

DETERMINING EQUIVALENCE IN LEARNING OUTCOMES FOR FRESHMAN-
LEVEL COMPOSITION COURSES TAUGHT ONLINE AND VIA FACE-TO-FACE
DELIVERY

by

Kimberly A. DeLauro, MA

DISSERTATION

Presented to the Faculty of
The University of Houston-Clear Lake
in Partial Fulfillment
of the Requirements
for the Degree

DOCTOR OF EDUCATION

THE UNIVERSITY OF HOUSTON-CLEAR LAKE

MAY, 2016

Copyright 2015, Kimberly Ann DeLauro*

*All Rights Reserved

DETERMINING EQUIVALENCE IN LEARNING OUTCOMES FOR FRESHMAN-
LEVEL COMPOSITION COURSES TAUGHT ONLINE AND VIA FACE-TO-FACE
DELIVERY

by

Kimberly A. DeLauro

APPROVED BY

Bryan Morgan, EdD, Chair

Carol A. Carman, PhD, Committee Member

Lillian McEnery, EdD, Committee Member

Jana Willis, PhD, Committee Member

RECEIVED BY THE SCHOOL OF EDUCATION:

Lawrence T. Kajs, EdD, Associate Dean

Mark D. Shermis, PhD, Dean

ACKNOWLEDGEMENTS

It is impossible to thank all of the people who have supported me throughout my life, including my graduate work and doctoral research. I know that I am not always an easy person to be around, especially when I have my nose in a book or am tapping away at a keyboard for most of the day. I thank everyone who has helped me on my never-ending educational journey and without your support, I could not have accomplished as much as I have in life.

To Andrew DeLauro, my sweet husband, I thank you so much for all the many ways you have supported me. You have worked hard to ferry the kids around and attend their sporting and educational events while I've been in class, you have often taken a back seat in many ways when I had to write, and you have been a continual support to me in more ways than I can count. Thanks for all you do and I love you very much.

To my children five—Catherine, Nicholas, Caroline, Rebecca, and Anthony DeLauro, each one of you fills me with pride for being exactly who you are. I thank you for being strong, independent, good kids, even when I asked you to make dinner for yourselves and then clean up the kitchen. A mother could never hope to have better kids than you.

To my parents, Steve and Gail Dee, thank you so much for giving me a happy childhood and a deep appreciation for books, a desire for writing, and a constant thirst for knowledge and learning, without which I could never have pursued advanced degrees.

To my extended family—thank you all for listening to my ramblings on data and inter rater reliability being injected into every conversation. Thanks to my brother, Brett Dee, his wife, Jessie, and my nephew, Eli—thank you for your prayers and continued encouragement (and Eli, thanks for just being adorable). Also, thanks to all of my extensive family of various DeLauri, Parolisi, Lodigiani, Paullisens, et al. for your continued support and love in all ways.

Many thanks to my Doctoral Cohort 6. When I began this journey, I had no idea that I would meet so many new friends. Thank you all for being so unique and fun in each of your own ways and working together so well. Thank you so much for challenging me and for listening to my stories. I am honored to know each of you and I am blessed to have been part of your lives.

Rebecca Stout, you are one of my closest friends and I appreciate all of the time you have spent talking with me about research, navigating “cage matches,” sharing articles and books, discussing Galveston real estate, and giving me advice on raising daughters. I look forward to many years of the same.

To my chair, Dr. Bryan Morgan, thank you for your wisdom and advice, not only with regards to this dissertation, but also with regards to a career in higher education in general. I love hearing your “You’ll-never-guess-what-happened-when-I-was-a-dean-have-I-told-you-this-one-before?” stories and learned more than you can imagine from each one.

To my methodologist, Dr. Carol Carman, thank you for all of your advice and guidance as I navigated through IRBs and data. Even before you were my methodologist,

you were easily one of the best instructors I have ever had. Anyone who can help this English professor understand statistics is a true master teacher.

