was utilized.

**Limitations**

During the Fall 2021 semester, all large-enrollment courses were required to be taught webcast, different from Spring 2022. Thus, the teaching modality could be a major
predictor of student attitudes, and if so, controlling for this variable will be needed in the MLM.

As with all quasi-experimental research, generalization of the results beyond this university is cautioned. The data for this proposed study will come from a nonrandomized group of students in two semesters using the FACs curriculum. Also, the ability to model the nested structure of recitation sections with random slopes could be limited if the sample size is too small (sparsity).

**Summary**

Using existing data, a quasi-experimental quantitative research design in two phases was conducted to analyze the impact of FACs in undergraduate large-enrollment introductory statistics course curricula on student attitudes toward statistics and student achievement. Research Question 1 used regression discontinuity to identify changes in student achievement around the ALEKS placement exam score of 30 between pre-FACs semesters and the semester where FACs was implemented. The ALEKS placement score determined whether the student placed in Stat 1045 or Stat 1040. The results and analysis of Research Question 1 provide feasibility: the successful import of FACs under ideal circumstances with experienced instructors.

Research Question 2 investigated student attitudes across the Fall 2021 and Spring 2022 semesters using the SATS-36 (Schau, 2003) instrument. To analyze these data, MLM was used to address the nesting structure of students within sections and the dependence among observations. Additionally, MLM measured the longitudinal change
from pre- and post-survey responses. This study contributes to the field of statistical education research by providing empirical evidence for utilizing formative assessment research in large-enrollment courses to improve the important student outcomes of students’ attitudes towards statistics and student achievement. Additionally, these results provide a foundation for other colleges and universities to study how formative assessment cycles can be employed in their large-enrollment courses. Thus, this research exemplifies a successful pathway for students to complete their undergraduate quantitative requirements, creating a positive experience that transcends the classroom to their citizenship in a data-centric world.
CHAPTER IV

RESULTS

“True scientists do not collect evidence in order to prove what they want to be true or want others to believe. That is a form of deception and manipulation called propaganda, and propaganda is not science.” (Cunningham, 2021, p. 10)

The purpose of this study was to investigate the impact of an embedded cycle of formative assessment with feedback and reassessment opportunities in the curriculum of large-enrollment introductory statistics courses on student attitudes toward statistics and student achievement scores. Using a quasi-experimental quantitative research design, this study sought to answer the following two research questions regarding the effects of FACs on student attitudes toward statistics and statistics achievement.

1. How do formative assessment cycles (FACs) affect student achievement in large-enrollment introductory statistics courses for different mathematically prepared students?

2. After allowing for student-to-student variability, which student attitude components change after a semester of a large-enrollment introductory statistics course with FACs? Also, how do demographic factors impact attitude, and do these effects change over time?

This chapter details the results of this analysis by research question. The Center for Student Analytics at Utah State University gathered and deidentified the pre-existing dataset for this analysis after IRB approval. Appendix A details the student data I obtained on Stat 1040 and Stat 1045 students from the Fall 2017, Spring 2018, Spring 2019, Fall 2021, and Spring 2022 semesters. The pre-existing dataset included achievement data gathered from all semesters. Student responses for the SATS-36 (Schau, 2003) pre-and post-surveys in Fall 2021 and Spring 2022 semesters measured the students’ attitudes toward statistics. The Center for Student Analytics also provided
demographic data such as sex, class rank, recitation section, and recitation teacher. Additionally, ALEKS math placement scores were matched to students who used the ALEKS exam for their prerequisite requirement. The next two sections detail the analyses for each research question. The chapter ends with a summary of these results.

**Analysis for Research Question 1**

Research Question 1 examined students’ achievement in semesters before and after FACs were implemented in large-enrollment introductory statistics courses. MLM was used to analyze the overall course achievement (final grade percentage, DV) on their math placement exam scores (ALEKS scores, IV) using the methodology of regression discontinuity. This methodology allowed me to explore the students’ achievement at the exogenous cutoff, placing students below this cutoff in Stat 1045 and students above this cutoff in Stat 1040. In this section, I provide the demographic and summary statistics, exploratory data analysis, and MLM results and conclusions. The complete analysis in an R markdown file is available upon request.

**Descriptive and Summary Statistics**

Students’ achievement scores from Fall 2017, Spring 2018, and Spring 2019 were collected, including the student’s prerequisite math course or placement exam score. Demographic data, such as sex, class rank, age, and the number of previous attempts at the course, were also obtained on these students. A full list of the variables collected on these students by semester is found in Appendix A. After I obtained the data, I removed the 549 students who did not use the ALEKS math placement exam for their prerequisite
from this dataset, leaving 730 students. Then I removed 50 students who had taken the class previously. Finally, I removed 104 students with a final grade percentage below 60% or who withdrew from the course during these semesters, leaving a sample of 576 students. This sample of 576 students were those who had an ALEKS math placement score and achieved a D- or better for their first time taking introductory statistics in one of Fall 2017, Spring 2018, or Spring 2019 semesters. Students in Fall 2017 and Spring 2018 semesters were assessed with two midterms, a final, and no formative assessments (pre-FACs). Students in Spring 2019 experienced a semester with a FAC-based curriculum, including formative assessments, feedback, and reassessments throughout the course. Table 6 shows the summary statistics of the 576 unique students.

Figure 7 displays the histograms of all ALEKS scores from the Fall 2017, Spring 2018, and Spring 2019 semesters of Stat 1040 and Stat 1045. Figure 7 shows that some students who scored over 30 chose to take Stat 1045 rather than Stat 1040. Additionally, there are extremely high ALEKS placement scores. Students can choose to take Stat 1040 or Stat 1045 even if they place into a higher introductory statistics course due to their ALEKS scores; some students’ majors only require Stat 1040 or Stat 1045, and therefore that is what they register for despite having had a strong mathematical background.

The distribution of the final grades of the 576 students in the sample is shown in Figure 8. The distributions appear similar between the courses.

**Crossovers and Extreme Cases**

An ALEKS score of 14 places students into the co-requisite introductory statistics course, Stat 1045, and a score of 30 places students in Stat 1040, although students are
Table 6

Descriptive and Summary Statistics of Participants for Research Question 1

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 576)</th>
<th>Pre-FACs (N = 383)</th>
<th>FACs (N = 193)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>361</td>
<td>62.7</td>
<td>233</td>
<td>60.8</td>
</tr>
<tr>
<td>Male</td>
<td>215</td>
<td>37.3</td>
<td>150</td>
<td>39.2</td>
</tr>
<tr>
<td>Course</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stat1040</td>
<td>345</td>
<td>59.9</td>
<td>230</td>
<td>60.1</td>
</tr>
<tr>
<td>Stat1045</td>
<td>231</td>
<td>40.1</td>
<td>153</td>
<td>39.9</td>
</tr>
<tr>
<td>Class rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>272</td>
<td>47.2</td>
<td>173</td>
<td>45.2</td>
</tr>
<tr>
<td>Sophomore</td>
<td>166</td>
<td>28.8</td>
<td>113</td>
<td>29.5</td>
</tr>
<tr>
<td>Junior</td>
<td>91</td>
<td>15.8</td>
<td>61</td>
<td>15.9</td>
</tr>
<tr>
<td>Senior</td>
<td>47</td>
<td>8.2</td>
<td>36</td>
<td>9.4</td>
</tr>
<tr>
<td>Age</td>
<td>20.70</td>
<td>3.94</td>
<td>20.69</td>
<td>3.18</td>
</tr>
<tr>
<td>Final grade %</td>
<td>82.19</td>
<td>9.82</td>
<td>81.78</td>
<td>9.60</td>
</tr>
<tr>
<td>ALEKS Score</td>
<td>34.27</td>
<td>11.42</td>
<td>34.32</td>
<td>10.81</td>
</tr>
</tbody>
</table>

Note. Pre-FACs semesters were the Fall 2017 and Spring 2018 students enrolled in Stat 1040 or Stat 1045. Spring 2019 was the FACs semester. All differences between the Pre-FACs and FACs groups are not statistically significant.

not forced to take Stat 1040 even if they place into it. If students want to take Stat 1045, they are allowed—these students are crossovers. There were 59 crossovers in the sample (10.2%). Figure 7 shows the extremely large ALEKS math placement scores and the number of crossovers above a score of 30 in Stat 1045.

Appendix I shows the summary and descriptive statistics when the 59 crossovers in Stat 1045 are removed. Table I.1 shows the descriptive statistics with crossovers removed with no significant differences between the Pre-FACs and FACs groups. Figure I.1 shows the distribution of ALEKS scores among Pre-FACs and FACs semesters with crossovers removed, and Figure I.2 shows the distribution of the final grade percentages.
Students with a minimum score of 46 on the ALEKS placement exam are placed into a higher-level introductory statistics course (Stat 2000) that requires a pre-calculus mathematical understanding. However, students are not required to skip Stat 1040 if they place in a course beyond this introductory level. Thus, even with a higher level of mathematical preparedness, students may take Stat 1040. I will call these students "extreme cases." Seventy-nine students (13.7\%) in the sample scored 46 or higher on the ALEKS math placement exam. Further restricting the ALEKS score to those below 46 provided a window of 15 points around the cutoff. Cunningham (2021) suggested that the
window around the cutoff must be narrowed as the regression discontinuity design only
provides causality around the cutoff. Limiting the ALEKS scores to 15 on either side of
the cutoff localized the ALEKS scores around the cutoff. Figures J.1 and J.2 in Appendix
J display the distributions of ALEKS scores and final grade percentages by Pre-FACs and
FACs groups. Table J.1 displays the descriptive and summary statistics of the students’
demographics. There were no significant differences between the Pre-FACs and FACs
groups with extreme cases removed.

Twenty students were both crossovers and extreme cases (3.5%), which accounts
for 33.9% of the crossovers and 25.3% of the extreme cases. With the crossovers and

---

**Figure 8**

*Histogram of the Final Grade Percentages by Course*
extreme cases removed \((n = 123)\), 453 students remained in the sample. Appendix K contains the descriptive statistics for the data with crossovers and extreme cases removed. Table K.1 shows the descriptive and summary statistics, histograms of ALEKS scores are provided in Figure K.1, and final grade percentages by Pre-FACs and FACs groups are found in Figure K.2. There were no significant differences between the Pre-FACs and FACs groups with the crossovers and extreme cases removed.

**MLM Analyses for Regression Discontinuity**

Using random intercepts to control for recitation section-to-section variability and centering the ALEKS scores at the cutoff of 30, the baseline model was fit to determine whether there was a change in the slopes from the pre-FACs to FACs achievement. I performed an Analysis of Variance (ANOVA) using Type III Sums of Squares with Satterthwaite’s method on each model to sequentially test the fixed effects for significance of the main effects since there was no significant interaction between the ALEKS score and whether or not the student’s score was 30 and above, see Table 7.

The slope at the discontinuity for the FACs group was significant \((b = -6.45, p = .033)\). Thus, there was a significant difference in achievement right below and right above the cutoff. The slope of the ALEKS scores was significant for the FACs group \((b = .45, p = .003)\). Using the regression estimates in Table 7, calculations for the final grade at the discontinuity for the pre-FACs and FACs groups can be made. For the pre-FACs group, at an ALEKS of 30, the Stat 1040 student averaged 80.04% in the class. Students who scored a 29 on the ALEKS placement exam were placed into Stat 1045 and averaged 81.04% in the class. However, in the FACs group, students with an ALEKS score of 30
Table 7

Regression Estimates for Baseline Models at the Cutoff (ALEKS = 30)

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pre-FACs</th>
<th>FACs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>81.15***</td>
<td>1.86</td>
</tr>
<tr>
<td>ALEKS scorea</td>
<td>0.11</td>
<td>0.19</td>
</tr>
<tr>
<td>30 &amp; aboveb</td>
<td>-1.11</td>
<td>2.04</td>
</tr>
<tr>
<td>ALEKS scorea × 30 &amp; aboveb</td>
<td>0.16</td>
<td>0.20</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation section</td>
<td>6.45</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>83.33</td>
<td></td>
</tr>
</tbody>
</table>

Sample size
- Recitation sections: 30 (Pre-FACs), 15 (FACs)
- Participants: 383 (Pre-FACs), 193 (FACs)

Note. Significance was determined by performing a Type III Sums of Squares ANOVA with Satterthwaite’s method.

aALEKS score was centered at 30.
b0 = false, 1 = true.
*p < .05. **p < .01. ***p < .001.

averaged 79.83% in the class, and students with an ALEKS score of 29 averaged 85.83% in the class.

Because the ALEKS scores at the discontinuity was significant for the FACs group in Table 7 but not significant for the interaction between the ALEKS score and whether the student’s ALEKS score was 30 and above, two-way interactions were fit. EQ(full) in Chapter III shows the equation for the model. The first interaction in the model is the ALEKS score and whether the student’s ALEKS score was 30 and above, and second interaction is whether the students’ ALEKS scores were 30 and above and whether the students were in the FACs semester. Both interactions were not significant in
the model, and Type III Sums of Squares ANOVA with Satterthwaite's method was performed to sequentially test the main effects for significance. Model 1 of Table 8 shows there was statistical significance for the slope of the ALEKS scores ($b = 0.24$, $p < .001$).

To visualize what is happening at the cutoff, I fit the data to a three-way interaction model between the variables of the students’ ALEKS scores, whether the scores are above 30, and whether the students were in the FACs semester (see Model 1 of Table 9). However, the three-way interaction did not show significance, so Type III Sums of Squares ANOVA with Satterthwaite's method was performed to test the fixed effects sequentially. Only the main effects of the ALEKS score ($b = .10$, $p < .001$) and whether the student’s ALEKS scores were above 30 ($b = -1.29$, $p = .031$) were statistically significant.

Then, I performed simple slopes analysis on the three-way interaction model. Although the three-way interaction was not significant, simple slopes allowed me to investigate the slopes at the cutoff within the Pre-FACs and FACs groups. Model 1 in Appendix L shows that the slopes significantly differ from 0 in pre-FACs and FACs groups for those who scored at least 30 on the ALEKS math placement exam.

The three-way interaction model allowed me to plot the student achievement on the ALEKS scores before and after the discontinuity for the pre-FACs and FACs groups. Figure 9 provides a visual estimate of the slopes of the achievement around the cutoff, suggesting that there are differences between the final course percentages at an ALEKS score of 30 despite the three-way interaction not being significant. It is important to note
### Table 8

**Models Fit to Equation 3**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th>Model 4</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( b )</td>
<td>( SE )</td>
<td>Variance</td>
<td>( n )</td>
<td>( b )</td>
<td>( SE )</td>
<td>Variance</td>
<td>( n )</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>82.33*** 1.62</td>
<td></td>
<td>82.36*** 1.63</td>
<td></td>
<td>82.86*** 1.65</td>
<td></td>
<td>82.77*** 1.63</td>
<td></td>
</tr>
<tr>
<td>ALEKS score(^a)</td>
<td>0.24*** 0.15</td>
<td></td>
<td>0.24*** 0.15</td>
<td></td>
<td>0.23*** 0.15</td>
<td></td>
<td>0.23*** 0.15</td>
<td></td>
</tr>
<tr>
<td>30 &amp; above(^b)</td>
<td>-2.72 1.77</td>
<td></td>
<td>-3.59* 1.89</td>
<td></td>
<td>-4.41* 1.90</td>
<td></td>
<td>-5.29** 1.98</td>
<td></td>
</tr>
<tr>
<td>FACs(^c)</td>
<td>1.90 1.90</td>
<td></td>
<td>1.64 1.91</td>
<td></td>
<td>1.60 1.95</td>
<td></td>
<td>1.63 1.88</td>
<td></td>
</tr>
<tr>
<td>ALEKS score × 30 &amp; above(^b)</td>
<td>0.06 0.16</td>
<td></td>
<td>0.20 0.20</td>
<td></td>
<td>0.12 0.16</td>
<td></td>
<td>0.30 0.20</td>
<td></td>
</tr>
<tr>
<td>30 &amp; above(^b) × FACs(^c)</td>
<td>-1.17 2.13</td>
<td></td>
<td>-1.40 2.21</td>
<td></td>
<td>-0.48 2.38</td>
<td></td>
<td>-1.28 2.34</td>
<td></td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation section</td>
<td>4.73</td>
<td></td>
<td>4.56</td>
<td></td>
<td>4.52</td>
<td></td>
<td>3.46</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>86.60</td>
<td></td>
<td>88.27</td>
<td></td>
<td>88.91</td>
<td></td>
<td>90.20</td>
<td></td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation sections</td>
<td>45</td>
<td></td>
<td>45</td>
<td></td>
<td>45</td>
<td></td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>Participants</td>
<td>576</td>
<td></td>
<td>497</td>
<td></td>
<td>517</td>
<td></td>
<td>453</td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Significance of fixed effects was determined by performing a Type III Sums of Squares ANOVA with Satterthwaite's method.

\(^a\)ALEKS score was centered at 30.

\(^b\)0 = false, 1 = true.

\(^c\)pre-FACs = 0, FACs = 1.

\(^*\)\( p < .05\). \(^*\)\( p < .01\). \(^***\)\( p < .001\).
### Table 9

#### Three-Way Interaction Models

<table>
<thead>
<tr>
<th>Effect</th>
<th>Model 1</th>
<th></th>
<th>Model 2</th>
<th></th>
<th>Model 3</th>
<th></th>
<th>Model 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>SE</td>
<td>Variance</td>
<td>n</td>
<td>b</td>
<td>SE</td>
<td>Variance</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>81.24***</td>
<td>1.86</td>
<td></td>
<td>81.25***</td>
<td>1.88</td>
<td></td>
<td>81.81***</td>
</tr>
<tr>
<td>ALEKS score(^a)</td>
<td>0.10***</td>
<td>0.19</td>
<td></td>
<td>0.10***</td>
<td>0.19</td>
<td></td>
<td>0.09***</td>
</tr>
<tr>
<td>30 &amp; above(^b)</td>
<td>-1.29*</td>
<td>2.04</td>
<td></td>
<td>-2.53*</td>
<td>2.18</td>
<td></td>
<td>-2.72**</td>
</tr>
<tr>
<td>FACs(^c)</td>
<td>5.17</td>
<td>3.33</td>
<td></td>
<td>4.95</td>
<td>3.36</td>
<td></td>
<td>4.80</td>
</tr>
<tr>
<td>ALEKS score(^a) × 30 &amp; above(^b)</td>
<td>0.16</td>
<td>0.20</td>
<td></td>
<td>0.35</td>
<td>0.25</td>
<td></td>
<td>0.19</td>
</tr>
<tr>
<td>ALEKS score(^a) × FACs(^c)</td>
<td>0.37</td>
<td>0.31</td>
<td></td>
<td>0.38</td>
<td>0.32</td>
<td></td>
<td>0.37</td>
</tr>
<tr>
<td>30 &amp; Above(^b) × FACs(^c)</td>
<td>-5.25</td>
<td>3.61</td>
<td></td>
<td>-4.54</td>
<td>3.84</td>
<td></td>
<td>-5.21</td>
</tr>
<tr>
<td>ALEKS score(^a) × 30 &amp; above(^b) × FACs(^c)</td>
<td>-0.29</td>
<td>0.33</td>
<td></td>
<td>-0.41</td>
<td>0.40</td>
<td></td>
<td>-0.21</td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation section</td>
<td>4.88</td>
<td></td>
<td>4.78</td>
<td></td>
<td>4.69</td>
<td></td>
<td>3.69</td>
</tr>
<tr>
<td>Residual</td>
<td>86.51</td>
<td></td>
<td>88.24</td>
<td></td>
<td>88.58</td>
<td></td>
<td>90.19</td>
</tr>
<tr>
<td>Sample size</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Recitation sections</td>
<td>45</td>
<td></td>
<td>45</td>
<td></td>
<td>45</td>
<td></td>
<td>45</td>
</tr>
<tr>
<td>Participants</td>
<td>576</td>
<td></td>
<td>497</td>
<td></td>
<td>517</td>
<td></td>
<td>453</td>
</tr>
</tbody>
</table>

*Note.* Significance was determined by performing a Type III Analysis of Variance Table with Satterthwaite’s method.

\(^a\)ALEKS score was centered at 30.

\(^b\)0 = false, 1 = true.

\(^c\)pre-FACs = 0, FACs = 1.

\(^d\)T1 = 0, T2 = 1

\(*p < .05, \ **p < .01, \ ***p < .001.*"
Figure 9

*Estimated Means of Final Grade Percentage by ALEKS scores Around the Cutoff, Pre-FACs and FACs*

![Graph showing estimated means of final grade percentage by ALEKS scores around the cutoff for Pre-FACs and FACs.](image)

*Note.* Model 1 was fit on 576 participants in 45 recitation sections. Bands are ± 1 SE from the estimated marginal mean. The vertical dotted line represents an ALEKS of 29.5.

that three-way interactions are often underpowered to achieve significance (Moonseong & Andrew, 2010). Recall that crossovers and extreme cases can cause bias to the regression estimates in regression discontinuity. Therefore, further limiting the data offered insight into the differences between slopes at the discontinuity.

**Crossovers and Extreme Cases**

To provide unbiased regression estimates, regression discontinuity requires that the difference between students above an ALEKS score of 30 and below an ALEKS
score of 30 is due to the math placement score and nothing else. Thus, the crossovers seen in Figure 7 must be eliminated from the data as they can bias the estimates (Cunningham, 2021). Additionally, it is important to view the data close to the cutoff score—as narrowly as possible—while preserving as much data as needed to model it appropriately (Lesik, 2006). Plotting the LOESS (locally weighted least squares regression) curve on the data showed that the slopes of the regression estimates were influenced by the extreme ALEKS scores (see Appendix M). Thus, I made additional models to remove crossovers and extreme cases. Model 1 refers to the models fit to the full dataset with crossovers and extreme cases included. I removed the 79 extreme cases from the dataset and referred to subsequent models fit to this dataset as Model 2. I removed the 59 crossovers from the dataset and referred to subsequent models fit to this dataset as Model 3. Lastly, for the fourth dataset, I removed a total of 123 crossovers and/or extreme cases from the dataset and referred subsequent models fit to this dataset as Model 4. The best dataset according to the regression discontinuity specifications was the dataset fit as Model 4 because crossovers and extreme cases are removed. However, the calculations made on all four datasets allow for comparisons.

I followed the same process for these datasets as the full dataset to provide comparisons across the regression parameters. Thus, these datasets were fit to model EQ(full), the full equation described in Chapter III, with the two-way interactions. The regression parameters for Models 2-4 are shown with Model 1 in Table 8. Like Model 1, none of the interactions in Models 2-4 were significant. I then used the Type III Sums of Squares ANOVA using Satterthwaite’s method to sequentially test the fixed effects on
each model for significance. Models 2, 3, and 4, all had significance for the slope of the ALEKS scores \( b = 0.24, p < .001 \); \( b = 0.23, p < .001 \); \( b = 0.23, p < .001 \) and at the discontinuity \( b = -3.59, p = .021 \); \( b = -4.41, p = .013 \); \( b = -5.29, p = .002 \) but no significance was found on either of the two-way interactions.

The three-way interaction was fit to these data, and again, no interactions were significant (see Table 9). Similar to Model 1, the slope of the ALEKS scores for Models 2-4 were significant \( p < .001 \) and the discontinuity was significant for Models 2-4 \( b = -2.53, p = .013 \); \( b = -2.72, p = .006 \); \( b = -4.13, p = .001 \). After this, I ran simple slopes analyses for Models 2-4, shown in Appendix L, in which the outputs mimicked the output for Model 1. I continued to use the three-way interaction model for each of Models 2-4 to plot the marginal means for the final grade percentage on the ALEKS scores for pre-FACs and FACs groups. The plots for all four models are found in Appendix P. The change in slopes at the cutoff for the pre-FACs and FACs semesters warranted post hoc analyses.

**Post Hoc Analyses for the Change in Final Grade Percentage at the Cutoff**

Noting the significance of the main effects of the two variables (the ALEKS score and whether students were at or below the cutoff ALEKS score of 30), I ran post hoc analyses. I calculated the standardized mean differences (SMD) to measure the size of the change in the average final course percentages at the cutoff for all four models. The SMD give Cohen’s \( d \)-like effect sizes by dividing point differences by the pooled standard deviation. The pooled standard deviation was estimated by pooling the variance
components of the MLM model (Brysbaert & Stevens, 2018). Pairwise comparisons used Fisher's least significant difference (LSD) procedure, utilizing the Kenward-Roger method for degrees of freedom (Luke, 2017). Post hoc analyses revealed statistically significant differences in the average final grade percentages for the FACs semester for Models 2-4 before and after the discontinuity, see Table 10.

Table 10

*Standardized Mean Differences and Pairwise t Tests for Change in Final Grade Before and After the Cutoff (Less than 30 – At least 30) for all Four Models*

<table>
<thead>
<tr>
<th>Model</th>
<th>Group</th>
<th>SMD</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pre-FACs</td>
<td>.06</td>
<td>.22</td>
<td>512</td>
<td>.825</td>
</tr>
<tr>
<td></td>
<td>FACs</td>
<td>.74</td>
<td>1.85</td>
<td>538</td>
<td>.064</td>
</tr>
<tr>
<td>2</td>
<td>Pre-FACs</td>
<td>.22</td>
<td>.93</td>
<td>374</td>
<td>.351</td>
</tr>
<tr>
<td></td>
<td>FACs</td>
<td>.74</td>
<td>0.21</td>
<td>160</td>
<td>.033</td>
</tr>
<tr>
<td>3</td>
<td>Pre-FACs</td>
<td>.21</td>
<td>.76</td>
<td>281</td>
<td>.446</td>
</tr>
<tr>
<td></td>
<td>FACs</td>
<td>.83</td>
<td>2.09</td>
<td>299</td>
<td>.037</td>
</tr>
<tr>
<td>4</td>
<td>Pre-FACs</td>
<td>.40</td>
<td>1.42</td>
<td>166</td>
<td>.102</td>
</tr>
<tr>
<td></td>
<td>FACs</td>
<td>.91</td>
<td>0.48</td>
<td>125</td>
<td>.010</td>
</tr>
</tbody>
</table>

*Note.* Model 1 was fit on 576 participants in 45 recitation sections. Model 2 was fit to the data with 79 extreme cases removed on 497 observations nested in 45 recitation sections. Model 3 was fit to the data with 59 crossovers removed on 517 participants nested in 45 recitation sections. Model 4 was fit to the data with crossovers and extreme cases removed on 453 participants nested in 45 recitation sections. Pair-wise t tests utilize Fisher’s LSD and are unadjusted p-values. Degrees of freedom utilize the Kenward-Roger method.

The pre-FACs and FACs semesters saw changes in course achievement across the cutoff. For the pre-FACs semesters, the students who scored just below 30 on the ALEKS math placement exam performed better in the course on average than those who scored 30 on the ALEKS. For Model 4, this difference is shown by the SMD of .40 in Table 10, which is not significant (p = .102). Using the parameter estimates in Table 9, this
difference in means for pre-FACs semesters across the cutoff can be quantified. For the pre-FACs students who scored an ALEKS of 29, the average final grade percentage was 81.65%, and for those who scored an ALEKS of 30, the average final grade percentage was 77.61%, a difference of 4.04%.

For the FACs semester, Model 4 in Table 10 shows that the mean difference of the final grade percentages across the cutoff was statistically significant ($SMD = 0.91$, $p = .010$). Thus, the SMD of .91 implies that the average standardized difference in final grade percentages for students experiencing FACs was nearly one standard deviation higher for those students scoring 29 on the ALEKS placement exam than those who scored 30. Although pre-FACs semesters’ difference in course achievement across the cutoff was not significant, it was statistically significant for the FACs semester. Figure 10 is a visualization of these changes at the cutoff for Model 4.

The change in course achievement from pre-FACs to FACs semesters was not significant. Table 11 shows the effect sizes from pre-FACs to FACs semesters at the cutoff. Below an ALEKS of 30 represents Stat 1045 students and ALEKS scores at least 30 are those in Stat 1040. None of the pairwise $t$-tests were significant. Although the Stat 1045 course’s differences from their course grade from pre-FACs to FACs semesters had a medium effect size for Model 4, it was not statistically significant ($SMD = .52$, $p = .159$).

Figure 11 visualizes these differences in final grade percentages in Stat 1040 and Stat 1045 from pre-FACs to FACs semesters at the cutoff for Model 4. These differences can also be quantified. FACs made no significant change in the final course percentage
Figure 10

Side by Side Plots of Final Grade Percentage at the Cutoff, Pre-FACs to FACs Semester for Model 4

Note. Model 4 was fit to the data with crossovers and extreme cases removed on 453 participants nested in 45 recitation sections. Bands are ± 1 SE from the estimated marginal mean. The vertical dotted line represents an ALEKS of 29.5.

for Stat 1040 students at an ALEKS of 30, as the achievement for pre-FACs (red line) and FACs (blue line) at an ALEKS of 30 is a difference of 0.04%. This difference is verified by the SMD of .01 (p = .968) for the “At Least 30” group of Model 4 in Table 11. However, the “Below 30” group shows a potential difference in slopes pre-FACs (red line) to FACs (blue line). From Table 11, the SMD shows this difference was about .52 standard deviations in the final grade percentage but is not significant (SMD = .52, p = .159). Although this difference is not statistically significant, using the parameter estimates in Table 9, this difference can also be realized. In the pre-FACs semesters, for
Table 11

*Standardized Mean Differences (Effect Sizes) and Pairwise t Tests for the Difference in Final Grade Change from Pre-FACs to FACs (Pre-FACs – FACs) for all Four Models*

<table>
<thead>
<tr>
<th>Model</th>
<th>ALEKS group</th>
<th>SMD</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Below 30</td>
<td>.71</td>
<td>1.50</td>
<td>495</td>
<td>.134</td>
</tr>
<tr>
<td></td>
<td>At least 30</td>
<td>.03</td>
<td>0.20</td>
<td>77</td>
<td>.840</td>
</tr>
<tr>
<td>2</td>
<td>Below 30</td>
<td>.56</td>
<td>1.45</td>
<td>330</td>
<td>.145</td>
</tr>
<tr>
<td></td>
<td>At least 30</td>
<td>.04</td>
<td>0.21</td>
<td>156</td>
<td>.834</td>
</tr>
<tr>
<td>3</td>
<td>Below 30</td>
<td>.63</td>
<td>1.40</td>
<td>360</td>
<td>.164</td>
</tr>
<tr>
<td></td>
<td>At least 30</td>
<td>.02</td>
<td>0.12</td>
<td>60</td>
<td>.910</td>
</tr>
<tr>
<td>4</td>
<td>Below 30</td>
<td>.52</td>
<td>1.41</td>
<td>227</td>
<td>.159</td>
</tr>
<tr>
<td></td>
<td>At least 30</td>
<td>.01</td>
<td>0.04</td>
<td>135</td>
<td>.968</td>
</tr>
</tbody>
</table>

*Note.* Model 1 was fit on 576 participants in 45 recitation sections. Model 2 was fit to the data with 79 extreme cases removed on 497 observations nested in 45 recitation sections. Model 3 was fit to the data with 59 crossovers removed on 517 participants nested in 45 recitation sections. Model 4 was fit to the data with crossovers and extreme cases removed on 453 participants nested in 45 recitation sections. Pair-wise t tests utilize Fisher’s LSD and are unadjusted p values. Degrees of freedom utilize the Kenward-Roger method. No p values were significant at α = .05.

an ALEKS score of 29, the average final grade percentage was 81.65%. However, students with the same ALEKS score of 29 who experienced FACs scored 86.08% in the class, a difference of about 4.43%. This difference is about one-half of a standard deviation for the overall final grade percentage (SD = 9.83%, see Table K). Indeed, final grade percentages were not statistically significant at the cutoff from pre-FACs to FACs semesters. Still, the difference in average final grade percentages between pre-FACs to FACs semesters for those who scored below 30 on ALEKS could be meaningful.
Figure 11

*Estimated Means of Final Grade Percentage by ALEKS Scores Around the Cutoff for Model 4*

![Graph showing estimated means of final grade percentage by ALEKS scores around the cutoff.](image)

*Note.* Model 4 was fit to the data with crossovers and extreme cases removed on 453 participants nested in 45 recitation sections. Bands are ± 1 SE from the estimated marginal mean. The vertical dotted line represents an ALEKS of 29.5.

**Summary**

Multilevel modeling was appropriate to model the change in final grade percentages from the pre-FACs to FACs semesters for different ALEKS math placement exam scores using regression discontinuity methodology. The three-way interaction model was fit to the data by deleting crossovers and extreme scores by minimizing the range of ALEKS scores to about 15 points on either side of the exogenous cutoff score, which places students below 30 into Stat 1045 and otherwise into Stat 1040. Model 4
found no interaction between the variables of the ALEKS score, the discontinuity, and whether the student was in the FACs semester. However, potential differences in average final grade percentages were seen before and after the cutoff and from pre-FACs to FACs semesters. Although there is an average improvement of 4.5% in final grade percentages from pre-FACs to FACs semesters for students who scored 29 on the ALEKS placement exam, it was not statistically significant (SMD = .52, p = .159). Significant differences in average final grade percentages were discovered at the cutoff in the FACs group. For an ALEKS score of 29, the Stat 1045 FACs group saw a significant average improvement of 8.5% in their final grade percentage compared to the average FACs Stat 1040 student with an ALEKS of 30 (SMD = .91, p = .01). The only significant difference found in course achievement was at the cutoff for the FACs semester. Although there was no statistical significance for average final grade percentages at the cutoff from pre-FACs to FACs semesters, the difference may still be meaningful and is discussed in the next chapter.

Analysis for Research Question 2

Research Question 2 investigated potential improvement in students’ attitudes toward statistics in large-enrollment introductory statistics courses with embedded Formative Assessment Cycles (FACs). I explain and provide the participants’ descriptive and summary statistics in this section. Then, I report the Cronbach’s alpha values that provided a measure of the internal consistency before interpreting the scores from my sample, followed by a visual assessment of the data through histograms and person-
profile plots. Next, using multilevel modeling (MLM), I discuss the analysis of each attitude component of the SATS-36 (Schau, 2003). Subsequently, I provide plots of the change in attitude scores from pre-to post-survey and discuss the practical and statistical significance of the effect sizes calculated. Then, I devote a section to the analysis of students who took the ALEKS math placement exam before the semester to determine whether their mathematical knowledge before the class impacted their attitudes. Finally, a summary of the findings for the second research question completes this section. The complete analysis in an R markdown file is available upon request.

**Descriptive and Summary Statistics**

I taught the two large-enrollment introductory statistics sections of Stat 1040 and Stat 1045 in Fall 2021 and Spring 2022 semesters. Students in both sections experienced FACs in their homework, quizzes, and examinations. These course assessments were identical in both sections and semesters. All students had the opportunity to take the SATS-36 pre-survey during the first two weeks of the semester and the SATS-36 post-survey during weeks 12 and 13 of the 15-week semester. Of the 531 enrolled students, 66 withdrew or failed, leaving 465 students (88%) who completed the semester and received the full FACs intervention. Approximately 95% of the 465 students completed at least one of the two surveys, and only 5% ($n = 24$) failed to respond, leaving 441 participants for analysis, see Table 12.

Table 13 displays the descriptive and summary statistics for the students who received the full FACs intervention. Of the 465 students, the majority were female (76.8%), which was significant associated with participation in the surveys ($p = .001$).
Table 12

*Students Participation in the SATS-36 Survey: Total and After Removing Student Responses Who Withdrew or Failed*

<table>
<thead>
<tr>
<th>Survey response</th>
<th>n</th>
<th>Percentage (N = 531)</th>
<th>Percentage (N = 465)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre Only</td>
<td>57</td>
<td>10.75</td>
<td>12.26</td>
</tr>
<tr>
<td>Pre &amp; Post</td>
<td>356</td>
<td>67.04</td>
<td>76.56</td>
</tr>
<tr>
<td>Post Only</td>
<td>28</td>
<td>5.27</td>
<td>6.02</td>
</tr>
<tr>
<td>Neither</td>
<td>24</td>
<td>4.52</td>
<td>5.16</td>
</tr>
<tr>
<td>Student withdrew/failed</td>
<td>66</td>
<td>12.43</td>
<td>--</td>
</tr>
<tr>
<td>Total</td>
<td>531</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

Most students who received the FACs intervention were from the fall semester (59.1%) and Stat 1040 (65.6%). Both the semester and course were not associated with participation in the surveys ($p = .390; p = .597$).

Freshmen accounted for 44.3% of students, while sophomores accounted for 34.6%. Juniors and seniors accounted for a total of 21% of the students. The average age of the students was 19.79 ($SD = 2.66$). Both class rank and age were insignificant ($p = .248; p = .125$). The final grade in the course averaged 83.86% ($SD = 9.46$%), which was significantly different for survey participation ($p < .001$). Both significant demographic and summary variables, sex and final grade, were important variables in the multilevel model analyses as a main factor and an interaction, respectively.

More descriptive and summary statistics are found in Appendix O and P. Appendix O shows the descriptive and summary statistics of the breakdown of the prerequisites the students used for placement into Stat 1040 and Stat 1045. The prerequisite was not significantly associated with survey participation ($p = .670$). Appendix
Table 13

Descriptive and Summary Statistics of the Participation in the Surveys

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 465)</th>
<th>Pre only (n = 57)</th>
<th>Post only (n = 28)</th>
<th>Pre &amp; post (n = 356)</th>
<th>Neither (n = 24)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M</td>
<td>SD</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>108</td>
<td>23.2</td>
<td>19</td>
<td>33.3</td>
<td>10</td>
<td>35.7</td>
</tr>
<tr>
<td>Female</td>
<td>357</td>
<td>76.8</td>
<td>38</td>
<td>66.7</td>
<td>18</td>
<td>64.3</td>
</tr>
<tr>
<td>Semester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>275</td>
<td>59.1</td>
<td>38</td>
<td>66.7</td>
<td>14</td>
<td>50.0</td>
</tr>
<tr>
<td>Spring</td>
<td>190</td>
<td>40.9</td>
<td>19</td>
<td>33.3</td>
<td>14</td>
<td>50.0</td>
</tr>
<tr>
<td>Course</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stat 1040</td>
<td>305</td>
<td>65.6</td>
<td>34</td>
<td>59.6</td>
<td>18</td>
<td>64.3</td>
</tr>
<tr>
<td>Stat 1045</td>
<td>160</td>
<td>34.4</td>
<td>23</td>
<td>40.4</td>
<td>10</td>
<td>35.7</td>
</tr>
<tr>
<td>Class rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>206</td>
<td>44.3</td>
<td>25</td>
<td>43.9</td>
<td>11</td>
<td>39.3</td>
</tr>
<tr>
<td>Sophomore</td>
<td>161</td>
<td>34.6</td>
<td>22</td>
<td>38.6</td>
<td>11</td>
<td>39.3</td>
</tr>
<tr>
<td>Junior</td>
<td>63</td>
<td>13.5</td>
<td>6</td>
<td>10.5</td>
<td>4</td>
<td>14.3</td>
</tr>
<tr>
<td>Senior</td>
<td>35</td>
<td>7.5</td>
<td>4</td>
<td>7</td>
<td>2</td>
<td>7.1</td>
</tr>
<tr>
<td>Age</td>
<td>19.79</td>
<td>2.66</td>
<td>20.02</td>
<td>2.18</td>
<td>19.75</td>
<td>1.86</td>
</tr>
<tr>
<td>Final grade %</td>
<td>83.86</td>
<td>9.46</td>
<td>80.93</td>
<td>10.51</td>
<td>77.59</td>
<td>9.41</td>
</tr>
</tbody>
</table>
P displays the descriptive and summary statistics for the subset of the students \((n = 274)\) who used the ALEKS math placement exam score as their prerequisite. Like Table 13, the significant factors across survey participation are sex \((p = .025)\) and final grade \((p < .001)\).

For those students who identified their major on the pre-survey, Appendix Q displays the self-reported majors across the participants. Major was not a significant factor across survey participation \((p = .43)\). Appendix R provides the breakdown of students’ participation in the surveys by the 13 recitation instructors (see Table R.1). The breakdown of students by recitation instructor (TA) for the students who did participate in the surveys is shown in Table R2. The distribution of students by TA was not significant \((p = .276)\).

**Internal Consistency**

I computed Cronbach’s alphas on the sample to identify if any attitude components did not meet the generally approved level of 0.70 (Adeniran, 2019). A Cronbach’s alpha over .70 shows that the components are internally consistent (see Table 14).

**Table 14**

*Cronbach’s Alphas for the SATS-36 Attitude Survey by Attitude Component*

<table>
<thead>
<tr>
<th>Component</th>
<th>Pre-Survey</th>
<th>Post-Survey</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td>.792</td>
<td>.870</td>
</tr>
<tr>
<td>Cognitive Competence</td>
<td>.828</td>
<td>.861</td>
</tr>
<tr>
<td>Value</td>
<td>.881</td>
<td>.892</td>
</tr>
<tr>
<td>Difficulty</td>
<td>.740</td>
<td>.781</td>
</tr>
<tr>
<td>Interest</td>
<td>.886</td>
<td>.920</td>
</tr>
<tr>
<td>Effort</td>
<td>.681</td>
<td>.711</td>
</tr>
</tbody>
</table>
The pre-survey Cronbach alpha values for all attitude components except for *effort* were in the acceptable range of .74-.89. The post-survey alpha values were all above .71.

**Exploratory Data Analysis for Attitude Components**

Research Question 2 investigated potential improvement in students’ attitudes toward statistics in large-enrollment introductory statistics courses with embedded FACs. Student attitude scores from the SATS-36 (Schau, 2003) pre-and post-surveys measured in Fall 2021 and Spring 2022 provided in the pre-existing dataset were analyzed. The six components of attitudes are *affect, interest, cognitive competence, value, difficulty,* and *effort*. All attitudes were measured on a 7-point Likert scale.

Before performing MLM analysis on each attitude component to determine whether the attitudes changed over time, I explored the data by graphing histograms of each attitude component to visualize any change from pre- to post-survey. The histograms for each component are displayed in Appendix S. Additionally, person-profile plots, also called spaghetti plots, allowed me to see the individual participant’s pre- to post-scores by the recitation instructor and the mean change by the recitation instructor. The person-profile plots for each attitude component are displayed in Figures 12-17. Recitation TA, the nesting variable in the MLM analysis, is the grouping variable for the person-profile plots. The differences in slopes from recitation instructor to recitation instructor (the black dotted lines) by attitude component show that MLM is appropriate for this analysis as MLM will be able to control for the dependent relationships between students in a recitation section.
Figure 12

Affect Person-Profile Plot by TA
Figure 13

Cognitive Competence Person-Profile Plot by TA
Figure 14

Difficulty Person-Profile Plot by TA
Figure 15

Value Person-Profile Plot by TA
Figure 16

Interest Person-Profile Plot by TA
Figure 17

Effort Person-Profile Plot by TA
Finally, overall means and standard deviations by attitude component and correlations between pre- and post-scores are displayed in Table 15. All correlations were statistically significant at the $p < .001$ level.