To Dr. Lillian McEnery, thank you for being so approachable and helpful throughout this process. I also thank you for your great feedback on writing and your attention to detail with everything you have read of mine. I appreciate the “love” you give all of your students’ work.

To Dr. Jana Willis, thank you for your guidance on books to consult regarding distance learning and also for encouraging me professionally. I appreciate your help very much.

To my colleagues and friends at San Jacinto College, it is a true gift to enjoy my job so much that I look forward to each day as I head to campus. Not everyone has that experience in their lives, so I am truly thankful that I have had the honor to work at such a great institution filled with so many genuine, supportive, caring, professional people.

ABSTRACT

DETERMINING EQUIVALENCE IN LEARNING OUTCOMES FOR FRESHMAN-LEVEL COMPOSITION COURSES TAUGHT ONLINE AND VIA FACE-TO-FACE DELIVERY

Kimberly A. DeLauro
University of Houston-Clear Lake, 2016

Dissertation Chair: Bryan Morgan, EdD

Distance learning has become an increasingly important element in the landscape of higher education, with more colleges and universities considering distance learning course delivery as part of their strategic plans. While distance-learning courses are often assumed to have equal learning outcomes as their face-to-face counterparts, very few studies have quantitatively explored whether this is indeed the case. This study sought to determine whether freshman-level composition courses taught via a traditional face-to-face delivery method and courses taught via online delivery had equivalent course outcomes. For this study, essays written by community college freshman students in the second half of the fall 2014 semester in courses using both delivery modes were assessed

by three independent raters, using a rubric. The two one-sided t test (TOST) method was used to determine equivalence and ascertain whether the students' scores fell within a zone of equivalence of $\pm 30\%$ of the mean for the scores. Because the face-to-face students had consistently higher means, t tests were then used to determine whether these differences were significant. The findings of this study revealed that the face-to-face and distance learning freshman composition courses were not equivalent for any of the rubric outcome areas of essay structure, essay content, essay clarity, and use of sources. The findings also indicated that for most of the outcome areas, the means were significantly different, with the face-to-face students out-performing the distance-learning students in all of the rubric outcome areas.

TABLE OF CONTENTS

Abstract	iv
List of Tables	x
List of Figures	xi
Chapter	Page
I. INTRODUCTION	1
Need for Study	3
Purpose of Study	4
Research Questions	4
Definitions	4
Summary	6
II. LITERATURE REVIEW	7
Student Perceptions	9
Effectiveness of DL Courses	11
Developmental Students and Attrition	14
Types of Students and Success Rates	17
Learning Outcome Measures	19
Complexities and Importance of English Composition	19
Summary	21
III. METHODOLOGY	23

Population and Sample	23
Measurement of Variables	25
Research Design and Data Collection.....	25
Instrumentation	28
Pilot Test	31
Hypothesis and Data Analysis	33
Summary	37
IV. RESULTS	40
Research Question and Hypothesis One	40
Research Question and Hypothesis Two	45
Research Question and Hypothesis Three	51
Summary	56
V. FINDINGS, IMPLICATIONS, AND CONCLUSIONS	58
Summary of Findings.....	58
Face-to-Face and DL Equivalency for Freshman Composition	59
Equivalency of Courses and Ethnicity	63
Equivalency of Courses and Sex.....	65
Limitations	66
Future Research	69
Implications and Conclusions	70
REFERENCES	77
APPENDIX A	84

LIST OF TABLES

Table	Page
1 Intra-Class Correlation (ICC) Values for Pilot Study	32
2 Intra-Class Correlation (ICC) Values for Main Study	33
3 Results of Equivalency Tests for DL and F2F Rubric Scores	42
4 Results of Equivalency Tests for DL and F2F by Ethnicity	46
5 Results of Equivalency Tests for DL and F2F by Ethnicity and Outcome.....	47
6 Results of t Tests Comparing Outcomes of White Students.....	50
7 Results of t Tests Comparing Outcomes of Hispanic Students	50
8 Results of t Tests Comparing Outcomes of Black Students	51
9 Results of Equivalency Tests for Female Rubric Scores.....	53
10 Results of Equivalency Tests for Male Rubric Scores.....	53
11 Result of t Tests Comparing Outcomes of Female Students	55
12 Results of t Tests Comparing Outcomes of Male Students	56