**Table 15**

*Summary of Dependent Variables*

<table>
<thead>
<tr>
<th>Attitude component</th>
<th>Pre score ($n = 413$)</th>
<th>Post score ($n = 384$)</th>
<th>Pre vs. post ($n = 356$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$M$</td>
<td>$SD$</td>
<td>$M$</td>
</tr>
<tr>
<td>Affect</td>
<td>3.84</td>
<td>1.02</td>
<td>4.49</td>
</tr>
<tr>
<td>Cognitive competence</td>
<td>4.69</td>
<td>1.00</td>
<td>5.22</td>
</tr>
<tr>
<td>Difficulty</td>
<td>3.40</td>
<td>0.75</td>
<td>4.01</td>
</tr>
<tr>
<td>Effort</td>
<td>6.64</td>
<td>0.48</td>
<td>6.12</td>
</tr>
<tr>
<td>Interest</td>
<td>4.87</td>
<td>1.25</td>
<td>4.46</td>
</tr>
<tr>
<td>Value</td>
<td>5.28</td>
<td>0.96</td>
<td>5.05</td>
</tr>
</tbody>
</table>

***$p < .001$.

**MLM Analyses of the Six Components of the SATS-36**

Analyses of the six components of attitudes toward statistics were completed via six individual MLMs (see Table 16). The “attempt” variable in the model is the time point: pre- or post-survey. Attempt was statistically significant in every model, noting significant change in the attitude component over the semester. After accounting for person-to-person variability, the baseline models for affect, effort, and value were further nested by the recitation instructor. The baseline model for cognitive competence was the only model to be further nested by course (Stat 1040 or Stat 1045). The MLM baseline models found that 35.7% to 57.8% of the variance in the pre-survey scores was attributable to student-to-student variability and further nesting in the models, given by the
Table 16

Baseline Models for Multilevel Models for Attitude Components

<table>
<thead>
<tr>
<th>Effect</th>
<th>Affect</th>
<th>Cognitive competence</th>
<th>Difficulty</th>
<th>Effort</th>
<th>Interest</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>SE</td>
<td>Var</td>
<td>b</td>
<td>SE</td>
<td>Var</td>
</tr>
<tr>
<td>Intercept</td>
<td>3.8***</td>
<td>0.08</td>
<td>4.6***</td>
<td>0.20</td>
<td>3.4***</td>
<td>0.04</td>
</tr>
<tr>
<td>Attempta</td>
<td>0.7***</td>
<td>0.06</td>
<td>0.5***</td>
<td>0.06</td>
<td>0.6***</td>
<td>0.05</td>
</tr>
</tbody>
</table>

Random effects

<table>
<thead>
<tr>
<th>Course</th>
<th>0.06</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA</td>
<td>0.03</td>
</tr>
<tr>
<td>0.09</td>
<td>0.27</td>
</tr>
<tr>
<td>0.15</td>
<td>0.41</td>
</tr>
<tr>
<td>0.15</td>
<td>0.29</td>
</tr>
<tr>
<td>0.77</td>
<td>0.41</td>
</tr>
</tbody>
</table>

ICC

<table>
<thead>
<tr>
<th>.410</th>
<th>.394</th>
</tr>
</thead>
<tbody>
<tr>
<td>.394</td>
<td>.394</td>
</tr>
<tr>
<td>.357</td>
<td>.578</td>
</tr>
<tr>
<td>.394</td>
<td></td>
</tr>
</tbody>
</table>

Note. Models for affect, cognitive competence, difficulty, and effort were fit on 441 participants and 797 observations. Model for interest was fit on 441 participants and 796 observations. Model for value was fit on 441 participants and 795 observations.

ICC = Intraclss correlation.

a0 = pre-survey, 1 = post-survey.

***p < .001.
intraclass correlation (ICC) in Table 16. This variability is visually depicted in the high degree of the vertical spread between lines compared to trend lines in the person-profile plots found in Figures 12-17.

The variances of the random effects are displayed in Table 16. Trying to account for recitation instructor variability in the difficulty and interest models caused a non-unique solution (i.e., a singularity issue). Thus, the models for difficulty and interest accounted for student-to-student variability only. Models for affect, effort, and value were further nested by recitation instructor, and the model for cognitive competence allowed further nesting by course (Stat 1040 or Stat 1045). Figure 18 displays the person-profile plot by course for cognitive competence, which shows the difference in the trend lines

**Figure 18**

*Person-Profile Violin Plot by Course for Cognitive Competence*

![Person-Profile Violin Plot by Course for Cognitive Competence](image)

*Note.* The thick black line is the course average of cognitive competence from pre- to post-survey. Gray lines represent individual students pre- to post-survey scores. The “violin” outlines show the shape of distributions—the wider the violin the denser the scores.
between Stat 1040 and Stat 1045. Person-profile violin plots of the remaining attitude components by course are shown in Appendix T.

Then, using a bottom-up approach, I determined the full MLM models by eventually trying all possible fixed effects and combinations of fixed effects as interactions to see if the variables of age, sex, major, course, semester, and final grade were significant main or moderating effects. The final models for each component were determined via testing pairs of models by likelihood ratio tests (LRT). If insignificant results were found, I used the smallest Akaike’s Information Criterion (AIC) of the possible models. The full MLM models are displayed in Table 17.

For all six components of attitude, student performance significantly moderated the change in attitude from pre- to post-survey. Additionally, sex was a significant main factor for each attitude component except effort. The variability between semesters was controlled for in the value model only. Value decreased the spring semester on average by 0.2 points ($p < .05$); semester was not a covariate in the other models. Figure 12 displays the plots of the estimated marginal mean changes in attitudes moderated by the course performance to visualize the interaction between attempt and final grade.

Figure 19 shows the significant change from pre- to post-survey, moderated by the students’ final grades for all attitude components. At the pre-survey, attitudes were similar on average. However, at the post-survey, students’ performance in the course affected their attitudes. For the attitudes that significantly increased over time, the students on track to score an A in the course had the greatest gain in attitude scores (as seen in the top three panels of Figure 19). Conversely, for attitude components that decreased over time, the
### Table 17

**Final Multilevel Models for Attitude Components**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Affect</th>
<th>Cognitive competence</th>
<th>Difficulty</th>
<th>Effort</th>
<th>Interest</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>SE</td>
<td>Var</td>
<td>Size</td>
<td>b</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.34***</td>
<td>0.10</td>
<td></td>
<td>0.14</td>
<td>4.97***</td>
<td>0.07</td>
</tr>
<tr>
<td>Attempt</td>
<td>0.65***</td>
<td>0.06</td>
<td></td>
<td>0.05</td>
<td>0.54***</td>
<td>0.04</td>
</tr>
<tr>
<td>Final grade</td>
<td>0.01</td>
<td>0.01</td>
<td></td>
<td>0.00</td>
<td>0.01***</td>
<td>0.00</td>
</tr>
<tr>
<td>Sex</td>
<td>-0.61***</td>
<td>0.10</td>
<td></td>
<td>0.09</td>
<td>-0.36***</td>
<td>0.08</td>
</tr>
<tr>
<td>Attempt x</td>
<td>0.06***</td>
<td>0.01</td>
<td></td>
<td>0.01</td>
<td>0.06***</td>
<td>0.01</td>
</tr>
<tr>
<td>Final grade</td>
<td>0.06***</td>
<td>0.01</td>
<td></td>
<td>0.01</td>
<td>0.06***</td>
<td>0.01</td>
</tr>
<tr>
<td>Semester</td>
<td>-0.20</td>
<td>0.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Random effects

<table>
<thead>
<tr>
<th>Course</th>
<th>0.02</th>
</tr>
</thead>
<tbody>
<tr>
<td>TA</td>
<td>0.02</td>
</tr>
<tr>
<td>Student</td>
<td>0.43</td>
</tr>
<tr>
<td>Residual</td>
<td>0.62</td>
</tr>
</tbody>
</table>

Participants 441 441 441 441 441 441 441
Observations 797 797 797 797 796 795

a0 = pre, 1 = post.
bFinal grade was centered at 85.
c0 = male, 1 = female.
d0 = fall, 1 = spring.

*p < .05, **p < .01, ***p < .001.
Figure 19

*Estimated Marginal Means for Males’ Changes in Attitude Components from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95%*

Students on track to score a C had the greatest decline in attitude scores (as seen in the bottom three panels of Figure 19). These plots illustrate the significant moderation of students’ final grade on the change in students’ attitudes from pre- to post-survey.

Consider *affect* in the upper lefthand corner of Figure 19 and the model parameters for *affect* in Table 17. A male student, whose final grade was 85% in the
course (final grade was centered at 85% in the analysis), had an average affect score of 4.34 ($M = 4.34, p < .001$) at the pre-survey. As Figure 19 shows, the pre-survey mean scores by final grade were similar. However, affect significantly improved in the post-survey based on the student's final grade. For each 1% increase above 85% in the final grade, affect increased by $b = 0.063 (p < .001)$. However, the overall average change for affect is $b = 0.65 (p < .001)$. Similarly, cognitive competence increased from pre- to post-survey on average by $b = 0.54 (p < .001)$ and increased by $b = 0.058 (p < .001)$ for each 1% increase above 85% in the final grade, as seen in the middle top panel of Figure 19. Finally, difficulty, the top righthand plot of Figure 19, had an overall mean increase in attitude from pre- to post-survey of $b = 0.61 (p < .001)$. This change was also significantly moderated by the final grade, such that for each 1% increase in final grade over 85%, the difficulty score increased by $b = .03 (p < .001)$. The A students showed significant positive change in the attitude components of affect, cognitive competence, and difficulty. The C students showed no significant change in affect and cognitive competence but did increase significantly for difficulty.

Three attitude components significantly decreased on average from pre- to post-survey: effort ($b = -0.52, p < .001$), interest ($b = -0.40, p < .001$), and value ($b = -0.23, p < .001$), and these changes in attitude from pre- to post-survey were also significantly moderated by final grade. For every 1% in the students’ final grade percentage below 85%, the post-survey effort scores fell an additional $b = 0.03, (p < .001)$, interest fell an additional $b = 0.04, (p < .001)$, and value fell an additional $b = 0.02, (p < .001)$. The bottom three plots of Figure 19 display the pre- to post-survey change in effort, interest,
and *value* scores, respectively. For *interest* and *value*, the A students showed no significant change from pre- to post-survey but decreased significantly for *effort*. However, C students significantly decreased in these three attitude components.

Sex was a significant covariate in every model except *effort*, see Table 17. Female students whose final grade was a B in the course scored on average 0.61 points lower than male students on *affect* pre-survey ($M = 3.7, p < .001$). For *cognitive competence*, female students averaged 0.36 lower than male students at the pre-survey ($M = 4.61, p < .001$). For *difficulty, interest, and value*, female students each averaged 0.17, 0.55, and 0.30 lower at pre-survey than male students, respectively ($M = 3.37, p < .05; M = 4.77, p < .001; M = 5.31, p < .01$). However, for *effort*, female students averaged .09 points higher at pre-survey than male students ($M = 6.73, p > .05$). Appendix U displays the estimated marginal mean effect for females for the six attitude components, and side-by-side plots of each attitude component’s marginal mean effect by sex can be seen in Appendix V.

**Effect Sizes for Change in Attitude Components from Pre- to Post-Survey**

Having found significant changes for all attitude components moderated by final grade percentage, standardized mean differences (SMD) were computed and displayed in Table 18. The SMD gives Cohen’s $d$-like effect sizes in which mean differences are divided by the pooled standard deviation. The pooled standard deviation was estimated by pooling the variance components of the best fitting MLM (Brysbaert & Stevens, 2018). Corresponding correlations are provided in Table 15 for the ability to compute
Table 18

Standardized Mean Differences (Effect Sizes) and Pairwise t tests for Change in Pre- to Post-Survey Attitudes by Final Grade

<table>
<thead>
<tr>
<th>Component</th>
<th>Final grade %</th>
<th>SMD</th>
<th>t</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.02</td>
<td>0.24</td>
<td>434</td>
<td>.812</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.63</td>
<td>11.38</td>
<td>397</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>1.24</td>
<td>14.98</td>
<td>388</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Cognitive competence</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.04</td>
<td>0.50</td>
<td>438</td>
<td>.653</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.57</td>
<td>10.11</td>
<td>400</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>1.18</td>
<td>13.96</td>
<td>390</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Difficulty</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.36</td>
<td>4.25</td>
<td>434</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.76</td>
<td>13.78</td>
<td>397</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>1.17</td>
<td>14.14</td>
<td>388</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Effort</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>1.20</td>
<td>13.76</td>
<td>441</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.81</td>
<td>13.86</td>
<td>401</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>0.41</td>
<td>4.72</td>
<td>392</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>Interest</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.59</td>
<td>8.09</td>
<td>416</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.31</td>
<td>6.43</td>
<td>386</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>0.03</td>
<td>0.39</td>
<td>379</td>
<td>.390</td>
<td></td>
</tr>
<tr>
<td>Value</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>75</td>
<td>0.45</td>
<td>6.41</td>
<td>412</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>85</td>
<td>0.23</td>
<td>5.01</td>
<td>383</td>
<td>&lt;.001</td>
<td></td>
</tr>
<tr>
<td>95</td>
<td>0.01</td>
<td>0.18</td>
<td>376</td>
<td>.861</td>
<td></td>
</tr>
</tbody>
</table>

Note. Pair-wise t tests utilize Fisher’s LSD and are unadjusted p values. Degrees of freedom utilize the Kenward-Roger method. All significant p values maintain statistical significance when adjusted by the Bonferroni correction of 18(p).

Cohen’s $d_m$ for future meta-analyses (Lakens, 2013). Post hoc pairwise t tests were computed utilizing Fisher’s LSD with the Kenward-Roger method for the degrees of freedom (Luke, 2017). The p values are unadjusted. Because the SMD is a standardized effect size for the difference in means, the SMD can be interpreted qualitatively. Cohen
(1997) used general benchmarks for the strength of an effect size. A small effect size was $d < 0.5$, medium was $0.5 < d < 0.7$, and large was $d \geq 0.8$. Cohen’s scale is used in the literature to measure effect sizes in students’ attitudes toward statistics from pre- to post-survey as small, medium, or large (e.g., Emmioğlu & Capa-Aydin, 2013; Schau & Emmioğlu, 2012). Additionally, Millar and Schau (2010) defined a practically significant change in attitudes if the average change from pre- to post-survey was at least 0.5, or a half-point. Schau and Emmioğlu also used this measure for practical significance. Schau (2003) defined “neutral attitudes” as those averages of 3.5 to 4.5, 0.5 from the average of 4 on the 7-point Likert scale. The analysis of the effect sizes and their practical significance are discussed for each attitude component using these constructs.

**Attitudes that increased post-survey, moderated by final grade.** *Affect* measures students’ feelings toward statistics: whether they like or fear statistics, enjoy or feel stressed in class, and feel insecure or frustrated doing statistics problems ([http://evaluationandstatistics.com](http://evaluationandstatistics.com)). (Negative feelings were negatively scored.) Table 17 shows that *affect* had a statistically and practically significant average change from pre-to post-survey ($b = 0.65, p < .001$). For students who would earn As in the course, Table 18 shows a large effect size of 1.24 ($p < .001$). The B students saw a medium effect size for an increase in *affect* ($SMD = .63, p < .001$). However, the C students did not experience a significant change in *affect* from pre- to post-survey ($SMD = .02, p > .05$). Thus, the A and B students experienced a practically and statistically significant positive increase in the *affect* component.

*Cognitive competence* measures students’ beliefs about their intellectual skills and
abilities, whether they expect to feel lost, confused, or make many mistakes (these are negatively scored, http://evaluationandstatistics.com). Table 17 shows that, on average, the B students significantly increased statistically and practically in cognitive competence ($b = 0.54, p < .001$). The A students’ average increased by 1.14 points, a large effect size ($SMD = 1.18, p < .001$). However, the C students experienced a nonsignificant change ($SMD = 0.04, p > .05$). The C students’ average scores at post-survey were still higher than neutral (over 4.5), thus suggesting a higher-than-neutral belief in their abilities to make statistical computations correctly. The A and B students’ cognitive competence statistically and practically significantly improved over the semester.

**Difficulty** measures students’ beliefs in how complicated it is to learn statistics. Some questions on the SATS-36 (Schau, 2003) ask students whether they believe statistics is more complicated, technical, and difficult, and these were negatively scored. The average difficulty pre-survey score for students who earned a B was 3.54 and improved on average by 0.61 ($p < .001$). This mean improvement was statistically and practically significant. The C students’ effect size was small ($SMD = 0.36, p < .001$), the B students’ effect size was medium ($SMD = 0.76, p < .001$), and the A students effect size was large ($SMD = 1.17, p < .001$). This positive change is visually depicted in the upper right plot in Figure 19. All students’ difficulty components significantly improved on average over the semester.

**Attitudes that decreased post-survey, moderated by final grade.** **Effort** assesses students’ plans to study, work hard, attend every class, and complete all their assignments (http://evaluationandstatistics.com). Initially, students’ expectation of their
effort was high, at an average of 6.67. All students’ effort scores dropped on average by $b = 0.52$ ($p < .001$) from pre- to post-survey, a statistically and practically significant change (see Table 17). The B and C students’ average effort scores significantly decreased as seen by the large effect sizes from Table 18 ($SMD = .81, p < .001$; $SMD = 1.20, p < .001$). The A students’ average change in effort also significantly decreased in at post-survey, although the effect size was small ($SMD = 0.41, p < .001$). No matter their final grade, all students had significant decreases in effort from pre- to post-survey.

The other two attitude components that decreased significantly from pre- to post-survey were interest ($b = -0.40, p < .001$) and value ($b = -0.23, p < .001$), neither of which were considered practically significant. As shown in the plots of Figure 19, the C students had the greatest decline in these two attitude components. Figure 19 also shows that the A students experienced no statistically significant change in these attitude components, which is also shown by their effect sizes, $SMD = 0.03$ ($p > .05$) for interest and $SMD = 0.01$ ($p > .05$) for value. Effect sizes by students’ final grades for value showed small effect sizes for both the C students ($SMD = 0.45, p < .001$) and the B students ($SMD = 0.23, p < .001$). For interest, the C students’ effect size was medium ($SMD = 0.59, p < .001$), and the B students’ effect size was small ($SMD = 0.31, p < .001$). The average decrease in student attitudes was not practically significant for both interest and value. For the C students, interest and value fell from high average scores to neutral attitudes from pre- to post-survey (see Figure 19).
Considering the Impact of Students’ ALEKS Math Placement Scores on Attitudes

The subset of the data from those students who used the ALEKS placement test as their prerequisite was also analyzed similarly to see if ALEKS placement scores were a significant factor in student attitudes. The baseline models were all significant for the sample of ALEKS-takers. The full models were fitted using a bottom-up approach, and parameters were estimated. However, the ALEKS scores did not significantly impact student attitudes except in the model for cognitive competence. In the cognitive competence model, for each point increase in ALEKS score, the cognitive competence post-survey score increased by $b = .01$ ($p < .01$), see Appendix W. The other finals models for the ALEKS subset of data were not different from the models fit with all the data.

Students’ final grades significantly moderated all the attitude components’ final models’ post-survey scores. Most of the attitude component models could not incorporate further nesting of the students by the recitation instructor due to sparsity. The parameter estimates for the final MLM models for the data in the ALEKS subset can be viewed in Appendix W. Due to the similarities between the final MLM models between the ALEKS subset and the full data, the inability of the ALEKS subset to nest by recitation instructor, and the ALEKS scores producing no significant impact concerning student attitudes other than cognitive competence, there was no need to further investigate prior mathematical knowledge as an independent variable in the analysis.
Summary

Multilevel modeling was appropriate for modeling the change in SATS-36 (Schau, 2003) attitude components pre-to post-survey. MLM was able to account for dependence among observations. Random intercepts were fit to accommodate further nesting of students by their recitation instructor. Affect, cognitive competence, and difficulty showed statistically significant improvement pre-to post-survey, with the students’ final grades moderating the improvement. Students performing well (85% and above) had better attitudes pre- to post-survey than those just passing the course. Female students’ attitude scores were significantly lower at baseline in these attitude components than male students and practically significant for affect. All average improvements in attitude were also practically significant, increasing by over a half-point on average pre- to post-survey.

The three attitude components that significantly decreased pre- to post-survey were effort, interest, and value. The significance was also moderated by the students’ final grades in the course. The greatest average declines in attitude components over time were seen in the C students. The smallest and insignificant average decreases were seen in the A students. Female students’ scores were significantly lower on average than male students for interest and value and were not significantly different for effort. Only the average decrease in effort for all students was considered practically significant. The discussion of possible reasons behind these significant changes in attitudes is discussed in Chapter V.
CHAPTER V
DISCUSSION

“But all can be judged by the same criteria: an open process, beginning with goals, that measures and enhances students’ mathematical performance; that draws valid inferences from multiple instruments; and that is used to improve instruction for all students.”

(Steen, 1999, p. 8)

This study investigated the impact of embedded formative assessment cycles with feedback and reassessment opportunities in the curriculum of large-enrollment introductory statistics courses on student attitudes toward statistics and student achievement scores. This chapter presents an overview of the study and a discussion of the results by research question. The chapter includes the study’s limitations and suggestions for further research.

Overview of the Study

This quantitative, quasi-experimental study sought to answer the following two research questions regarding the effects of FACs on student attitudes toward statistics and student achievement.

1. How do formative assessment cycles (FACs) impact student achievement in large-enrollment introductory statistics courses for different mathematically prepared students?

2. After allowing for student-to-student variability, which student attitude components change after a semester of a large-enrollment introductory statistics course with FACs? Also, how do demographic factors impact attitude, and do these effects change over time?

The methodology for Research Question 1 utilized regression discontinuity. Using ALEKS math placement scores, the regression discontinuity analysis revealed changes in
achievement across the threshold of a score of 30—the cutoff for placement into Stat 1040. The methodology for Research Question 2 employed a pre-and post-survey design to assess the changes in students’ attitudes across a semester of introductory statistics with FACs. Multilevel modeling (MLM), also called hierarchical linear models (HLM), accounted for the dependence between observations due to students being nested in recitation sections by incorporating random intercepts into the model. The discussion of the analysis from Chapter IV follows.

**Discussion of the Analysis of Research Question 1**

Regression discontinuity is a unique methodology because it can provide unbiased estimates of a treatment effect without randomly assigning subjects to experimental groups (Lesik, 2006). Using the exogenous cutoff score of 30, where students who score above 30 are placed in Stat 1040 and students who score below 30 are placed in Stat 1045, regression discontinuity provides causal inference by examining students on either side of the cutoff to determine changes in achievement between students in the pre-FACs (Fall 2017 and Spring 2018) semester and students in the FACs (Spring 2019) semester. The regression estimates at the cutoff can provide an unbiased estimate of a treatment effect. Because there were no significant differences between the demographics of the pre-FACs and the FACs groups, it is as if the students were randomly assigned to take their introductory statistics class in the pre-FACs or FACs semesters. Furthermore, student assignment into Stat 1040 and Stat 1045 is because of the exogenous continuous variable of the ALEKS placement exam measured before the semester. Thus, the cutoff
can provide unbiased estimates of the effects of the treatment.

The analysis in Chapter IV for Research Question 1 showed the following.

1. No significant difference in course achievement between Stat 1040 and Stat 1045 groups at the cutoff for the pre-FACs semesters,

2. A significant difference in course achievement between Stat 1040 and Stat 1045 at the cutoff for the FACs semester, and

3. No significant difference in course achievement between Stat 1040 and Stat 1045 from pre-FACs to FACs at the cutoff.

This section will discuss the practical significance and meaningfulness to the student at the cutoff and the implications of these results going forward.

**No Significance Difference in Course Achievement at the Cutoff in the Pre-FACs Semester**

Stat 1045 was designed as a co-requisite course, meaning that students who could not place into Stat 1040 via their mathematical knowledge before the course could learn the mathematics necessary for Stat 1040 as the course progressed. Thus, Stat 1045 has more contact hours per week than Stat 1040 (approximately 70-90 minutes more a week). However, the material is otherwise the same.

The first panel of Figure 10 displays the change in average final grade percentage for pre-FACs semesters, and the difference at the cutoff is shown. Although that difference in course achievement at the cutoff is not statistically significant, the estimated means for the students’ achievement does suggest that Stat 1045 (the solid line) did help students earn at least a B- or better, versus the Stat 1040 students who averaged below 80% at the cutoff. As discussed in Chapter IV, the average final course percentage for Stat 1045 was 81.65%, and the average final grade percentage for Stat 1040 was 77.61%,
a difference of 4.04%. Although that difference was not statistically significant, the practical significance of Stat 1045 students improving by 4.04% on average cannot be understated. Those students did not waste an extra semester of time and money to take and pass the needed mathematics courses to get into Stat 1040. In addition, several majors require that students earn at least a B- (80% or better) in Stat 1040/1045, and had the student taken the mathematics prerequisite course to finally place into Stat 1040 with a 30-35 ALEKS score, they would have, on average, scored below 80% in the course requiring them to retake Stat 1040. This 4.04% difference is the difference in having to retake the introductory statistics course, effectively saving the Stat 1045 students two additional semesters to make it through this quantitative literacy requirement. This difference not only saved students time but also saved them thousands of dollars in tuition. This finding suggests that the department should explore whether an ALEKS score of 30 is an appropriate cutoff for Stat 1040 or if it should be lowered.

The Significant Difference in Course Achievement at the Cutoff for the FACs Semester

FACs were implemented in the introductory statistics courses to answer the call to improve students’ pathways for completion in introductory statistics courses (Peck, 2019) by improving students’ achievement and experiences. The pane on the right-hand side of Figure 10 shows students’ average course achievement at the cutoff for the FACs semester. Chapter IV investigated this statistically significant difference in course achievement in the FACs semester between Stat 1045 and Stat 1040 at the cutoff. The average difference of 8.42% between the final grade percentages was statistically
significant—nearly a full standard deviation apart at the cutoff—a difference between a course grade of a C+ and a B, approximately two letter grades, in favor of students in Stat 1045. This finding provides important information to mathematics and statistics departments on the effectiveness of the ALEKS cutoff. These data suggest that students in Stat 1040 with an ALEKS score between 30-40 could benefit academically from the Stat 1045 course, potentially saving time and money rather than retaking the course to achieve a better grade.

**No Significant Difference in Course Achievement Between Stat 1040 and Stat 1045 from Pre-FACs to FACs at the Cutoff**

The purpose of Research Question 1 was to assess whether student achievement improved after implementing FACs in the large-enrollment introductory statistics courses. These data did not provide statistically significant p-values in course achievement from the pre-FACs semesters to the FACs semester. Upon visually inspecting the plots in Figures 10 and 11, there is a change in slope for the Stat 1045 group from pre-FACs to FACs and a slightly steeper slope for the Stat 1040 group from pre-FACs to FACs. First, for the Stat 1045 group, the change in achievement at an ALEKS of 29 is noticeable, with an average difference of 4.43% at the cutoff. Although not statistically significant, this change in the overall course grade in Stat 1045 at the cutoff is meaningful. Students’ average grade in the pre-FACs semester was a B- and in the FACs semester, it was a B on average. These results suggest that students in Stat 1045 took advantage of formative assessments with feedback and reassessment in the FACs semester, improving their course grades and further impacting their GPAs. Similarly,
Chiesi and Primi (2010) divided their introductory statistics students into two groups based on the student’s mathematical preparedness before the course. They found that reassessment gave the lower mathematically prepared students additional chances to succeed.

Interestingly, the Stat 1040 students with a higher ALEKS score (above 35) appeared to have a higher average final grade percentage in the FACs semester than the pre-FACs semester, as the steeper slope suggests in Figure 11. Still, at the cutoff, the difference in final grade percentages was minuscule (0.05% higher for the students in the FACs semester). Again, these data point to the need to question whether the cutoff of 30 for Stat 1040 is truly appropriate. An interesting question remains: why didn’t these students in Stat 1040 at the cutoff improve with the opportunity to retake exams and improve their overall grade like the students appeared to do in Stat 1045?

One possibility for this discrepancy in achievement at the cutoff is the “frog pond effect.” The frog pond effect suggests that a variable like a student’s “mathematical knowledge” could depend on the average mathematical knowledge of those around the student (Hox, 2018). For example, place a medium-sized frog in a pond with larger-than-average-sized frogs. The medium-sized frog feels small. Conversely, place the medium-sized frog in a pond with smaller-than-average-sized frogs, and the medium-sized frog feels large. The frog pond effect translates to this situation—the student with an ALEKS score at the cutoff placed in Stat 1040 is analogous to the medium-sized frog in the large pond. This student could feel unmotivated in Stat 1040, having their preconceived feelings about their mathematical abilities reinforced as they feel like they have the least
mathematical knowledge in the room. This lack of motivation creates a defeatist attitude—why would using the feedback and reassessments help them? The student with the ALEKS score at the cutoff in Stat 1045 begins to recognize that their mathematical knowledge seems at least as similar as others around them, providing confidence and motivation to use the feedback and reassessment opportunities to improve their grade. This study found that the Stat 1045 students at the cutoff had a higher average final grade percentage than Stat 1040 students in both the pre-FACs and FACs semesters. Although the improvement in student achievement in the FACs semester was not statistically significant, the difference in improvement between a B- and a B had practical importance to the Stat 1045 student in the FACs semester.

Summary

Meaningful and practical implications were found in the students’ course achievement in the large-enrollment introductory statistics courses. First, this study provided evidence that a co-requisite course can make a difference in students’ grades for students of lower mathematical preparedness, eliminating the so-called “bottleneck” that many students experience in completing the quantitative requirement for their academic trajectory. Second, this study provided evidence for departments evaluating the cutoff for math placement scores to determine the appropriateness of the student’s placement into introductory statistics courses. Academic advisors can also use this information to counsel students on the appropriate quantitative literacy course, given their math placement scores. Third, this study showed that FACs could impact students’ grades, especially those in co-requisite courses or those with lower math placement scores.
Students in the Stat 1045 co-requisite course took advantage of the opportunities to reassess and learn from mistakes, earning final grades like those in Stat 1040 who have higher math placement scores. This implies the frog pond effect can be a substantial issue for those with less mathematical preparedness than the average, possibly confirming students’ biases toward their abilities to understand the course, seeing no point in utilizing FACs to improve.

**Discussion of the Analysis of Research Question 2**

Research Question 2 investigated the change in student attitudes after a semester of a large-enrollment introductory statistics course with FACs, accounting for student-to-student variability. The full MLM models included the significant demographic variables and interactions that impacted and moderated attitude scores from pre- to post-survey.

Recall the SATS-36 (Schau, 2003) instrument measured students’ attitudes at two time points on a 7-point Likert scale measuring six attitude components: *affect, cognitive competence, interest, value, difficulty,* and *effort*. The 7-point Likert scale measured attitudes from 1 = “Strongly Disagree,” 4 = “Neither Disagree nor Agree” (neutral), to 7 = “Strongly Agree.”

All attitude components changed significantly from pre- to post-survey. On average, student attitudes improved significantly in *affect, cognitive competence,* and *difficulty,* decreasing significantly in the *effort, interest,* and *value* components. Notably, students’ performance in the course significantly moderated their post-survey scores. Additionally, sex was a significant main effect for all components but *effort*—the female
students’ attitudes were significantly lower on average for all attitude components except effort, where females insignificantly scored higher on average. The discussion of these results of the change in students’ attitudes follows.

The Change in Post-Survey Attitudes

Final grade percentages were provided in the pre-existing dataset. This variable specified the students’ overall course grades after the final exam. Students took the post-survey at weeks 12 and 13 while the final exam was still 3-4 weeks away. Student performance in the course by weeks 12 and 13 significantly impacted their attitudes toward statistics. This section discusses these findings considering the literature on students’ attitudes toward statistics.

Attitudes That Increased Post-Survey

On average, scores for affect, cognitive competence, and difficulty statistically and practically increased from pre- to post-survey. The students’ final course percentages moderated these changes. Using the definitions of the attitude components from Schau (2003), an increase in affect suggests that students enjoyed the class more than they initially thought they would. Notably, students were less stressed, more secure, and found greater enjoyment in the statistics course. An increase in cognitive competence demonstrates that students gained belief in their ability to learn and understand statistics, statistical concepts, and equations. And a post-survey increase in difficulty suggests the students believed learning statistics was easier, less difficult, and less complicated than they believed at the pre-survey. These three attitudes were significantly moderated by
final grade such that the A and B students’ change was greater than the C students’ change for each attitude. The C students’ changes in each attitude component were not significant. Chiesi and Primi (2010) found less mathematically prepared students did not change their feelings in affect or difficulty. The C students in this study could be those who came to the course with less mathematical preparedness than the A students.

The average change from pre- to post-survey for these attitudes which improved in this study were about five times higher than reported in the large study by Schau and Emmioğlu (2012). Schau and Emmioğlu (2012) studied attitudes from the SATS-36 on a large study of 101 sections of introductory statistics in the United States. They reported an average increase in affect of 0.13 points, 0.10 points in cognitive competence, and 0.15 points in difficulty. In this current study, the average increase in attitude scores were: 0.65 points for affect, 0.53 points for cognitive competence, and 0.61 points for difficulty (see Table 15). These average changes in affect, cognitive competence, and difficulty were also considered practically significant (greater than .5; Millar & Schau, 2010). This current study’s change in affect was five times the increase reported in Schau and Emmioğlu. Additionally, cognitive competence increased 5.3 times the reported increase in Schau and Emmioğlu, and the change in difficulty was 4 times greater. The increase seen in this study is important because Schau and Emmioğlu believed they showed that introductory statistics courses “are not enough to improve (or, for some attitude components, to even maintain) students’ attitudes toward statistics” (p. 93). They issued a call for undergraduate educators to identify factors that improve attitudes and design courses accordingly. The results of this current study provide evidence that FACs could
be such a factor in improving student attitudes.

The significant change in *affect*, from this current study, could be attributed to the many opportunities for the students in the course to attempt tests and homework questions multiple times and not be punished for initially incorrect work. Where *affect* measures students’ enjoyment of the statistics class, students on track to earn As and Bs in the course will find the course more enjoyable than students struggling to pass. Chesi and Primi (2010) saw positive increases in the four components of *affect, cognitive competence, difficulty*, and *value* for the higher mathematically prepared students. This current study only saw significant increases in attitudes for A and B students in the first three components, verifying Chiesi and Primi’s findings if, in fact, students who earn As and Bs in the course are those who were more mathematically prepared before the course. Certainly, providing students with more opportunities to succeed through FACs had no negative impact on students’ *affect* scores.

Like *affect*, there was not a statistically significant change for C students’ *cognitive competence* attitude component. Yet, the C students’ average scores were still higher than neutral attitudes, suggesting a higher-than-average belief in their abilities to perform statistical computations in the pre- and post-surveys. The A and B students’ large effect sizes in *cognitive competence* imply a significant gain in students’ feelings toward their intellectual knowledge and an improved attitude toward their ability in their mathematical skills. These findings also suggest the students felt they understood and could perform the statistical calculations. These high feelings of *cognitive competence* could be due to the feedback in the course providing ways for students to increase their
self-confidence in their cognitive abilities noted also in the study by Chiesi and Primi (2010). These findings align with other SATS research (Chiesi & Primi, 2010, Emmioglu & Capa-Aydin, 2012, Schau, 2003). In this current study and previous studies, students’ change in cognitive competence scores were significantly positive and correlated with achievement. Empowering students through feedback and reassessment through FACs in the introductory statistics curriculum provides opportunities for students to experience improvement in their mathematical abilities without the frustration over high-stakes assessments.

Regarding difficulty, students, on average, felt that the difficulty of learning statistics was less than they initially believed. On average, this change in attitude was also practically significant. This improvement could be attributed to the FACs embedded in the course that entice students to learn from mistakes and reassess their learning. Having reassessment opportunities built into the course allows students to make mistakes and creates feelings of empowerment and motivation to do statistical calculations (Freeman et al., 2014). These feelings improve students’ cognitive competence, which in turn, affects the students’ beliefs regarding the difficulty of the course. As students begin to experience success in statistical calculations, they perceive that the course is not as arduous as they first believed it would be. Several past studies saw no negative change in difficulty (Schau, 2003; Schau & Emmioglu, 2012). The studies that measured prior mathematical success in students found difficulty improved in those students with successful past experiences (Chiesi & Primi, 2010; Ramirez, 2012).

This current study’s practical and statistical gain in difficulty differed from past
studies, suggesting that FACs could have helped shift these attitudes. Notably, this current study found neutral scores in difficulty at pre-survey. The pre-survey average difficulty score was the lowest initial score of all six attitude components. This finding suggests that students expected the course to be hard, despite their confidence in their mathematical and statistical competency (cognitive competence). No matter what the students’ final grades were, their average difficulty scores increased. The A students saw a greater than a half-point change from pre- to post-survey, showing practical significance. On average, students believed that statistical computations were less difficult at the post-survey; their successes in the course helped make significant gains in their difficulty attitudes.

**Effort Decreased Post-Survey**

Effort assesses the students’ plans to study, work, attend every class, and complete all their assignments (https://www.evaluationandstatistics.com/). Initially, students’ expectation of their effort was nearly 7, the ceiling of the Likert scale ($M = 6.67$). This average pre-survey score suggests that students initially believed they planned to put nearly 100% effort into the course. All students’ effort scores dropped on average, with the C students experiencing the largest decrease in effort. This average decrease across all students suggests that either the students’ effort in the course was lacking, leading to a poorer grade, or the course did not require as much effort to pass the class as they initially thought. Students tend to have high goals for themselves at the beginning of the semester, as it is a new beginning and a fresh start. After a while, students’ efforts wane, and their ability to persist wanes. This behavior in students could be exaggerated in
The A students’ effort also significantly decreased post-survey, but smaller than the B and C students’ decreases. This smaller decrease suggests that the A students did not need to put in the amount of effort they felt was needed initially to perform well in the course, but the course still required their effort to be successful. These results are consistent with the analysis of difficulty: all students felt that statistics was easier on average post-survey (difficulty increased), and they also felt that it required less effort of them than they initially thought (effort decreased). For all students, the effort post-survey score was still high (above 4.5), suggesting that all students believed that the introductory statistics course required more effort than an average class. Schau and Emmioğlu (2012) reported an average change of -0.48 points in effort in their large-scale study. This study’s average change of -0.52 points is both statistically and practically significant, but consistent with the literature on student attitudes in the U.S.

I expected to see a decrease in effort across all students because FACs would likely decrease the student’s experience of stress and cramming compared to studying for high-stakes examinations, however, the workload required of the students to prepare for the reassessment still requires a higher-than-average effort. Preparing for the reassessments takes self-regulation. The student must act on the feedback through self-assessment to learn from their mistakes to perform better on the reassessment. This change in effort is shown by the effect sizes: the A students' average decrease in effort was small, whereas the C students’ average decrease in effort was large. This difference also points to achievement being a function of the effort the student puts into the course.
rather than the student’s mathematical ability. As the C students self-reported lower average effort scores at the post-survey, it is interesting to note that their course achievement might have been higher had they put in similar effort to the A students.

*Interest Decreased Post-Survey*

*Interest* assesses the students’ interest in their abilities to communicate statistical information, use statistics, and understand statistical information ([https://www.evaluationandstatistics.com/](https://www.evaluationandstatistics.com/)). Students’ *interest* attitudes, on average, significantly fell, 0.41 points from pre- to post-survey. Schau and Emmioğlu (2012) reported an average decrease of 0.50 points in *interest*, which is a 1.2 times greater decrease than this current study.

Students’ average *interest* was high at the pre-survey, and the A students maintained high *interest* at the post-survey. The greatest change was that of the C students, whose *interest* average fell nearly a half-point more than B students post-survey, falling into the neutral zone of attitudes. The A students’ lack of change suggests that the A students felt similar *interest* at the post-survey as they did at the beginning of the course. The B and C students experienced decreased *interest* over the course, suggesting that their desire to continue learning statistics or spend more time learning statistical concepts to portray that knowledge to others or themselves declined. Schau and Emmioğlu (2012) found that *interest* also decreased but stayed neutral: students didn’t rate interest high or low, on average. This current study differed in these results, for the A and B students’ average *interest* scores stayed high post-survey. This was surprising that all students felt higher-than-neutral *interest* towards the course at pre-survey, especially
given the hyperboles, “I hate statistics” or “statistics is the most failed class,” overheard in undergraduate students’ conversations. Even with these narratives among students, these high average scores in interest show that students began the course with higher interest than in past studies, and students mostly maintained these high attitudes. This current study’s results could be due to the increased learning opportunities to gain statistical understanding in the FACs-embedded introductory statistics curriculum—students spent more time on statistical concepts because they utilized their opportunities to improve their grades with feedback and reassessments. Thus, the A and B students may have felt more confident in their ability to explain statistics to others. Staying involved in statistical concepts through time and effort may have an increased benefit of holding students’ interest as well as improving students’ achievement.

Value Decreased Post-Survey

Value measures the students’ attitudes about the worth of statistics in their personal lives and relevance in their professional lives (https://www.evaluationandstatistics.com/). Students scored value particularly high pre-survey, above neutral for all students. The A students’ average change in value was insignificant. The insignificant change in the A students’ average value score suggests that students maintained their attitudes toward the value of statistics in their personal and professional lives. In contrast, the C students felt that statistics were less valuable to their personal and professional lives than they had believed at the beginning of the course, although value was still rated above neutral. Schau and Emmioğlu (2012) found similar results with value, but a quarter of the students had dropped to neutral attitudes post-course. Additionally, the average reported
decrease of .32 points in value was a 1.4 times higher average decrease than this current study.

Schau (2003) experienced comparable results to this current study, with value decreasing post-survey but maintaining a higher-than-neutral overall average, and she wondered why students began the course with these high average attitudes in value. It may be that introductory statistics students are familiar with polls and headlines, being the information age generation, inundated with data and information continuously. But after experiencing the course, students realize there are “lies, damn lies, and statistics” (Twain, 1906, as cited in White, 1964, p. 15). Students in the introductory course may begin to be critical of polls, headlines, and studies and recognize that not all these data are as important to their lives as they first thought. This study’s findings also suggest that the A students learned to navigate the media critically after the introductory statistics course and maintained their feelings about the value of statistics in society and careers. Although B and C students’ average value scores declined at post-survey, only the C students’ decline was practically significant. For all students, the effect size was small, and students stayed above neutral in their value attitudes. The small effect size suggests that the students’ experience in the introductory course impacted them enough to maintain their high attitudes in value.

It may seem contradictory that affect improved over time, yet the average value scored declined. Value measures attitudes regarding students’ beliefs in statistics being applicable in their lives or careers, where affect concerns their experience and feelings about being in the course. As the audience for the introductory statistics course is
generally a non-STEM student, the science of statistics may seem unnecessary for their future career, despite the interdisciplinary nature of an introductory statistics course.

**Summary**

This study’s significant changes in the SATS-36 (Schau, 2003) attitude components differ from past analyses. Schau and Emmioğlu (2012) investigated 101 sections of introductory statistics in the U.S. They found that *affect, cognitive competence, and difficulty* experienced no practically significant change and stayed neutral (3.5-4.5) on average. Additionally, *effort, interest, and value* decreased in their study on average. Where this current study also found a decrease in the attitude components of *effort, interest, and value*, the average decrease was less than Schau and Emmioğlu reported in their study for *interest* and *value*, and similar to *effort*. Moreover, this study showed an average improvement of about five times the increase Schau and Emmioğlu found for *affect, cognitive competence, and difficulty*. The average change in these three components is also practically significant in this current study, with more than a half-point change (Millar & Schau, 2010). The attitudes in Schau and Emmioğlu that decreased are the same ones in this current study, but the average decrease was about 1.2 to 1.4 times more for *interest* and *value* than in this current study. This study's attitude changes differ from the past literature on students’ attitudes, pointing to FACs as a course design that can positively impact student attitudes.

The consensus of the literature on attitudes is that attitudes are important to improve, and the introductory statistics course in post-secondary education is currently not doing enough to effect change (Ramirez et al., 2012; Schau & Emmioğlu, 2012, Xu
This study has shown a statistical and practical increase in affect, cognitive competence, and difficulty and a less than previously reported decrease in effort, interest, and value. Assessments for Learning could be the key to these major shifts in student attitudes toward statistics. Past literature showed gains in attitudes that this current study found, but this study’s attitude gains were much greater and could be due to students experiencing a course designed with FACs. In contrast, past studies which showed declines in attitude components experienced greater decreases than the students in curriculum with FACs. Thus, FACs deserve further study as the impetus for these changes in attitudes in introductory statistics students in other student populations.