LIST OF FIGURES

Figure	Page
1 Zone of Equivalence for TOST Tests	34
2 Confidence Intervals for F2F and DL Total Scores	43
3 Confidence Intervals for F2F and DL by Outcome	44
4 Confidence Intervals for White, Hispanic, and Black Total Scores	48
5 Confidence Intervals for Outcome Scores by Ethnicity	49
6 Confidence Intervals of Rubric Outcomes by Sex	54

CHAPTER 1

INTRODUCTION

Distance learning is not a new concept—since its early roots in postal correspondence courses in the late 1800’s, distance learning has transformed and adapted alongside technology into radio and television courses to the “real time” experiences of audio-conferencing in the 1980’s, and then into the Internet versions found today (Moore & Kearsley, 1996). Currently, distance learning courses have now become so prevalent that one is hard-pressed to find an institution that does not offer at least some type of online course, if not entire online degree programs.

Allen and Seaman (2013) reported that the number of students who were enrolled in least one online class in 2012 increased to 6.9 million and the proportion of all students who were taking an online class had risen to 32 percent, which is an all-time high. Institutions of higher education have also begun to view distance learning as an integral part of their strategic plans. In 2002, when the 2,800 institutions surveyed by Babson Research Group, in conjunction with the College Board, were asked whether they saw distance learning as part of their long-term strategy, less than half of them agreed. By contrast, when the same group was asked this same question in 2012, 69.1 percent of the institutions surveyed felt that distance learning should be an important consideration for their long-term strategy.

Massive Open Online Courses (MOOCs) have also received a great deal of press recently and are an example of the impact of distance learning in higher education. Originating in 2008, MOOCs are online courses that emerged commercially by companies who were associated with top universities and thus allowed students around the world to partake, in an online format, of courses taught at some of the best institutions in the world (Harris, 2013). In 2013, 160,000 students signed up for Stanford's Artificial Intelligence course, run by Sebastian Thrun, and 60,000 signed up for Duke's Introduction to Astronomy course on Coursera. Affordability is the primary appeal of MOOCs, which made their debut into the distance-learning scene at a time of rising education costs. In addition, MOOCs run by many prestigious institutions such as Harvard and Massachusetts Institute of Technology are attractive to people who otherwise could not afford or could not academically qualify to attend these institutions in the traditional sense. On the downside, however, the attrition rate for MOOCs has been an ongoing problem. As an example, the edX MOOC started by MIT and Harvard University offered a physics course that, in 2012, attracted 155,000 people from around the world, yet only 7,000 people finished the class. Other researchers have found similar results with regard to high attrition rates in distance learning courses (Dutton, Dutton, & Perry, 2001; Zavarella & Ignash, 2009; Ashby, Sadara, & McNary, 2011).

These programs also have the potential to make tremendous amounts of money for the institutions that sponsor them. The sheer volume of students attending these courses allow for even modest tuition charges, which can result in substantial profits. In addition to this, some of the future business models include credentialing of students, charging head-hunting companies for contact information of the most successful MOOC

users, or using a MOOC as a referral service to brick-and-mortar schools for students might be better served attending a face-to-face course (Kolowich, 2012).