**Sex Differences in the Change of Students’ Attitudes**

A meta-analysis of the SATS (Schau, 2003) surveys in the United States showed that, generally, male and female students' attitudes are indifferent (Ramirez et al., 2012). Specifically, Schau (2003) found that female and male students had similar pre-test attitude scores on average. Still, the male students were about a third of a point higher than females on affect and cognitive competence in the post-survey, which is not considered practically significant. In an analysis of 101 sections of introductory statistics in the United States, Schau and Emmioğlu (2012) did not report any sex differences, yet wondered why some sections exhibited such differences in some attitudes (could that have been a sex effect?). Xu et al. (2020) reported similar gender differences to Ramirez et al. Xu et al. found significantly higher attitudes across interest, difficulty, cognitive competence, and affect for the male students, although none were practically significant and at most 0.2 points. The sex differences seen in studies on attitudes from the SATS
surveys in the U.S. have been unremarkable.

Sex differences were notable in this current study. Sex was a statistically significant main effect for affect, cognitive competence, interest, value, and difficulty ($p < .05$). Sex was an insignificant main effect for effort ($p > .05$). The female students scored lower on average for all attitude components but effort. In addition, the differences in sex for affect and interest are practically significant as the differences are over a half-point. Seeing a sex effect in this study where sex was not seen as significant in meta-analyses could be due to the similarities among the student population at Utah State University. It is important to note that female students comprised most of the participants. Out of 441 students who took at least one part of the survey, 78% were females. Additionally, these students are overwhelmingly white and from the state of Utah (https://www.usu.edu/aaa/enrollmentsataglance.cfm). Utah has a homogenous population regarding its culture, religion, and politics. These differences in Utahns from the population of the United States could be the factor in the sex difference. Because these sex differences were seen at the pre-survey, FACs did not affect the gender gap in attitudes. This study needs to be replicated at universities across the U.S. to investigate the sex difference in attitudes in different areas and regions.

**The Connection of Students’ Achievement and Their Attitudes in the Analysis of the SATS**

Achievement in the course, measured by final course grade, was statistically significant in this current study as a moderator of attitudes. Students with A and B final grades increased in affect, cognitive competence, and difficulty. These increases were
practically significant, whereas students who earned a C saw practically significant decreases in *effort, interest, and value*. In the spring semester, the B students experienced a practically significant drop in *value*. The significance of the final course grade as a highly significant interaction in the analysis for every attitude component deserves further discussion. Many studies have demonstrated a positive association between students’ attitudes and achievement in introductory statistics (Chiesi & Primi, 2010; Emmioğlu & Capa-Aydin, 2012; Ramirez et al., 2012; Schau, 2003). Ramirez et al. (2012) found 17 studies that evaluated the relationships between SATS and achievement, where 15 had significant positive associations in *affect, cognitive competence, and value*. Few studies related *difficulty* to achievement. Contrasting the Ramirez meta-analysis with this study, the current study did find students’ final grades significantly moderating the increases in the attitudes of *affect, cognitive competence, and difficulty*. Although *value* and *interest* decreased on average, these components were also significantly moderated by course grade. These decreases were less than observed in prior studies, showing that experiencing FACs in this study helped to minimize decreases in attitudes.

Emmioğlu and Capa-Aydin (2012) performed a meta-analysis of the research on student attitudes on the components of *affect, cognitive competence, difficulty, and value*, and their association with student achievement. In the U.S., all four attitude components had statistically significant correlations with student achievement, and effect sizes were double that of non-U.S. students. This current study also saw all four attitude components significantly moderated by final grade. The literature does suggest a relationship between attitudes and achievement. However, I know of no other studies that used the final grade
as a variable in a multilevel regression model, only correlational analyses between post-survey scores and final grades.

This study found that students who experienced success in their introductory statistics courses experienced greater increases in *affect*, *cognitive competence*, and *difficulty* than in past studies on attitudes assessed by the SATS (Schau, 2003). These students additionally experienced less of a decrease in *effort*, *interest*, and *value* than students who did not achieve as high of a final course grade. This study focused on FACs being integral to increasing student success and achievement in the introductory statistics curriculum. Recall this study’s conceptual framework in Figure 2. There is a bidirectional association between student attitudes and achievement, and the results of this study have provided more evidence for that association. The conceptual framework’s focus is that FACs, as an integral instructional intervention, impacts student attitudes and student achievement. A FACs-embedded curriculum uses formative assessments as Assessment for Learning where students are empowered to utilize the feedback on the course objectives to reassess their learning and improve their understanding. Research Question 1 provided evidence of the FACs framework improving student achievement in students with lower mathematical preparedness in a corequisite course environment, and Research Question 2 showed that students’ attitudes toward statistics were moderated by the student’s achievement when experiencing a FACs-centered curriculum. This study underscored the connection between students’ attitudes towards statistics and their course achievement. Moreover, as *interest* and *value* did not decrease for the A students, this study suggests that higher achievement had a protective effect on those attitude
components that decreased on average overall.

FACs create opportunities for improved student achievement and greater satisfaction in student understanding through feedback and reassessment. Thus, improving pathways for student success could significantly improve or maintain students’ attitudes toward statistics. Attitudes can remain with students throughout their lives, affecting future generations (Ramirez et al., 2012). This study’s conceptual framework grounds future empirical studies using FACs to investigate student attitudes and achievement in introductory statistics. FACs implemented in an introductory statistics course as an assessment intervention could engender better course attitudes, affecting success in future classes and statistical experiences.

**Limitations**

Limitations to this study include those inherent in multilevel modeling that was used to analyze both research questions, including its ability to control for dependence. This study could not control for the two main instructors of the courses which could confound students’ attitudes and achievement. The limitations of each research question are discussed next.

**Limitations of Multilevel Modeling**

One limitation of educational, quasi-experimental studies is the dependence among observations. Students can work together, experience similar instruction in small sections, and all experience the same instructor for the large section. This study used multilevel modeling to attempt to control for these dependencies through random
intercepts. Additionally, for Research Question 2, other covariates like sex and semester of the course were added to control for demographic differences in the nesting variables.

Multilevel modeling also has a limitation of convergence. Models can fail to converge or have non-unique solutions, i.e., singularities. In Research Question 2, this study experienced singularities trying to further nest some models by recitation instructor. Thus, not all models could control for recitation teacher variability in the model.

**Effect of Instructors on Student Attitudes and Achievement**

Confounding can cause an association between attitudes and achievement not due to FACs. It is important to consider what other factors besides FACs could impact course performance and attitudes. One factor could be an instructor effect. Xu et al. (2020) studied the influence of instructors on student attitudes. The authors discovered that post-survey attitudes differed significantly across different instructors. They found that instructor effectiveness was associated with all five attitude components (eliminating effort due to the extreme scores and non-normality of responses). Instructor impact on student attitudes ranged between 0.18 to 0.36 $SD$s depending on the instructor’s effectiveness. Additionally, instructor-associated gains in student attitudes were positively associated with students expected grades, i.e., students expected their grades to be better in courses where the instructors’ effectiveness produced gains in student attitudes.

The students’ attitudes investigated in Research Question 2 all experienced the same large-section instructor in both semesters. Thus, there was no comparison or control...
for the main instructor. Students’ achievement in Research Question 1 had only two instructors. It is important to note that the two instructors in this study have received numerous teaching awards, with over 50 years of combined experience teaching introductory statistics courses. I was the main instructor for the semesters from which attitude data were obtained for Research Question 2. I was one of the two instructors who taught the courses from which student achievement data were obtained for Research Question 1. As such, instructor effectiveness could have impacted student attitudes. Notably, the main section was divided into recitation sections, and random intercepts modeled this dependence to control for recitation instructor-to-instructor variability for both research questions. However, due to sparsity, models for difficulty and interest could not account for this dependence in Research Question 2.

Limitations of the Analysis of Research Question 1

Students in the study who were retaking the class, withdrew, or failed were removed from the data prior to analysis. Then, students who did not take the ALEKS placement exam as their prerequisite were removed. The students using other prerequisites were not investigated in this current study. Thus, the results of this study may not extend to all students.

A total of 123 crossovers and extreme cases were investigated and subsequently removed from the sample to prevent bias in the final model’s regression estimates. Demographics were studied and showed that students in the pre-FACs semesters were similar to the students in the FACs semesters. Parameters were compared in the models with and without the crossover and extreme cases in Chapter IV. However, due to the
COVID-19 pandemic in March 2020, the Spring 2020 and subsequent semesters could not be part of the FACs semester data. Thus, only 153 students were in the FACs semester, nearly half the number of students in the pre-FACs semesters, which could have impacted the power of the analysis, limiting the ability of the regression to find a significant interaction between the discontinuity and whether the students were in the FACs semester.

Limitations of Likert Scales on the Analysis for Research Question 2

Likert scales can have truncation issues due to ceiling effects. *Effort* has a truncation issue, where at the beginning of the semester, students overestimate the effort they plan to exert in the course by selecting 7, the ceiling of the 7-point Likert scale. Thus, it is hard to know whether the change in *effort* is real or attributable to regression to the mean. Regression to the mean can happen for any attitude component: in test-retest situations, extreme values, on average, tend to regress to the mean. Thus, students severely overestimating or underestimating their pre-survey attitudes could result in an average decrease or improvement in the post-survey attitudes due to regression to the mean.

Generalizability

Research Question 1 did not investigate student achievement for those students who did not take the ALEKS math placement exam. The research question examined student achievement for those who had an ALEKS placement score. Thus,
generalizability to students who did not take the ALEKS placement exam is cautioned.

Research Question 2 did not investigate variables influencing the pre-survey attitude or what prior mathematical experiences students bring to the course. How students’ prior mathematics successes impact pre-survey attitudes would be interesting to investigate and could provide explanations for the moderation of attitudes by course achievement. Additionally, as sex was a significant fixed effect that impacted attitudes overall, future studies must investigate the difference between female and male students in other areas of the country to see if sex differences impact attitudes toward statistics, contrary to the meta-analyses. Generalizability of attitudes beyond this sample is cautioned.

The majority of this study’s participants were white female underclassmen, mostly from a single western state and culture. The diversity of ethnicities, races, and cultures was not studied. Therefore, generalizability to different populations is cautioned. Future studies could investigate the impact of FACs on marginalized groups and its impact on their attitudes and success in subsequent classes.

Implementing the Formative Assessment Cycle in Large-Enrollment Introductory Statistics Courses

The large-enrollment statistics courses in this study used technology to implement FACs. Using Harlan’s (2012) framework of Assessment for Learning, I employed several types of formative assessments in the course, from in-class informal formative assessments to informal summative assessments (the unit tests in the FACs semester).
Several practitioner papers and book chapters provided examples of large-enrollment introductory statistics courses implementing elements of the formative assessment cycle using technology. These papers are excellent resources for ways to apply computer-based formative assessments in the introductory statistics curriculum. I discuss how I implemented FACs in this study and provide these additional resources in this section.

Online student response systems (OSRSs) such as clickers or polling software import formative assessments with automatic feedback and the opportunity to reassess students’ understanding throughout a lecture. The current study used an OSRS daily to continuously provide in-class informal formative assessments to guide instruction and provide important feedback to students in the FACs semesters. The questions posed in the OSRS in this current study asked students about concepts they learned from their reading, the past lecture, or what was barely discussed to assess whether the information was being received by students correctly or if further discussion was needed. Additionally, the OSRS allowed for ease in collecting student data for use in class demonstrations and activities anonymously. Several of these activities are explained in detail in an article I co-authored (Schneiter et al., 2022). Full lesson plans that include OSRS questions and how the instructors used them in their classes for exam reviews, data exploration, and lectures are available from other papers (Bruff, 2011; Gunderson & McGowan, 2011; Murphy et al., 2011). In addition, Peck (2011) created a website with links to OSRS questions and other online resources (http://mathquest.carroll.edu/resources.html). An informal experiment by Peck (2011) found that students who used an OSRS had higher average final exam scores than those who did not. An important confounder of this
informal study is that students in the OSRS sections of introductory statistics had higher attendance. Using OSRSs daily provides students and instructors with feedback regarding instruction and student learning.

This current study utilized formal formative assessments through the online homework system in the LMS. Students received up to three attempts to obtain the correct answer on homework and reading quizzes. Having multiple attempts allowed students to take advantage of the automated feedback available in the online environment. There are also many options for online and web-based homework systems for introductory statistics, from textbook publishers to free and open-source homework systems mentioned in Lunsford and Pendergrass (2016). The authors also discussed the importance of allowing for multiple attempts on each homework question. Implementing web-based homework benefited students when the curriculum provided students the ability to resubmit homework.

Last, but importantly, this current study implemented FACs in large-enrollment introductory statistics with frequent, low-stakes examinations through computer-generated testing and grading. Informal summative assessments are formative in the Assessment for Learning framework (Harlan, 2012). Initially, creating individualized testing can be time-consuming but can provide easy grading for the instructor of large-enrollment courses and relay immediate, individualized feedback to the students. The benefits make the time commitment worth the effort to import formative assessments for all (Spencer, 2010; Stirling, 2010). Creating test banks, randomizing questions, and providing individualized assessments is an investment of instructor and departmental
resources, but one that allows students to experience automatic feedback, self-
assessment, and retakes or reassessments in a large-enrollment learning environment
(Simonite & Targett, 2010). Computer-based testing makes grading and reassessments
trivial in large-enrollment courses. The practical implications of this study can be made
available to all students by implementing FACs.

Future Research

This quasi-experimental study found that students with lower math placement
scores significantly improved their overall course grades in the FACs semester.
Additional research could use the regression discontinuity methodology to investigate
whether homework scores, course achievement before the final, or final exam scores as a
summative assessment after experiencing a course with FACs improve. This study also
found that all attitudes in the SATS-36 (Schau, 2003) were significantly impacted by
experiencing a semester of introductory statistics embedded with FACs. High effect sizes
were found as student achievement moderated attitudes. These results created an
important foundation as a feasibility study to investigate FACs’ impact on student
attitudes and achievement. The following research questions are offered for future study.

• How does the implementation of FACs in large-enrollment introductory
  statistics courses support achievement?
  o Do pass/fail/withdrawal rates improve?
  o Do assessment (homework, quiz, exam) scores improve when offered
    reassessment?
  o What are the characteristics of students who utilize reassessment?
  o Do FACs improve summative scores?
Do FACs improve statistical thinking and statistical understanding?

What are the impacts on student attitudes toward statistics, statistics anxiety, self-efficacy, self-regulation, and other cognitive attributes after experiencing a large-enrollment course with FACs?

Do FACs provide an equitable student experience among students of diverse mathematical backgrounds, majors, genders, ethnicities, and interests?

Of interest would be an investigation into the characteristics of students who took advantage of the retake opportunities and whether the retake opportunities significantly improved students’ grades. Further research could also investigate whether the opportunity for reassessment lowered students’ statistical anxieties about assessment. Qualitative research is also important to identify students’ self-efficacy and beliefs surrounding their statistical abilities and achievement. Finally, further research into those students whose course achievement improved in a course with FACs and their corresponding changes in attitudes toward statistics is needed.

Conclusions

Overall, this study provided a basis as a feasibility study for a large-scale endeavor of implementing FACs and studying students’ achievement and attitudes. It showed that the regression discontinuity methodology could help identify achievement differences with an exogenous “cutoff” variable common to undergraduate mathematics and statistics courses. Further, this methodology allows future studies to make valid and reliable conclusions about student achievement. This study found that students in the co-requisite introductory statistics course had the greatest change in achievement in the FACs semester, suggesting the need for departments to offer co-requisite courses to help
students successfully complete their quantitative requirements and provide greater learning gains.

Students’ attitudes towards statistics are decidedly an outcome of introductory statistics courses that the statistics education community has deemed valuable to a students’ introductory statistics experience and aspires to improve (Xu & Schau, 2019). This study found that all attitude components significantly changed from pre- to post-survey and that final course grades significantly moderated that change. The average increases in affect, cognitive competence, and difficulty in this study were larger than in past studies. Students performing well in the course did not experience the average decline in the components of interest, effort, and value as other students in this study. These differences in this current study compared to past literature on the SATS surveys are strikingly different. Thus, there is a need for further studies using FACs as a curricular intervention that potentially improves student attitudes toward statistics, especially as FACs can impact student achievement, moderating the students’ attitudes. Improving student attitudes fosters an appreciation for statistics and statistical thinking that extends beyond the classroom, especially for students with diverse and varying mathematical backgrounds typically found in introductory statistics courses.

The effects of FACs have far-reaching implications for instructors, students, statistics departments, and curriculum designers. Currently, instructors tend to be their own barrier to implementing a change in assessment in a classroom of several hundred students (Cash et al., 2017; Garfield et al., 2002, 2011). The GAISE College Report (ASA Committee, 2016) recommends using assessments in statistics courses to improve student
outcomes and evaluate achievement. Currently, assessment is not a continuous, daily practice by most introductory statistics instructors (Xu et al., 2020). Instructors require support for applying evidence-based pedagogical transformations to their courses, stating the “lack of time to plan, practice, use and reflect” being of great concern regarding making those changes (Ghaicha, 2016, p. 222). Instructors must also respond to student feedback constructively and without judgment, which is difficult with large-enrollment courses without computer-assisted testing (Shute, 2008). To address these concerns, mathematics and statistics departments can allocate resources for instructors to create and implement FACs as previously suggested in large-enrollment courses. Additionally, this study provided a roadmap for implementing FACs in large-enrollment courses and the type of technology and commitment required of departments and instructors to sustain formative assessments with reassessments in large-enrollment capacities. Moreover, with future research, introductory statistics curricula and course designers will more readily design course curricula employing FACs, creating a learning environment that empowers all adult learners.

This study aimed to investigate the impact of FACs in the introductory statistics curriculum on the important student outcomes of attitudes and achievement. Research has suggested that formative aspects like feedback improve students’ performance on assignments (Abell et al., 2018; ASA Revision Committee, 2016; Shute, 2008). Specifically, this study used immediate, automated feedback connected to learning outcomes, which is also associated with increased performance in other studies (Shute, 2008). As evidenced by the literature, the individual elements of FACs can improve the
outcomes of attitudes toward statistics and achievement. Thus, embedding FACs can potentially renovate the introductory statistics curriculum, providing all students with successful pathways to achieve their quantitative literacy requirement in higher education.

With the current focus on improving statistics achievement and creating successful student experiences, formative assessment must be integrated as a continuous cycle—the foundation of the introductory statistics experience (ASA Revision Committee, 2016; Steen, 1999). FACs can impact student achievement and attitudes during the course, and due to improved attitudes toward statistics, throughout their lives. As students experience increased opportunities for success from formative feedback and repeating assessment opportunities, FACs could reduce statistics anxiety, improve attitudes towards statistics, and increase statistics achievement, preparing a new generation of statistically literate citizens for a data-driven world.
REFERENCES


Bruff, D. (2011). Engaging statistics students with classroom response systems. In K. Cline & H. Zullo (Eds.), *Teaching mathematics with classroom voting with and without clickers* (pp. 61–70). The Mathematical Association of America. [https://doi.org/10.1017/CBO9781113443018.010](https://doi.org/10.1017/CBO9781113443018.010)


APPENDICES
Appendix A

Data Sources, Deidentified for Proposed Study

Pre-FACs

<table>
<thead>
<tr>
<th>Semester</th>
<th>Course</th>
<th>Instructor</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sp 2018(^a)</td>
<td>1040-001</td>
<td>T1</td>
<td>Student Data (gender, class standing, major, pre-requisite type or ALEKS score, whether student was retaking the class); Course data (Midterm 1 percentage, Midterm 2 percentage, final exam percentage, final course percentage)</td>
</tr>
<tr>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fa 2017(^a)</td>
<td>1040-001</td>
<td>T2</td>
<td>Student Data (gender, class standing, major, pre-requisite type or ALEKS score, whether student was retaking the class); Course data (Midterm 1 percentage, Midterm 2 percentage, final exam percentage, final course percentage)</td>
</tr>
<tr>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FAC Creation Period

<table>
<thead>
<tr>
<th>Semester</th>
<th>Course</th>
<th>Instructor</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa 2018</td>
<td>1040-001</td>
<td>T1</td>
<td>No data; Test Bank Creation</td>
</tr>
<tr>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sp 2019</td>
<td>1040-001</td>
<td>T1</td>
<td>No data; Test Bank Creation</td>
</tr>
<tr>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

FACs

<table>
<thead>
<tr>
<th>Semester</th>
<th>Course</th>
<th>Instructor</th>
<th>Data Sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fa 2019(^a)</td>
<td>1040-001</td>
<td>T2</td>
<td>Student data (sex, class standing, major, pre-requisite type or ALEKS score, whether student was retaking the class); Course data (all test percentages for all attempts, final exam percentage, final course percentage)</td>
</tr>
<tr>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sp 2020</td>
<td>COVID</td>
<td></td>
<td>No data; Not included in study due to pandemic, sent all students home in March 2020</td>
</tr>
<tr>
<td>Fa 2020</td>
<td>COVID</td>
<td></td>
<td>No data; Not included due to pandemic</td>
</tr>
<tr>
<td>Sp 2021</td>
<td>1040 XL</td>
<td>T1</td>
<td>No data; Not included due to continued pandemic conditions which continued to cause disruption to students and classes, illness, etc.</td>
</tr>
<tr>
<td>1045 XL</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fa 2021(^b)</td>
<td>1040</td>
<td>T1</td>
<td>Student data (sex, class standing, major, pre-requisite type or ALEKS score); SATS Pre-Survey Responses; SATS Post-Survey Responses, Recitation Section and TA; final course percentage</td>
</tr>
<tr>
<td>1045</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sp 2022(^b)</td>
<td>1040-003</td>
<td>T1</td>
<td>Student data (sex, class standing, major, pre-requisite type or ALEKS score); SATS Pre-Survey Responses; SATS Post-Survey Responses, Recitation Section and TA; final course percentage</td>
</tr>
<tr>
<td>1045-001</td>
<td>T1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. T1 and T2 are two different instructors.

\(^a\)Denotes data for Research Question 1. \(^b\)Denotes data for Research Question 2.
Appendix B

Permission to Use SATS-36

[EXT] Re: Register

Candace Schau <cschau@comcast.net>
Tue 10/27/2020 2:14 PM
To: KimberLeigh Hadfield <k.hadfield@usu.edu>

Hi, KimberLeigh,

Yes, you are welcome to continue to use the SATS. You can add the extra demographic questions.

I'm confused about the article with the SAT-M. The main SATS-M I know about refers to a model based on Expectancy Value Theory but with the SATS-36 in it. Can you let me know if I'm not understanding your questions?

Take care,
Candace

On 10/26/2020 2:21 PM KimberLeigh Hadfield <k.hadfield@usu.edu> wrote:

Dear Dr. Schau,

I hope this email finds you well.

I would like to continue to use the SATS-36 for my large-section introductory course to measure attitudes for an eventual dissertation.

I have no funding, and I will absolutely inform you of any results I obtain from my dissertation (2021-2022).

Also, I read an article about SATS-M and was wondering if I could view that instrument to see if it would be more appropriate to my dissertation?

Lastly, may I ask if I can add a few more demographic-like questions to the end of the pre and post-tests? If not, it is okay, as I would have the questions asked at a different time in the course. I didn't know if that would cause a problem, and I don't want to alter your instrument in any way.

All the best,
KimberLeigh

KimberLeigh Hadfield
Senior Lecturer
Dept of Mathematics & Statistics
AnSci 315
k.hadfield@usu.edu
Appendix C

SATS-36 Pre-Survey
DIRECTIONS: The statements below are designed to identify your attitudes about statistics. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

<table>
<thead>
<tr>
<th></th>
<th>Strongly disagree</th>
<th>Neither disagree nor agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I plan to complete all of my statistics assignments.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I plan to work hard in my statistics course.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will like statistics.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will feel insecure when I have to do statistics problems.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will have trouble understanding statistics because of how I think.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics formulas are easy to understand.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics is worthless.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics is a complicated subject.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics should be a required part of my professional training.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistical skills will make me more employable.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will have no idea of what's going on in this statistics course.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I am interested in being able to communicate statistical information to others.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statistics is not useful to the typical professional.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I plan to study hard for every statistics test.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>I will get frustrated going over statistics tests in class.</td>
<td>1  2  3  4  5  6  7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Statement</td>
<td>Strongly disagree</td>
<td>Neither disagree nor agree</td>
<td>Strongly agree</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>----------------------------</td>
<td>----------------</td>
</tr>
<tr>
<td>Statistical thinking is not applicable in my life outside my job.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I use statistics in my everyday life</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I will be under stress during statistics class.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I will enjoy taking statistics courses.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I am interested in using statistics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Statistics conclusions are rarely presented in everyday life.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Statistics is a subject quickly learned by most people.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I am interested in understanding statistical information.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Learning statistics requires a great deal of discipline.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I will have no application for statistics in my profession.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I will make a lot of math errors in statistics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I plan to attend every statistics class session.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I am scared by statistics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I am interested in learning statistics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Statistics involves massive computations.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I can learn statistics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I will understand statistics equations.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Statistics is irrelevant in my life.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Statistics is highly technical.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>I will find it difficult to understand statistical concepts.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Most people have to learn a new way of thinking to do statistics.</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
</tbody>
</table>
Please note that the labels for each scale on the rest of this page change from item to item.

How well did you do in mathematics courses you have taken in the past?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very poorly</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Very well</td>
</tr>
<tr>
<td>Very good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How good at mathematics are you?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very poor</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Very good</td>
</tr>
<tr>
<td>Very good</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the field in which you hope to be employed when you finish school, how much will you use statistics?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Great deal</td>
</tr>
<tr>
<td>Very confident</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

How confident are you that you can master introductory statistics material?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all confident</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Very confident</td>
</tr>
<tr>
<td>Very confident</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Are you required to take this statistics course (or one like it) to complete your degree program?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Don’t know</td>
</tr>
<tr>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

If the choice had been yours, how likely is it that you would have chosen to take any course in statistics?

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not at all likely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Very likely</td>
</tr>
<tr>
<td>Very likely</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

DIRECTIONS: For each of the following statements mark the one best response. Notice that the response scale changes on each item.

What is your major? If you have a double major, pick the one that bests represents your interests.

1. Arts/Humanities 6. Education
2. Biology 7. Engineering
4. Chemistry 9. Medicine/Pre-Medicine
5. Economics 10. Psychology
11. Sociology/Social Work
12. Statistics
13. Other

Current grade point average (please estimate if you don’t know; give only one single numeric response: e.g., 3.52). If you do not yet have a grade point average, please enter 99:  _______
For each of the following three items, give one single numeric response (e.g., 26). Please estimate if you don’t know exactly.

Number of credit hours earned toward the degree you are currently seeking (don’t count this semester):

________

Number of high school mathematics and/or statistics courses completed:

________

Number of college mathematics and/or statistics courses completed: (don’t count this semester):

________

Degree you are currently seeking:

1. Associate  5. Certification
2. Bachelors  6. Post-bachelor's Licensure
3. Masters  7. Specialist
4. Doctorate  8. Other

What grade do you expect to receive in this course?


In order to describe the characteristics of your class as a whole, we need your responses to the following items.

Your gender:  1. Male  2. Female  3. Other

Your citizenship:  1. US citizen  2. Foreign student  3. Other

Your age (in years): ______

THANKS FOR YOUR HELP!
Appendix D

SATS-36 Post-Survey
DIRECTIONS: The statements below are designed to identify your attitudes about statistics. Each item has 7 possible responses. The responses range from 1 (strongly disagree) through 4 (neither disagree nor agree) to 7 (strongly agree). If you have no opinion, choose response 4. Please read each statement. Mark the one response that most clearly represents your degree of agreement or disagreement with that statement. Try not to think too deeply about each response. Record your answer and move quickly to the next item. Please respond to all of the statements.

<table>
<thead>
<tr>
<th>Strongly disagree</th>
<th>Neither disagree nor agree</th>
<th>Strongly agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>I tried to complete all of my statistics assignments.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I worked hard in my statistics course.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I like statistics.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I feel insecure when I have to do statistics problems.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I have trouble understanding statistics because of how I think.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statistics formulas are easy to understand.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statistics is worthless.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statistics is a complicated subject.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statistics should be a required part of my professional training.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statistical skills will make me more employable.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I have no idea of what's going on in this statistics course.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>I am interested in being able to communicate statistical information to others.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statistics is not useful to the typical professional.</td>
<td>1 2 3 4 5 6 7</td>
<td></td>
</tr>
<tr>
<td>Statement</td>
<td>Strongly disagree</td>
<td>Neither disagree nor agree</td>
</tr>
<tr>
<td>---------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>-----------------------------</td>
</tr>
<tr>
<td>I tried to study hard for every statistics test.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I get frustrated going over statistics tests in class.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Statistical thinking is not applicable in my life outside my job.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I use statistics in my everyday life</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am under stress during statistics class.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I enjoy taking statistics courses.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am interested in using statistics.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Statistics conclusions are rarely presented in everyday life.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Statistics is a subject quickly learned by most people.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am interested in understanding statistical information.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Learning statistics requires a great deal of discipline.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I will have no application for statistics in my profession.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I make a lot of math errors in statistics.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I tried to attend every statistics class session.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am scared by statistics.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I am interested in learning statistics.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Statistics involves massive computations.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I can learn statistics.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I understand statistics equations.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Statistics is irrelevant in my life.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Perception</td>
<td>Strongly disagree</td>
<td>Neither disagree nor agree</td>
</tr>
<tr>
<td>----------------------------------------------------------------------------</td>
<td>-------------------</td>
<td>---------------------------</td>
</tr>
<tr>
<td>Statistics is highly technical.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>I find it difficult to understand statistical concepts.</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Most people have to learn a new way of thinking to do statistics.</td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

NOTICE that the labels for the scale on each of the following items differ from those used above.

<table>
<thead>
<tr>
<th>Question</th>
<th>Scale</th>
<th>Label</th>
</tr>
</thead>
<tbody>
<tr>
<td>How good at mathematics are you?</td>
<td>Very poor</td>
<td>Very poor</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>In the field in which you hope to be employed when you finish school, how much will you use statistics?</td>
<td>Not at all employed</td>
<td>Great deal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>How confident are you that you have mastered introductory statistics material?</td>
<td>Not at all confident</td>
<td>Very confident</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>As you complete the remainder of your degree program, how much will you use statistics?</td>
<td>Not at all employed</td>
<td>Great deal</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>If you could, how likely is it that you would choose to take another course in statistics?</td>
<td>Not at all likely</td>
<td>Very likely</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>How difficult for you is the material currently being covered in this course?</td>
<td>Very easy</td>
<td>Very difficult</td>
</tr>
<tr>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3</td>
<td></td>
</tr>
<tr>
<td></td>
<td>4</td>
<td></td>
</tr>
<tr>
<td></td>
<td>5</td>
<td></td>
</tr>
<tr>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>7</td>
<td></td>
</tr>
</tbody>
</table>
DIRECTIONS: For each of the following statements mark the one best response. Notice that the response scale changes on each item.

Do you know definitely what grade you will receive in this course?
1. Yes  2. No

What grade do you expect to receive in this course?


In a usual week, how many hours did you spend outside of class studying statistics? Give only one single numeric response that is a whole number ____________

In the past week, how would you describe your overall stress level?

1 2 3 4 5 6 7

THANKS FOR YOUR HELP!
## Appendix E

**Additional Questions Asked on SATS-36 Pre- and Post-Surveys**

### Additional Questions Pre-Survey

<table>
<thead>
<tr>
<th>How anxious are you about taking statistics exams coming into this class?</th>
<th>Not at all anxious</th>
<th>Very anxious</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How anxious are you about taking statistics exams if the option to take exams multiple times were available?</th>
<th>Not at all anxious</th>
<th>Very anxious</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>How much do you think that having exam retakes will help your grade this semester?</th>
<th>No help at all</th>
<th>A major help to my overall grade</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

### Additional Questions Post-Survey

<table>
<thead>
<tr>
<th>Rate your level of anxiety toward statistics before you took this class.</th>
<th>Not at all anxious</th>
<th>Very anxious</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Rate your level of anxiety toward statistics now the course is near completion.</th>
<th>Not at all anxious</th>
<th>Very anxious</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>No help at all</td>
<td></td>
</tr>
<tr>
<td>--------------------------</td>
<td>----------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>How much do you think having exam retakes helped your anxiety towards statistics this semester?</td>
<td>1   2   3  4  5  6  7</td>
<td></td>
</tr>
<tr>
<td>How much do you think that having exam retakes and multiple chances on homework helped your grade this semester?</td>
<td>1   2   3  4  5  6  7</td>
<td></td>
</tr>
</tbody>
</table>
Appendix F

Introductory Statistics Course Objectives

Experimental Design:
Given the description of conducted research,
1.1 Distinguish between a controlled experiment and an observational study.
1.2 Identify the treatment group and the control group.
   • 1.2.1 Distinguish between the treatment and response.
1.3 Evaluate the use of a placebo (including efficacy and ethics).
1.4 Evaluate and distinguish between the use of blind and double-blind procedures.
1.5 Distinguish between longitudinal research and cross-sectional research
1.6 Determine possible confounding factors and describe their effect on both the treatment and the outcome.
1.7 Evaluate the possible causal relationship between the treatment and the outcome:
   • Association between the treatment and outcome is not causation.
   • Randomized, controlled experiments allow a careful inference of causation.
1.8 Evaluate the importance and benefits of using randomization in experimental design.
   • 1.8.1 Describe the weaknesses of various kinds of non-randomized studies (such as historical control groups or non-randomized control groups)
1.9 Understand and recognize Simpson's Paradox.

Graphical and Numerical Summaries of Data:
2.1 Distinguish between qualitative and quantitative (discrete and continuous) variables.
2.2 Histograms
   • 2.2.1 Construct and label a density histogram.
   • 2.2.2 Use a density histogram to estimate the percentage of data within a specified interval.
   • 2.2.3 Analyze the visual distribution of data in a histogram based upon its symmetry, tails, and bell shape.
   • 2.2.4: Calculate the percent, width and height of intervals of a histogram.
2.3 Numerical Summaries

- 2.3.1 Calculate the average, standard deviation, and median of a list of numbers and be able to interpret them in context.
- 2.3.2 Given a histogram, visually estimate the average, standard deviation, and median of the data.

2.4 For a described list of data,

- 2.4.1 Calculate the change in the average, standard deviation, and median that will occur when a single number is added or multiplied to each number in the list.
- 2.4.2 Estimate the change in the average and standard deviation that will occur when another data set is combined with the list.

**Normal Approximation and Estimation:**

2.5 Given the average and standard deviation of a normally distributed data set

- 2.5.1 Calculate the percentage of the data that falls within a specified range.
- 2.5.2 Calculate the percentile rank for a given data point.
- 2.5.3 Calculate the raw score for a given percentile rank.

2.6 Given the histogram for a normally distributed data set,

- use the empirical rule to visually estimate the average and standard deviation of the data,
- use the empirical rule to estimate areas under the normal curve, and
- use the empirical rule to identify outliers.

2.7 Given a list of repeated measurements, calculate the best guess and estimated error for the next measurement.

2.8 Understand the difference between chance error and bias.

**Correlation:**

3.1 Given a scatter plot of two associated variables, $x$ and $y$,

3.1.1 Estimate the five-number summary of the data (average of $x$, average of $y$, SD of $x$, SD of $y$, $r$).

3.1.2 Distinguish between a linear and non-linear association.

3.1.3 Distinguish between a strong and weak association.

3.1.4 Evaluate the possible causal relationship between $x$ and $y$. 
3.2 Given the description of a linear association of two variables, interpret the sign and strength of the correlation coefficient.

3.3 Given a small set of data, calculate a correlation coefficient.

3.4 Given the description of a linear association between two variables based on averages, recognize, and explain the effect of ecological correlation.

3.5 Given the description of a linear association between two variables (including \( r \)), predict the value of the correlation coefficient after

- a change of scale in either variable.
- an interchanging of the two variables.
- an outlier is included or removed from the data.

**Regression:**

3.6 Given the five-number summary of the data (average of \( x \), average of \( y \), \( \text{SD of } x \), \( \text{SD of } y \), \( r \))

3.6.1 Calculate the equation of the regression line, be able to interpret slope and intercept in context of the data.

3.6.2 Estimate the average \( y \) from \( x \); predict a value from a given \( x \) value; find a regression estimate. Be able to interpret in context.

3.6.3 Calculate the r.m.s. error for the regression line and its measure of chance error for the regression estimate.

3.6.4 Use the r.m.s. error to determine if an observation is an outlier.

3.6.5 Distinguish between the regression line and the SD line both visually and by the slope calculation.

3.7 Given a scatter plot, evaluate the appropriateness of a linear regression model for the data (based on outliers, linearity, and homoscedasticity).

3.8 Given a scenario involving a test-retest situation, recognize and explain the regression effect and the regression fallacy.

3.9 Understand danger of extrapolation.

3.10 Calculate a residual; use residual plots to determine if least squares regression line is appropriate for the data.

**Probability:**

4.1 Represent chance as a ratio of desired outcomes over possible outcomes.

4.2 Recognize independent, dependent, and mutually exclusive events.
4.3 Calculate the chance of independent and dependent events using the multiplication rule.

4.4 Calculate probabilities for mutually exclusive events using the addition rule.

4.5 Calculate the chance of the opposite of an event using the subtraction rule.

4.6 List the possible outcomes of a chance process (tossing 3 coins, rolling 2 dice, drawing a ticket from Box A and Box B, etc.)

**Chance Variability:**

5.1 Construct a box model for chance processes

5.2 Calculate the average and standard deviation of the tickets in the box.

5.3 Be able to explain the Law of Averages as it pertains to the long run of events and their probabilities.
   
   - 5.3.1 Use the law of averages to explain the likelihood of getting certain outcomes in the short run versus the long run.
   - 5.3.2 Use the law of averages to explain the Gambler's Fallacy

5.4 Calculate the EVsum and SEsum for the sum of draws from a box.

5.5 Be able to set up chance processes and games (like Roulette) using a box model analogy.

5.6 Be able to find the chance of an observed sum or more/less using the probability histogram

5.7 Understand that the probability histogram for the sum of the draws looks like the normal curve, even if the tickets in the box are not normally distributed, if the draws are sufficiently large.

**Sample Surveys:**

6.1 Distinguish between a population, a parameter, a sample, and a statistic.

6.2 Evaluate a sampling situation based on the sampling methods used
   
   - Simple random samples
   - Quota samples
   - Cluster samples
   - Samples of convenience

6.3 Predict potentially resulting types of bias in different sampling situations
   
   - Response bias
• Non-response bias
• Selection bias
• Volunteer-response bias

**Chance Errors in Sampling:**

6.4 Describe and use the sampling distribution for a sample percentage.

• Create a box model.
• Calculate the EV% and SE%.
• Calculate the chance of observing a specified range of outcomes.

6.5 Describe and use the sampling distribution for a sample average.

• Create a box model.
• Calculate the EVave and SEave.
• Calculate the chance of observing a specified range of outcomes.

6.6 Evaluate the relative accuracy of two samples based on their sample size.

6.7 Describe the shape, center, and spread of a sampling distribution (CLT) for sample percentages and sample averages.

6.8 Calculate the new SE using the Square Root Law.

**Confidence Intervals:**

6.9 Given sample statistics, calculate a confidence interval for the percentage at a given confidence level.

6.10 Given sample statistics, calculate a confidence interval for the average at a given confidence level.

6.11 Given a scenario involving sample data, interpret a confidence interval.

6.12 Be able to know whether a confidence interval is appropriate for a situation.

**One Sample Z-tests:**

7.1 Given the claim about a parameter and information about a sample (data or statistics), conduct an appropriate hypothesis test to evaluate the claim, demonstrating all the following:

• 7.1.1 Recognize the use of a one-sample z-test for percentage
• 7.1.2 Construct appropriate null and alternative hypotheses.
• 7.1.3 Calculate an appropriate test statistic.
• 7.1.4 Calculate a p-value.
• 7.1.5 Reject or do not reject the null hypothesis based on the p-value.
• 7.1.6 Write an appropriate conclusion in context.

7.2 Given the claim about a parameter and information about a sample (data or statistics), conduct an appropriate hypothesis test to evaluate the claim, demonstrating all the following:

• 7.2.1 Recognize the use of a one-sample z-test for average
• 7.2.2 Construct appropriate null and alternative hypotheses.
• 7.2.3 Calculate an appropriate test statistic.
• 7.2.4 Calculate a p-value.
• 7.2.5 Reject or do not reject the null hypothesis based on the p-value
• 7.2.6 Write an appropriate conclusion in context.

One Sample T-tests for Averages:

7.3 Given the claim about a parameter and information about a sample (data or statistics), conduct an appropriate hypothesis test to evaluate the claim, demonstrating all the following:

• 7.3.1. Recognize the use of a one-sample t-test for average
• 7.3.2. Construct appropriate null and alternative hypotheses.
• 7.3.3. Calculate an appropriate test statistic (using SD+ for the SEave).
• 7.3.4. Calculate a p-value (with the appropriate degrees of freedom).
• 7.3.5. Reject or do not reject the null hypothesis based on the p-value.
• 7.3.6. Write an appropriate conclusion in context

7.4.1 Understand the assumptions of a one-sample z-test: a large, random sample

7.4.2 Understand the assumptions for a one-sample t-test: a small, random sample from a population which follows the normal curve.

Two Sample Z-tests for Percentages and Averages:

7.5 Given quantitative information about two samples (data or statistics), conduct an appropriate hypothesis test, demonstrating all the following:

• 7.5.1. Recognize the use of a two-sample z-test for averages
• 7.5.2. Construct appropriate null and alternative hypotheses.
• 7.5.3. Calculate an appropriate test statistic.
• 7.5.4. Calculate a p-value.
• 7.5.5. Reject or do not reject the null hypothesis based on the p-value.
• 7.5.6. Write an appropriate conclusion in context

7.6 Given qualitative information about two samples (data or statistics), conduct an appropriate hypothesis test, demonstrating all the following:
• 7.6.1. Recognize the use of a two-sample z-test for percentages
• 7.6.2. Construct appropriate null and alternative hypotheses.
• 7.6.3. Calculate an appropriate test statistic.
• 7.6.4. Calculate a p-value.
• 7.6.5. Reject or do not reject the null hypothesis based on the p-value.
• 7.6.6. Write an appropriate conclusion in context

7.7 Understand that the assumptions for a two-sample z-test require two, large, independent samples

Chi-Square Tests:

7.8 Given a claim or model and qualitative information about a sample (data or statistics), conduct an appropriate hypothesis test to evaluate the claim, demonstrating all the following:

• 7.8.1 Recognize the use of a chi-squared goodness of fit test
• 7.8.2. Construct appropriate null and alternative hypotheses.
• 7.8.3. Calculate an appropriate test statistic.
• 7.8.4. Calculate a p-value.
• 7.8.5. Reject or do not reject the null hypothesis based on the p-value.
• 7.8.6. Write an appropriate conclusion in context.

7.9 Given two qualitative variables about a sample, conduct an appropriate hypothesis test to evaluate the independence of the two variables, demonstrating all the following:

• 7.9.1 Recognize the use of a chi-squared test for independence
• 7.9.2. Construct appropriate null and alternative hypotheses.
• 7.9.3. Calculate an appropriate test statistic.
• 7.9.4. Calculate a p-value.
• 7.9.5. Reject or do not reject the null hypothesis based on the p-value.
• 7.9.6. Write an appropriate conclusion in context.