Need for the Study

Distance learning is now a large part of the higher education landscape and there is reason to suspect that it will continue to expand and impact higher education in future years. With this increase in distance learning across higher education, research has been done to explore some issues related to distance learning course delivery. There have been research studies in the areas of student perceptions of online courses, specifically dealing with isolation and disconnection (Kwon, Han, Bang, & Armstrong, 2010; Reilley, Gallagher-Lepack, & Killion, 2012; Phelan, 2012), studies that explore various pedagogical and andragogical approaches of online course design and implementation (Garrison & Arbaugh, 2007; Yi, & Durrington, 2010; Akyol & Garrison, 2011; D. Armstrong, 2011; Warnock, Bingham, Driscoll, Fromal, & Rouse, 2012), as well as studies that examine success outcomes and comparisons with face-to-face and online courses, with some of them focusing on hybrid course delivery systems as well (Dutton et al., 2001; Frederickson, Reed, & Clifford, 2005; Herman & Banister, 2007; Gerlich & Sollosy, 2009). The bulk of the comparison studies examine higher-level undergraduate coursework or graduate level course work, and the vast majority of them examined math courses, with most of them focusing on higher levels of math, computer science courses, and health-science courses. One study, conducted by Zavarella and Ignash (2009), examined developmental math courses, although this was not a direct comparison study between the two delivery methods. Almost none of the studies thus far, however, have examined the mastery of learning outcomes for a freshman level writing course, yet these

courses are frequently taught online at many institutions and are assumed to be equivalent to their face-to-face versions of the course.

Purpose of the Study

The purpose of this study is to determine if courses delivered via a distance-learning platform have equivalent student outcomes as courses delivered via traditional face-to-face methods.

Research Questions

This study addresses the following research questions:

1. Do freshman-level English composition courses taught online have equivalent outcomes as English composition courses taught via face-to-face delivery?
2. Are the outcomes for English composition courses taught via online delivery and face-to-face delivery equivalent to the outcomes for the ethnicities of White, Hispanic, and Black students?
3. Are the outcomes for English composition courses taught via face-to-face delivery equivalent to the outcomes for male and female students?

Definitions

In order to ensure a consistent understanding of all of the terms mentioned in this study, the following terms are defined:

Online Courses or Distance Learning Courses: are used interchangeably in this study.

Both terms are defined as having 80% or more of the course material delivered online and such courses typically have no face-to-face meetings (Allen & Seaman, 2013).

Face-To-Face Delivery Courses: are defined by the Sloan Consortium of “web-facilitated” courses, in which 1%-29% of the course material is delivered online rather

than the definition of “traditional” course delivery, in which no material (0%) is delivered online, including the posting of syllabi or the use of online course management systems (Allen & Seaman, 2013). The reason for this distinction is because all public institutions in the state where this study took place are required to post syllabi online, thus making the traditional definition inaccurate. Furthermore, at the institution involved in this study, instructors are required to have, using a course management system, at least a minimal online presence in terms of posting their contact information, syllabus, and grades in all courses. Therefore, this study adopted the definition of “web-facilitated courses,” even though the bulk of the material is delivered via face-to-face meetings that occur on a weekly basis.

Freshman English Composition Course: in this study is defined as a freshman level composition and rhetoric course which includes the study and practice of the writing process, the study of rhetorical choices (including determining audience, purpose, structure of argument, and style), a focus on writing academic essays, and conducting critical analysis (Texas Higher Education Coordinating Board, 2013). The general learning outcomes for this course are as follows:

- Demonstrate knowledge of individual and collaborative writing processes.
Develop ideas with appropriate support and attribution.
- Write in a style appropriate to audience and purpose.
- Read, reflect, and respond critically to a variety of texts.
- Use Edited American English in academic essays (Texas Higher Education Coordinating Board, 2013).

Rubrics: refers to guidelines that rate student competencies or performance (Moskal, 2000).

Competency Outcomes: in this study are defined as the student display of the learning outcomes of a course and refers to what a student knows, has learned, and is able to do at the end of a course (Shoikova & Krumova, 2010).

Equivalence: refers to the condition of being the same or of being “capable of being placed in one-to-one correspondence” (Merriam-Webster, 1988).