Cautions and Recommendations:

7.10 Evaluate the appropriate use of hypothesis tests and the conclusions that can be drawn from them.

• 7.10.1 Recognize that a P-value of 4.99% and a P-value of 5.01% are almost equally convincing even though only the former is "statistically significant".
• 7.10.2 Recognize that even "statistically significant" results can be due to chance and know that this happens 5% of the time.
• 7.10.3 Identify situations of "data snooping" where researchers test many different hypotheses and focus only on those that are "statistically significant" at the 5% level.
• 7.10.4 Recognize that Z-tests and t-tests can be either one-tailed or two-tailed depending on the alternative hypothesis.
• 7.10.5 Know how to calculate the P-value for both one-tailed and 2-tailed tests.
• 7.10.6 Recognize that "statistical significance and importance are different and know how to identify the difference.
• 7.10.7 Understand the role of a box model and be able to recognize when tests of significance are being wrongly used for data that represent the entire population.
Appendix G

Learning Mastery Gradebook Examples: Teacher and Student View

Figure G.1

Teacher View of the LMG for the Course (Students’ Names are Erased)
Figure G.2

Test Student View of the LMG

Grades for Test Student

Arrange By
Due Date
Apply

Assignments Learning Mastery

Fall 2021 STAT-1040-MW1 XL
0 OF 1 MASTERCED

Unit 1 - Experimental Design
0 OF 10 MASTERCED

1.1: Distinguish between a controlled experiment and an observational study
2 alignments
NOT MASTERCED

1.2: Identify the treatment group and the control group.
1 alignment
NOT MASTERCED

1.2.1 Distinguish between the treatment and response.
No alignments
NOT MASTERCED

1.3: Evaluate the use of a placebo (including efficacy and ethics).
No alignments
NOT MASTERCED
Appendix H

Units Covered in the Stat 1040/1045 Courses

Unit 1:

Design of experiments: controls, randomization, blind and double-blind, placebos. The Salk vaccine trial, historical controls.

Observational studies: association and causation, confounding factors, Simpson’s Paradox

Unit 2:

Descriptive statistics: the histogram, the density scale, cross-tabulation, the average and the SD and their relationship to the histogram, the median, standardization, the normal approximation, percentiles, percentiles and the normal curve, measurement error, outliers, bias versus chance error.

Unit 3:

Correlation: the scatter diagram, the correlation coefficient, properties of the correlation coefficient (invariance to change of location and scale, symmetry), ecological correlations, correlation does not imply causation, and the SD line.

Regression: the graph of averages, regression to the mean, the regression method for individuals, the regression fallacy, the two regression lines, the r.m.s. error, plotting residuals, hetero and homoscedasticity, looking at vertical strips, the slope and the intercept, the method of least squares, and regression cautions.

Unit 4:

Probability: the long run argument, conditional probabilities, the multiplication rule, the subtraction rule, independence, listing the ways, the addition rule, and mutually exclusive events. What does the law of averages really say?

Unit 5:

Chance processes: making box models, the sum of the draws, the expected value, the standard error, probability histograms, the normal approximation for probability histograms, and the Central Limit Theorem.
Unit 6:

Sampling: parameters versus statistics, The Literary Digest poll, the year the polls elected Dewey, using chance in survey work, how well probability methods work, the importance of a simple random sample, a close look at the Gallup poll, telephone surveys, chance error, non-response bias, sampling bias, quota samples, samples of convenience, voluntary response samples, multi-stage cluster samples.

Chance errors in sampling: the standard error, the accuracy of percentages, the accuracy of averages, the $SE$ and expected value, confidence intervals, and their interpretation.

Unit 7:

Tests of significance: null and alternative hypotheses, test statistics and significance levels, the role of the box model, zero-one boxes, the one-sample $z$-tests and $t$-tests, $p$-values, and statistical significance. The standard error for a difference, comparing two sample averages, comparing two proportions, experiments. The $\chi^2$-tests for independence and goodness-of-fit. Is the result significant? Is it important? Data snooping, the importance of a box model.
# Appendix I

**Numerical and Graphical Summaries with Crossovers Removed**

## Table I.1

*Descriptive and Summary Statistics of Participants for Research Question 1, Crossovers Removed*

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 517)</th>
<th>Pre-FACs (N=345)</th>
<th>FACs (N = 172)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>318</td>
<td>61.5</td>
<td>207</td>
<td>60.0</td>
</tr>
<tr>
<td>Male</td>
<td>199</td>
<td>38.5</td>
<td>138</td>
<td>40.0</td>
</tr>
<tr>
<td>Course</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stat1040</td>
<td>345</td>
<td>66.7</td>
<td>230</td>
<td>66.7</td>
</tr>
<tr>
<td>Stat1045</td>
<td>172</td>
<td>33.3</td>
<td>115</td>
<td>33.3</td>
</tr>
<tr>
<td>Class rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>252</td>
<td>48.7</td>
<td>160</td>
<td>46.4</td>
</tr>
<tr>
<td>Sophomore</td>
<td>145</td>
<td>28.0</td>
<td>101</td>
<td>29.3</td>
</tr>
<tr>
<td>Junior</td>
<td>79</td>
<td>15.3</td>
<td>52</td>
<td>15.1</td>
</tr>
<tr>
<td>Senior</td>
<td>41</td>
<td>7.9</td>
<td>32</td>
<td>9.3</td>
</tr>
<tr>
<td>Age</td>
<td>20.75</td>
<td>4.09</td>
<td>20.72</td>
<td>3.27</td>
</tr>
<tr>
<td>Final grade %</td>
<td>81.86</td>
<td>10.00</td>
<td>81.41</td>
<td>9.75</td>
</tr>
<tr>
<td>ALEKS Score</td>
<td>33.65</td>
<td>11.47</td>
<td>33.80</td>
<td>10.94</td>
</tr>
</tbody>
</table>

*Note.* Pre-FACs semesters were the Fall 2017 and Spring 2018 students enrolled in Stat 1040 or Stat 1045. Spring 2019 was the FACs semester. All differences between the Pre-FACs and FACs groups are not statistically significant.
Figure I.1

Histogram of ALEKS Scores in Pre-FACs and FACs Semesters, Crossovers Removed

Note. The line at 29.5 represents the discontinuity, the ALEKS score which separates the two introductory statistics courses. $N = 517.$
Figure I2

Histogram of Final Grades in Pre-FACs and FACs Semesters, Crossovers Removed

Note. N = 517.
Appendix J

Numerical and Graphical Summaries with Extreme Cases Removed

Table J.1

Descriptive and Summary Statistics of Participants for Research Question 1, Extreme Cases Removed

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 497)</th>
<th>Pre-FACs (N = 327)</th>
<th>FACs (N = 170)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>307 61.8</td>
<td>196 59.9</td>
<td>111 65.3</td>
<td>.285</td>
</tr>
<tr>
<td>Male</td>
<td>190 38.2</td>
<td>131 40.1</td>
<td>59 34.7</td>
<td></td>
</tr>
<tr>
<td>Course</td>
<td></td>
<td></td>
<td></td>
<td>.999</td>
</tr>
<tr>
<td>Stat1040</td>
<td>281 56.5</td>
<td>185 56.6</td>
<td>96 56.5</td>
<td></td>
</tr>
<tr>
<td>Stat 1045</td>
<td>216 43.5</td>
<td>142 43.4</td>
<td>74 43.5</td>
<td></td>
</tr>
<tr>
<td>Class rank</td>
<td></td>
<td></td>
<td></td>
<td>.591</td>
</tr>
<tr>
<td>Freshman</td>
<td>242 48.7</td>
<td>153 46.8</td>
<td>89 52.4</td>
<td></td>
</tr>
<tr>
<td>Sophomore</td>
<td>138 27.8</td>
<td>95 29.1</td>
<td>43 25.3</td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>78 15.7</td>
<td>51 15.6</td>
<td>27 15.9</td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td>39 7.8</td>
<td>28 8.6</td>
<td>11 6.5</td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>20.76 4.01</td>
<td>20.67 3.04</td>
<td>20.92 5.41</td>
<td>.576</td>
</tr>
<tr>
<td>Final grade %</td>
<td>81.32 9.72</td>
<td>81.06 9.60</td>
<td>81.81 9.96</td>
<td>.418</td>
</tr>
<tr>
<td>ALEKS Score</td>
<td>31.10 8.12</td>
<td>31.20 7.85</td>
<td>30.90 8.63</td>
<td>.697</td>
</tr>
</tbody>
</table>

Note: Pre-FACs semesters were the Fall 2017 and Spring 2018 students enrolled in Stat 1040 or Stat 1045. Spring 2019 was the FACs semester. All differences between the Pre-FACs and FACs groups are not statistically significant. N = 497.
Figure J.1

Histogram of ALEKS Scores in Pre-FACs and FACs Semesters, Extreme Cases Removed

Note. The line at 29.5 represents the discontinuity, the ALEKS score which separates the two introductory statistics courses. $N = 497$. 

Figure J.2

Histogram of Final Grades in Pre-FACs and FACs Semesters, Crossovers Removed

Note. N = 497.
### Table K.1

**Descriptive and Summary Statistics of Participants for Research Question 1, Crossovers and Extreme Cases Removed**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 453)</th>
<th>Pre-FACs (N=300)</th>
<th>FACs (N = 153)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M</td>
<td>SD</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>276</td>
<td>60.9</td>
<td>178</td>
<td>59.3</td>
</tr>
<tr>
<td>Male</td>
<td>177</td>
<td>39.1</td>
<td>122</td>
<td>40.7</td>
</tr>
<tr>
<td>Course</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stat1040</td>
<td>281</td>
<td>62.0</td>
<td>185</td>
<td>61.7</td>
</tr>
<tr>
<td>Stat 1045</td>
<td>172</td>
<td>38.0</td>
<td>115</td>
<td>38.3</td>
</tr>
<tr>
<td>Class rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>226</td>
<td>49.9</td>
<td>142</td>
<td>47.3</td>
</tr>
<tr>
<td>Sophomore</td>
<td>125</td>
<td>27.6</td>
<td>89</td>
<td>29.7</td>
</tr>
<tr>
<td>Junior</td>
<td>68</td>
<td>15</td>
<td>44</td>
<td>14.7</td>
</tr>
<tr>
<td>Senior</td>
<td>34</td>
<td>7.5</td>
<td>25</td>
<td>8.3</td>
</tr>
<tr>
<td>Age</td>
<td>20.79</td>
<td>4.14</td>
<td>20.69</td>
<td>3.09</td>
</tr>
<tr>
<td>Final grade %</td>
<td>80.98</td>
<td>9.83</td>
<td>80.73</td>
<td>9.68</td>
</tr>
<tr>
<td>ALEKS Score</td>
<td>30.71</td>
<td>8.32</td>
<td>30.91</td>
<td>8.08</td>
</tr>
</tbody>
</table>

*Note.* Pre-FACs semesters were the Fall 2017 and Spring 2018 students enrolled in Stat 1040 or Stat 1045. Spring 2019 was the FACs semester. All differences between the Pre-FACs and FACs groups are not statistically significant. N = 453.
Figure K.1

Histogram of ALEKS Scores in Pre-FACs and FACs Semesters, Crossovers and Extreme Cases Removed

Note. The line at 29.5 represents the discontinuity, the ALEKS score which separates the two introductory statistics courses. $N=453$. 
Figure K.2

Histogram of Final Grades in Pre-FACs and FACs Semesters for Students with Crossovers and Extreme Cases Removed

Note. N = 453.
Appendix L

Simple Slopes Analysis

Table L.1

Simple Slopes Analysis

<table>
<thead>
<tr>
<th>Model</th>
<th>ALEKS group</th>
<th>Pre-FACs</th>
<th>FACs</th>
<th>Pre-FACs</th>
<th>FACs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>b</td>
<td>SE</td>
<td>t</td>
<td>p</td>
</tr>
<tr>
<td>1</td>
<td>Below 30</td>
<td>0.10</td>
<td>0.19</td>
<td>0.53</td>
<td>.60</td>
</tr>
<tr>
<td></td>
<td>At Least 30</td>
<td>0.26</td>
<td>0.07</td>
<td>3.84</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>2</td>
<td>Below 30</td>
<td>0.10</td>
<td>0.19</td>
<td>0.53</td>
<td>.59</td>
</tr>
<tr>
<td></td>
<td>At Least 30</td>
<td>0.46</td>
<td>0.16</td>
<td>2.95</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>3</td>
<td>Below 30</td>
<td>0.09</td>
<td>0.19</td>
<td>0.49</td>
<td>0.63</td>
</tr>
<tr>
<td></td>
<td>At Least 30</td>
<td>0.29</td>
<td>0.07</td>
<td>3.80</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>4</td>
<td>Below 30</td>
<td>0.09</td>
<td>0.19</td>
<td>0.47</td>
<td>0.64</td>
</tr>
<tr>
<td></td>
<td>At Least 30</td>
<td>0.51</td>
<td>0.16</td>
<td>3.10</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Model 1 was fit on 576 participants in 45 recitation sections. Model 2 was fit to the data with 79 extreme cases removed on 497 observations nested in 45 recitation sections. Model 3 was fit to the data with 59 crossovers removed on 517 participants nested in 45 recitation sections. Model 4 was fit to the data with crossovers and extreme cases removed on 453 participants nested in 45 recitation sections.
Appendix M

LOESS Curve on Model 1

Figure M.1

LOESS Curve on Model 1
Appendix N

Side by Side Plots of Final Grade Percentage at the Cutoff, Pre-FACs to FACs Semester

Figure N.1

Side by Side Plots of Final Grade Percentage at the Cutoff, Pre-FACs to FACs Semester

Note. Model 1 was fit on 576 participants in 45 recitation sections. Model 2 was fit to the data with 79 extreme cases removed on 497 observations nested in 45 recitation sections. Model 3 was fit to the data with 59 crossovers removed on 517 participants nested in 45 recitation sections. Model 4 was fit to the data with crossovers and extreme cases removed on 453 participants nested in 45 recitation sections.
Appendix O

Descriptive and Summary Statistics by Prerequisite

Table O.1

Descriptive and Summary Statistics by Prerequisite

<table>
<thead>
<tr>
<th>Prerequisite</th>
<th>Total</th>
<th>Neither</th>
<th>Post Only</th>
<th>Pre &amp; Post</th>
<th>Pre Only</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N = 465</td>
<td>N = 24</td>
<td>N = 28</td>
<td>N = 356</td>
<td>N = 57</td>
</tr>
<tr>
<td>ALEKS</td>
<td>274</td>
<td>18</td>
<td>21</td>
<td>200</td>
<td>35</td>
</tr>
<tr>
<td></td>
<td>58.9%</td>
<td>75%</td>
<td>75%</td>
<td>56.2%</td>
<td>61.4%</td>
</tr>
<tr>
<td>ACT</td>
<td>80</td>
<td>3</td>
<td>3</td>
<td>66</td>
<td>8</td>
</tr>
<tr>
<td></td>
<td>17.2%</td>
<td>12.5%</td>
<td>10.7%</td>
<td>18.5%</td>
<td>14.0%</td>
</tr>
<tr>
<td>Math 1050</td>
<td>57</td>
<td>1</td>
<td>0</td>
<td>47</td>
<td>9</td>
</tr>
<tr>
<td></td>
<td>12.3%</td>
<td>4.2%</td>
<td>0.0%</td>
<td>13.2%</td>
<td>15.8%</td>
</tr>
<tr>
<td>Math 0995</td>
<td>18</td>
<td>0</td>
<td>2</td>
<td>15</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>3.9%</td>
<td>0.0%</td>
<td>7.1%</td>
<td>4.2%</td>
<td>1.8%</td>
</tr>
<tr>
<td>Math 1210</td>
<td>20</td>
<td>1</td>
<td>0</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td></td>
<td>4.3%</td>
<td>4.2%</td>
<td>0.0%</td>
<td>4.8%</td>
<td>3.5%</td>
</tr>
<tr>
<td>NA</td>
<td>10</td>
<td>1</td>
<td>2</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>2.2%</td>
<td>4.2%</td>
<td>7.1%</td>
<td>1.7%</td>
<td>1.8%</td>
</tr>
<tr>
<td>RSAT</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td></td>
<td>0.6%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.6%</td>
<td>1.8%</td>
</tr>
<tr>
<td>AP Calc</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.4%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.6%</td>
<td>0.0%</td>
</tr>
<tr>
<td>Math 1010</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>0.2%</td>
<td>0.0%</td>
<td>0.0%</td>
<td>0.3%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

p = .670
Appendix P

Descriptive and Summary Statistics of Students who Took the ALEKS Math Placement Exam
Table P.1

Descriptive and Summary Statistics of Students who Took the ALEKS Math Placement Exam

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (N = 274)</th>
<th>Neither (N = 18)</th>
<th>Post only (N = 21)</th>
<th>Pre &amp; post (N = 200)</th>
<th>Pre only (N = 35)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>M</td>
<td>SD</td>
<td>n</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>76</td>
<td>27.7</td>
<td>9</td>
<td>50.0</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>198</td>
<td>72.3</td>
<td>9</td>
<td>50.0</td>
<td>13</td>
</tr>
<tr>
<td>Semester</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fall</td>
<td>124</td>
<td>45.3</td>
<td>1</td>
<td>55.6</td>
<td>8</td>
</tr>
<tr>
<td>Spring</td>
<td>150</td>
<td>54.7</td>
<td>8</td>
<td>44.4</td>
<td>13</td>
</tr>
<tr>
<td>Course</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Stat1040</td>
<td>154</td>
<td>56.2</td>
<td>1</td>
<td>66.7</td>
<td>12</td>
</tr>
<tr>
<td>Stat1045</td>
<td>120</td>
<td>43.8</td>
<td>6</td>
<td>33.3</td>
<td>9</td>
</tr>
<tr>
<td>Class Rank</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>98</td>
<td>35.8</td>
<td>1</td>
<td>5.6</td>
<td>6</td>
</tr>
<tr>
<td>Sophomore</td>
<td>100</td>
<td>36.5</td>
<td>8</td>
<td>44.4</td>
<td>10</td>
</tr>
<tr>
<td>Junior</td>
<td>45</td>
<td>16.4</td>
<td>4</td>
<td>22.2</td>
<td>3</td>
</tr>
<tr>
<td>Senior</td>
<td>31</td>
<td>11.3</td>
<td>5</td>
<td>27.8</td>
<td>2</td>
</tr>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ALEKS Score</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Final Grade %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix Q
Additional Descriptive and Summary Statistics

Table Q.1
Descriptive and Summary Statistics of Participants by Major

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (n = 413)</th>
<th>Pre only (n = 57)</th>
<th>Pre &amp; post (n = 356)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Major</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Arts/Humanities</td>
<td>38</td>
<td>9.2</td>
<td>7</td>
<td>12.3</td>
</tr>
<tr>
<td>Biology</td>
<td>8</td>
<td>1.9</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>Business</td>
<td>12</td>
<td>2.9</td>
<td>3</td>
<td>5.3</td>
</tr>
<tr>
<td>Chemistry</td>
<td>2</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Economics</td>
<td>2</td>
<td>0.5</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Education</td>
<td>24</td>
<td>5.8</td>
<td>1</td>
<td>1.8</td>
</tr>
<tr>
<td>Engineering</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Medicine/Pre-Medicine</td>
<td>90</td>
<td>21.8</td>
<td>18</td>
<td>31.6</td>
</tr>
<tr>
<td>Other</td>
<td>135</td>
<td>32.7</td>
<td>12</td>
<td>21.1</td>
</tr>
<tr>
<td>Psychology</td>
<td>72</td>
<td>17.4</td>
<td>12</td>
<td>21.1</td>
</tr>
<tr>
<td>Sociology/Social Work</td>
<td>28</td>
<td>6.8</td>
<td>3</td>
<td>5.3</td>
</tr>
<tr>
<td>Statistics</td>
<td>1</td>
<td>0.2</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>
Appendix R

Summary Statistics of All Students by Recitation TA

Table R.1

Student Participation by Recitation Teacher (TA)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total</th>
<th>Pre Only</th>
<th>Post Only</th>
<th>Pre &amp; Post</th>
<th>Neither</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(n = 465)</td>
<td>(n = 57)</td>
<td>(n = 28)</td>
<td>(n = 356)</td>
<td>(n = 24)</td>
</tr>
<tr>
<td>TA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>20 4.3</td>
<td>5 8.8</td>
<td>2 7.1</td>
<td>13 3.7</td>
<td>0 0.0</td>
</tr>
<tr>
<td>2</td>
<td>55 11.8</td>
<td>8 14</td>
<td>1 3.6</td>
<td>42 11.8</td>
<td>4 16.7</td>
</tr>
<tr>
<td>3</td>
<td>36 7.7</td>
<td>5 8.8</td>
<td>3 10.7</td>
<td>26 7.3</td>
<td>2 8.3</td>
</tr>
<tr>
<td>4</td>
<td>25 5.4</td>
<td>1 1.8</td>
<td>4 14.3</td>
<td>19 5.3</td>
<td>1 4.2</td>
</tr>
<tr>
<td>5</td>
<td>57 12.3</td>
<td>4 7.0</td>
<td>4 14.3</td>
<td>44 12.4</td>
<td>5 20.8</td>
</tr>
<tr>
<td>6</td>
<td>21 4.5</td>
<td>3 5.3</td>
<td>0 0.0</td>
<td>17 4.8</td>
<td>1 4.2</td>
</tr>
<tr>
<td>7</td>
<td>53 11.4</td>
<td>9 15.8</td>
<td>2 7.1</td>
<td>41 11.5</td>
<td>1 4.2</td>
</tr>
<tr>
<td>8</td>
<td>17 3.7</td>
<td>0 0.0</td>
<td>1 3.6</td>
<td>13 3.7</td>
<td>3 12.5</td>
</tr>
<tr>
<td>9</td>
<td>38 8.2</td>
<td>1 1.8</td>
<td>3 10.7</td>
<td>32 9.0</td>
<td>2 8.3</td>
</tr>
<tr>
<td>10</td>
<td>14 3.0</td>
<td>4 7.0</td>
<td>0 0.0</td>
<td>10 2.8</td>
<td>0 0.0</td>
</tr>
<tr>
<td>11</td>
<td>104 22.4</td>
<td>15 26.3</td>
<td>6 21.4</td>
<td>79 22.2</td>
<td>4 16.7</td>
</tr>
<tr>
<td>12</td>
<td>5 1.1</td>
<td>0 0.0</td>
<td>1 3.6</td>
<td>3 0.8</td>
<td>1 4.2</td>
</tr>
<tr>
<td>13</td>
<td>20 4.3</td>
<td>2 3.5</td>
<td>1 3.6</td>
<td>17 4.8</td>
<td>0 0.0</td>
</tr>
</tbody>
</table>
Table R.2

Participants in the Analysis by Recitation TA

<table>
<thead>
<tr>
<th>Variable</th>
<th>Total (n = 441)</th>
<th>Pre only (n = 57)</th>
<th>Post only (n = 28)</th>
<th>Pre &amp; post (n = 356)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
<td>n</td>
</tr>
<tr>
<td>TA</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TA 1</td>
<td>20</td>
<td>4.5</td>
<td>5</td>
<td>8.8</td>
<td>2</td>
</tr>
<tr>
<td>TA 2</td>
<td>51</td>
<td>11.6</td>
<td>8</td>
<td>14.0</td>
<td>1</td>
</tr>
<tr>
<td>TA 3</td>
<td>34</td>
<td>7.7</td>
<td>5</td>
<td>8.8</td>
<td>3</td>
</tr>
<tr>
<td>TA 4</td>
<td>24</td>
<td>5.4</td>
<td>1</td>
<td>1.8</td>
<td>4</td>
</tr>
<tr>
<td>TA 5</td>
<td>52</td>
<td>11.8</td>
<td>4</td>
<td>7.0</td>
<td>4</td>
</tr>
<tr>
<td>TA 6</td>
<td>20</td>
<td>4.5</td>
<td>3</td>
<td>5.3</td>
<td>0</td>
</tr>
<tr>
<td>TA 7</td>
<td>52</td>
<td>11.8</td>
<td>9</td>
<td>15.8</td>
<td>2</td>
</tr>
<tr>
<td>TA 8</td>
<td>14</td>
<td>3.2</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>TA 9</td>
<td>36</td>
<td>8.2</td>
<td>1</td>
<td>1.8</td>
<td>3</td>
</tr>
<tr>
<td>TA 10</td>
<td>14</td>
<td>3.2</td>
<td>4</td>
<td>7.0</td>
<td>0</td>
</tr>
<tr>
<td>TA 11</td>
<td>100</td>
<td>22.7</td>
<td>15</td>
<td>26.3</td>
<td>6</td>
</tr>
<tr>
<td>TA 12</td>
<td>4</td>
<td>0.9</td>
<td>0</td>
<td>0.0</td>
<td>1</td>
</tr>
<tr>
<td>TA 13</td>
<td>20</td>
<td>4.5</td>
<td>2</td>
<td>3.5</td>
<td>1</td>
</tr>
</tbody>
</table>
Appendix S

Histograms of Attitude Components

Figure S.1

*Histograms of Students’ Pre-and Post-Survey Affect Scores*
Figure S.2

*Histograms of Students’ Pre-and Post-Survey Cognitive Competence Scores*
Figure S.3

Histograms of Students’ Pre-and Post-Survey Difficulty Scores
Figure S.4

Histograms of Students’ Pre-and Post-Survey Effort Scores
Figure S.5

Histograms of Students’ Pre-and Post-Survey Interest Scores
Figure S.6

Histograms of Students’ Pre-and Post-Survey Value Scores
Appendix T

Person-Profile Violin Plots by Course for Attitude Components

Figure T.1

*Person-Profile Violin Plot by Course for Affect*
Figure T.2

Person-Profile Violin Plot by Course for Difficulty
Figure T.3
Person-Profile Violin Plot by Course for Effort
Figure T.4

Person-Profile Violin Plot by Course for Interest
Figure T.5

*Person-Profile Violin Plot by Course for Value*
Appendix U

Estimated Marginal Means for Female Students’ Changes in Attitude Components from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95%

Figure U.1

Estimated Marginal Means for Female Students’ Changes in Attitude Components from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95%

![Diagram showing changes in attitude components from pre- to post-survey for different final course grades.](image-url)
Appendix V

Estimated Marginal Mean Plots by Sex on Attitude Components Change from Pre-
to Post-Survey for Final Course Grades of 75%, 85%, and 95%
Figure V.1

Estimated Marginal Mean Plots by Sex on Attitude Components Change from Pre- to Post-Survey for Final Course Grades of 75%, 85%, and 95%
Appendix W

MLM Models for ALEKS Subset by Attitude Component
Table W.1

**MLM Models for ALEKS Subset by Attitude Component**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Affect</th>
<th>Cognitive competence</th>
<th>Difficulty</th>
<th>Effort</th>
<th>Interest</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>SE</td>
<td>Var</td>
<td>Size</td>
<td>b</td>
<td>SE</td>
</tr>
<tr>
<td>Fixed effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept</td>
<td>4.30***</td>
<td>0.13</td>
<td>4.60***</td>
<td>0.18</td>
<td>3.60***</td>
<td>0.09</td>
</tr>
<tr>
<td>Attempt&lt;sup&gt;a&lt;/sup&gt;</td>
<td>0.70***</td>
<td>0.08</td>
<td>0.59***</td>
<td>0.07</td>
<td>0.56***</td>
<td>0.05</td>
</tr>
<tr>
<td>Final grade&lt;sup&gt;b&lt;/sup&gt;</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>-0.00</td>
<td>0.01</td>
</tr>
<tr>
<td>Sex&lt;sup&gt;c&lt;/sup&gt;</td>
<td>-0.63***</td>
<td>0.13</td>
<td>-0.47***</td>
<td>0.12</td>
<td>-0.25*</td>
<td>0.10</td>
</tr>
<tr>
<td>Attempt&lt;sup&gt;a&lt;/sup&gt; x Final grade</td>
<td>0.06***</td>
<td>0.01</td>
<td>0.05***</td>
<td>0.01</td>
<td>0.03***</td>
<td>0.01</td>
</tr>
<tr>
<td>ALEKS score</td>
<td>0.01**</td>
<td>0.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Random effects</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TA</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Student</td>
<td>0.47</td>
<td>0.36</td>
<td>0.32</td>
<td></td>
<td>0.15</td>
<td></td>
</tr>
<tr>
<td>Residual</td>
<td>0.62</td>
<td>0.49</td>
<td>0.30</td>
<td>0.22</td>
<td>0.67</td>
<td>0.39</td>
</tr>
<tr>
<td>Participants</td>
<td>441</td>
<td>441</td>
<td>441</td>
<td>441</td>
<td>441</td>
<td>441</td>
</tr>
<tr>
<td>Observations</td>
<td>797</td>
<td>797</td>
<td>797</td>
<td>797</td>
<td>797</td>
<td>796</td>
</tr>
</tbody>
</table>

<sup>a</sup>0 = pre-survey, 1 = post-survey.  
<sup>b</sup>Final grade is centered at 85.  
<sup>c</sup>0 = male, 1 = female.  
*p < .05. **p < .01. ***p < .001.
Appendix X

Permission to Reprint from *Statistics Education Research Journal*

From: Peters, Susan <s.peters@louisville.edu>
Sent: Thursday, June 15, 2023 2:08 PM
To: KimberLeigh Hadfield <k.hadfield@usu.edu>; Noleine Fitzallen <n.fitzallen@unsw.edu.au>
Cc: Ayse Bilgin <ayse.bilgin@mq.edu.au>
Subject: Re: Conflict with Article and Dissertation?

Hi KimberLeigh,

Congratulations on your successful defense!

I made a similar inquiry to IASE and ISI regarding an ICOTS paper that was linked to an author's dissertation. In that case, the person wanted to include the ICOTS paper verbatim in their dissertation, and the word that I got back at that point was: "For the ones related to PhD thesis, it is okay to include the papers in the thesis provided that they include all information about where the paper was published (citation) and a web link." Given this previous response, I think that the acknowledgement you mention (along with the full reference) should be sufficient. I have cc'd Ayse Bilgin as the person from IASE that I contacted previously to weigh in should my interpretation be wrong.

Sue
CURRICULUM VITAE

KIMBERLEIGH FELIX HADFIELD

Senior Lecturer, Department of Mathematics & Statistics
Utah State University
3900 Old Main Hill, Logan, UT 84322
Email: k.hadfield@usu.edu

EDUCATION

Ph.D. in Education, (August 2023)
Utah State University, Logan, Utah
Specialization: Curriculum and Instruction
Concentration: Mathematics Education and Leadership
Dissertation Committee Chair: Katherine Vela, Ph.D.

Master of Arts, Mathematics, August 1996
Brigham Young University, Provo, Utah
Minor: Statistics
Advisor: Steven R. Williams, Ph.D.

Bachelor of Science, Mathematics Education, December 1994
Utah State University, Logan, Utah
Minor: English (Magna Cum Laude)

CERTIFICATIONS
2017 Scientific Teaching Fellow, Yale Center for Teaching Excellence
1996 Utah State Secondary Teaching Certificate--Level 4 Math and English Endorsements

AWARDS & RECOGNITION

Department of Mathematics and Statistics Faculty Service Award, 2021.
College of Science Teacher of the Year, Utah State University, 2020.
Department of Mathematics and Statistics Teacher of the Year, Utah State University, 2020.
Featured Instructor, 2019 Inclusive Excellence Symposium: “Disrupt.” Distinguished as one of five USU faculty for “designing [lectures] to the edges,” supporting students
with disabilities, which benefits all students.

**Invited Commencement Speaker**, USU Integrated Studies Bachelor’s Degree and Associate Degree Graduation Program (May 2, 2019).

**The Utah Statesman Top 10 Most Influential Faculty** (2018) nominated by students through *The Utah Statesman* Newspaper, Utah State University.

**Top Professor Award**, Mortar Board (2006), nominated by student Kyler E. Ovard, Utah State University.

**USU Department of Mathematics and Statistics Teaching Award** (1995).

**Drake Outstanding Student Teacher Award** (1995) ($150). College of Education, Utah State University

**USU Department of Mathematics and Statistics Academics Award** (1994).

**PROFESSIONAL HISTORY**

**UTAH STATE UNIVERSITY**, Logan, Utah

Department of Mathematics and Statistics

**Senior Lecture and Course Coordinator (2020 – present).**
Responsibilities include teaching two large-student section mathematics or statistics courses each semester; Introduction to Statistics course development and coordination for state-wise campuses, online and broadcast, including concurrent enrollment instructors; writing the departmental final for introduction to statistics; train and manage recitation leaders for the small section recitations. Use Microsoft Powerpoint, classroom management systems such as i-Clicker and TopHat, and TI graphing calculators as well as R, SAS and Minitab. Stat 1040/1045 Course Coordinator for Statewide USU campuses.

**Stat 1040 Concurrent Enrollment Coordinator (2016 – 2021).**
Responsibilities include visitations to high school CE Stat 1040 instructors and classrooms once a year; provide training, exam materials and departmental final to instructors; coordinate and provide curriculum.

**Lecturer and Course Coordinator (2017 – 2020).**
Responsibilities include teaching two large-student section mathematics or statistics courses each semester; Introduction to Statistics (Stat 1040/1045 Course Coordinator for Statewide USU campuses, online and broadcast), including concurrent enrollment instructors; writing the departmental final for introduction to statistics; train and manage recitation leaders for the small section recitations.

**Temporary Lecturer and Course Coordinator (2016 – 2017).**
Responsibilities include teaching two large student section mathematics or statistics courses each semester; managing small section recitations and training graduate student recitation leaders; coordinated the instruction for on-campus faculty for Introduction to
Statistics courses; wrote the departmental midterms and finals.

**Temporary Lecturer (2013 – 2016).**
Responsibilities included teaching one to two large student section mathematics or statistics courses each semester; managing small section recitations and training graduate student recitation leaders.

**Instructor (2000 – 2006).**
Responsibilities included teaching three small student section introductory statistics courses per semester.

**Teaching Assistant (1995).**
Responsibilities included teaching small student sections of college algebra.

Jon Huntsman School of Business, *China Cooperative Academic Program*

**Lead Instructor (2015 – 2021).**
Responsibilities include communicating with up to four Chinese campuses and their on-site instructors via email at least weekly providing materials, examinations, and grading of exams or providing grading rubrics for the teaching of college algebra to approximately 300-350 Chinese students each fall semester; traveling to China every few years. Travelled to Beijing November 2015. Taught the College Algebra and Business Calculus courses to students at Beijing Institute of Technology. Met with on-site instructors and directors at BIT in the CCAP program.

**WESTMINSTER COLLEGE**, Salt Lake City, Utah
School of Mathematics

**Instructor (2000 – 2002).**
Responsibilities included teaching small student section courses of introductory statistics, discrete mathematics, calculus and trigonometry. Coordinated a chapter meeting of the American Statistical Association at the college.

**BRIGHAM YOUNG UNIVERSITY**, Provo, Utah
Department of Statistics

**Visiting Lecturer and Course Manager (1998-2000).**
Responsibilities included teaching large student section introductory statistics, and small student sections of college algebra. Taught Teaching Preparation 1 and 2 (TA Preparation Courses) to teaching assistants for the small section recitations of introductory statistics. Managed of all the introductory statistics courses each semester and their teaching assistants, co-wrote departmental exams.

Member of research team: The Learning Research Initiative led by Sterling C. Hilton
Department of Mathematics

**Teaching Assistant (1995 – 1996).**
Responsibilities included teaching college algebra.

**UTAH VALLEY UNIVERSITY, Orem, Utah**
Department of Mathematics

**Adjunct Faculty (1996 – 1999).**
Responsibilities included teaching small student sections of statistics, college algebra, intermediate algebra, and trigonometry.

**ALPINE SCHOOL DISTRICT, Utah**

**District In-Service Teacher and Curriculum Coordinator for AP Statistics (1997 – 1998).**
Researched textbooks and software and wrote the curriculum for the AP Statistics course for Alpine District.
Responsible for the creation and teaching of in-service course, *Teaching AP Statistics* to district AP teachers.

**Teacher (1996 – 1998).**
**American Fork High School, American Fork, Utah**
Responsibilities included teaching AP Calculus, AP Statistics, Honors Pre-Calculus with Trigonometry, and Geometry; advised the Math Club and math competition teams.

**UTAH STATE OFFICE OF EDUCATION, Utah**

**Online Curriculum Design and Instruction (2000-2012).**
**Electronic High School, Salt Lake City, Utah; http://ehs.uen.org**
Responsibilities included developing the online curriculum and content for the online AP Calculus course; online AP Calculus instructor.

**UNIVERSITY TEACHING**

Department of Mathematics and Statistics

**STAT 1040 – Introduction to Statistics**
Descriptive and inferential statistical methods. Emphasis on conceptual understanding and statistical thinking. Examples presented from many different areas.

**STAT 1045 – Introduction to Statistics with Elements of Algebra**
Intro to statistics with an emphasis on conceptual understanding and statistical reasoning.
Foundational algebra, types of studies, summarizing data, probability, hypothesis testing.

STAT 5810 – Introduction to Data Analysis
Built exclusively for prospective CE STAT 1040 teachers, it includes material on the special pedagogy of STAT 1040 and some introductory statistics. The remainder of the course extends the topics of comparing means of two samples or groups to comparing multiple groups in a designed experiment and comparing two proportions and elementary chi-square tests to more general categorical data analysis.

MATH 1050 – College Algebra

Westminster College, Salt Lake City, Utah (2000-2002)
School of Mathematics

MATH 142 – Trigonometry
The study of trigonometric functions and their graphs, applications to navigation and surveying problems, modeling cyclic behavior, complex numbers, polar coordinates, and vectors.

MATH 150 – Elementary Statistics
An introduction to the use of statistics as a valuable tool for analyzing data in a variety of fields. Topics in elementary descriptive and inferential statistics, including the normal, binomial, Student t, and chi-square distributions, correlation and regression, confidence intervals, and hypothesis testing.

MATH 201 – Calculus I
Functions, graphs and limits. Differential calculus of algebraic, trigonometric, exponential, and logarithmic functions with applications to geometry, the physical and life sciences, and economics.

MATH 210 – Discrete Mathematics I
Topics in sets, logic, elementary counting including permutations and combinations, finite probability, sequences and mathematical induction.

Department of Statistics

STAT 121 (221) – Principles of Statistics
Graphical displays and numerical summaries, data collection methods, probability, sampling distributions, confidence intervals and hypothesis testing involving one or two means and proportions, contingency tables, correlation and simple linear regression.
STAT 510 and 511 – TA Orientation 1 and 2.

**Department of Mathematics**

**MATH 110, 110H – College Algebra, Honors College Algebra**
Functions, polynomials, theory of equations, exponential and logarithmic functions, matrices, systems of linear equations, permutations, combinations, binomial theorem.

**Utah Valley University, Orem, Utah (1996-1999)**

**Department of Mathematics**

**MATH 1010 – Intermediate Algebra**
Expands and covers in more depth basic algebra concepts introduced in Beginning Algebra. Includes linear and quadratic equations and inequalities, polynomials and rational expressions, radical and exponential expressions and equations, complex numbers, systems of linear and nonlinear equations, functions, conic sections, and real-world applications of algebra.

**MATH 1050 – College Algebra**
Includes inequalities, functions and their graphs, polynomial and rational functions, exponential and logarithmic functions, systems of linear and nonlinear equations, matrices and determinants, arithmetic and geometric sequences, and the Binomial Theorem.

**MATH 1060 – Trigonometry**
Includes the unit circle and right triangle definitions of the trigonometric functions, graphing trigonometric functions, trigonometric identities, trigonometric equations, inverse trigonometric functions, the Law of Sines and the Law of Cosines, vectors, complex numbers, polar coordinates, and rotation of axes.

**STAT 1040 – Introduction to Statistics**
A quantitative literacy course with a statistical theme. Includes descriptive statistics, sampling, and inferential methods. Emphasizes problem solving and critical thinking.

**DIRECTED STUDENT LEARNING**

**Melissa Hansen**, Masters of Statistics, Committee Member (2023-present)

**Asher Hanson**, Masters of Statistics, Committee Member (2023-present)

**Erica Hurst**, Masters of Statistics, Committee Member (2019-present)

**Bri Smith**, Honors in Practice, Mentor (2022)
INSTITUTIONAL LEADERSHIP & SERVICE

NATIONAL

External Peer Reviewer, (2021), University of Nevada-Reno.


UTAH STATE UNIVERSITY

Robins Award Selection Committee Member (2022).

Search Committee Member, Dean of the College of Science (2020-2021).

Faculty Senator, USU Faculty Senate (Fall 2019-2021).

Committee Member, USU Athletics Council (Fall 2019-2021).

Committee Member, USU Strategic Enrollment Management Process (SEMP)-Exploratory Advising Committee (Spring 2019-2020).

Reviewer, USU Honors Program (2018, 2019, 2020). Responsibilities include annually reviewing 30-40 applications for acceptance to the USU Honors Program.

UTAH STATE UNIVERSITY, Department of Mathematics and Statistics

Stat 1040/1045 Course Coordinator, USU Math and Stat Department (2017-present). Responsibilities include the coordination of teaching and materials for all instructors of the introductory statistics courses across on-campus, online and state-wide campus instructors, writing the departmental final exam for these courses each semester. Coordinate and train instructors, manage questions, concerns, pedagogical issues. Created and provide a dynamic departmental teaching repository for Stat 1040/1045 instructors.

Mentor, Graduate and Undergraduate Statistics Recitation Leaders, USU Math and Stat Department (2017-present). Evaluate the recitation leaders’ teaching of introductory statistics
recitations. Provide mentoring and leadership for the improvement of their recitation sections.

**Mentor and Trainer of Lead Recitation Leader**, USU Math and Stat Department (2017-present).
Responsibilities include training and mentoring a recitation leader who has taught recitations two or more semesters to serve as a Lead Recitation Leader for Introductory Statistics, one who is a resource for newer recitation leaders by modeling teaching, student interactions, and recitation leader responsibilities; organizes the lecture set-up and take down for instructor.

**Search Committee Member**, Math Education Professional Practice Hire (2021-2022).

**Search Committee Chair**, USU-Brigham City, Mathematics and Statistics Lecturer (2020).

**Curriculum Committee, Statistics** (2020-present).
Meet to discuss statistics education related curriculum issues pertaining to the department.

**Search Committee Member**, Math Education Professional Practice Hire (2019-2020).

**Committee Member**, Math & Stat Education Group (MSEG) (2017-present).
Meet to discuss mathematics and statistics education related issues pertaining to the department, pre-service teacher outreach, Math Ed course discussions and revisions.


**RESEARCH**

Research interests include the field of undergraduate statistics education, specifically, improving student understanding, student attitudes in statistics and decreasing statistics anxiety in large-student sections of undergraduate introductory statistics courses through innovative instruction and continuous assessment.

**GRANTS (Funded)**

**Contributing Faculty ($33,700,000) (2018-2025)**. *Utah STARS! GEAR UP Cohort 4*. U.S. Department of Education. Project Goal: A seven-year public school/industry/university partnership designed to increase middle and high school student college readiness in Granite School district. (PI – Jim Dorward, Utah State University).
PUBLICATIONS

JOURNALS


CONFERENCE PROCEEDINGS


MANUSCRIPTS


PRESENTATIONS

NATIONAL PRESENTATIONS


DEPARTMENT PRESENTATIONS, Utah State University, Logan, Utah

Hadfield, K. (2023, March). *Providing Ability to Probability: Reducing Cognitive Load*
with Worked Examples, Recitation Leader Workshop, USU Math and Stat Department, Logan, Utah


SCHOLARSHIPS


CONTINUOUS LEARNING AND SELF-DEVELOPMENT

CONFERENCES ATTENDED

Top Hat Summer Camp, Online, July 2021
Spring Intermountain MAA Section Meeting, Online, March 2021
Mathematics Teacher Educator Partnership (MTE-P), St. Louis, Missouri, June 2019
Enhance Statistics Teacher Education with E-Modules (ESTEEM), Orlando, Florida, February 2019
Association of Mathematics Teacher Educators (AMTE), Orlando, Florida, February 2019
Mathematics Teacher Education-Partnership (MTE-P), Denver, Colorado, July 2018
Scientific Teaching, Yale Center for Teaching Excellence, May 2017
Joint Statistical Meetings, Baltimore, Maryland, August 1999
Teaching Statistics with Technology Summer Institute, Berkeley, CA 1997
National Council of Teachers of Mathematics (NCTM)
  Regional Conference in Salt Lake City, February 1997
  National Conference in San Diego, April 1996
  Regional Conference in Boise, October 1994

PROFESSIONAL AND HONOR SOCIETIES

Association of Mathematics Teacher Educators
Mathematics Teacher Education-Partnership
American Statistical Association
National Council of Teachers of Mathematics
Phi Kappa Phi National Honor Society
Golden Key National Honor Society
Blue Key National Honor Society