Summary

This chapter discussed the prevalence of distance learning in higher education and pointed out that despite the increase in distance learning course offerings, few studies exist that examine whether distance-learning courses are equivalent to face-to-face versions of the courses. This chapter then stated that the purpose of this study was to determine if freshman composition courses taught online have equivalent outcomes to those taught via a face-to-face delivery method. Next, the research question was given and in order to ensure a consistent understanding of the concepts and terms, definitions were given for the key terms mentioned in the study.

CHAPTER 2

LITERATURE REVIEW

Moore and Kearsley (1996) stated that the most basic definition of distance learning is teaching and learning that occurs when the student is removed from the instructor by location and thus requires the use of technology in order to facilitate learning. Moore and Kearsley's (1996) definition, then, illustrates how distance learning is not a new phenomenon, even before the advent of the web and media technology. Distance learning emerged from the late 1880's by way of postal correspondence courses and can be divided into five different generations of distance education. Early correspondence study was the first generation, and this eventually led to the second generation of broadcast radio and television, including the relatively late arrival of cable-TV and Instructional TV (ITV) distance learning methods. Then, in the late 1960s and into the early 1970's, a time of change occurred with distance education offerings when the first distance learning universities opened their doors. Most of these universities were located in countries where political inequities created a market for people who could not attend traditional universities and thus saw distance learning as a viable answer. The first of these included Open Universities in Great Britain, China, Bangladesh, Korea, Pakistan, China, and others. The United States' model of "distance learning university" never took off in the same way as other countries and the United States never had a true "Open University," the closest model instead being the "consortium" forms, which involved

groups of traditional universities and offered their students some distance learning. One of the first of these consortiums was the University of Mid-America (UMA), which was a consortium designed by nine Midwest universities and based at the University of Nebraska. UMA was discontinued in 1982 due to lack of enrollment and high expenses associated with video production costs, which could be quite costly. From there, distance learning progressed alongside technological advances and included the fourth generation—teleconferencing—and finally, the computer and internet-based virtual classes that are present today.

One area of expansion is that of the Massive Open Online Courses (MOOCs). While MOOCs have received a great deal of recent attention as being a possible solution for everything from teacher shortages to a solution to financial woes, a recent study done on edX's (a joint venture between MIT and Harvard) first MOOC, "circuits and electronics," revealed that when students "worked offline with anyone on the MITx material," or "worked with someone who teaches or has expertise in this area," and with all other factors being equal, these students had a predicted score of nearly three points higher than students who did not work with someone outside of the course (Breslow, Pritchard, DeBoer, Stump, Ho, & Seaton, 2013, p. 20). The study also revealed—as many other studies similarly revealed—that attrition was a factor and the study revealed that less than 5% of the students completed the course. Of the 154,763 students who registered for the course, only 23,349 tried the first problem set, 10,547 made it to the mid-term, 9,318 passed the midterm, 8,240 took the final exam, and a mere 7,157 earned a certificate.

Another study examined using MOOCs to reach lower level math students at San Jose State University. Rather than doing a comparison of students who passed or failed the course (although this data was one component of the study), this study sought to explore the contributing factors of why students were or were not successful. They concluded that students who spent more time in their online class were more likely to pass the class and students who used support services offered also fared better in the classes (Firmin et al., 2014). This study also examined the student perspectives and described their experiences, but found that the interview sample was overrepresented by students who had fared well in the class, with the poor performing students not filling out the survey. The larger question that emerged throughout this study was how to encourage students to make use of the support services offered so that success rates could be improved.

Student Perceptions

Much of the research done comparing distance learning and face-to-face learning methods have been qualitative studies assessing student perceptions of online experiences and how these perceptions can impact success. One often-studied area of distance learning is that of student isolation and its effect on student learning. Kwon et al. (2010) conducted a case study exploring three Asian students' perceptions of an online course taken at a major research university and found that all three students had a negative perception of their social interactions in the online course and felt disconnected from other students in the class. This study also revealed that written messages to classmates also contain "deep nuance" that is not always understood by foreign students. Additionally, one of the students interviewed pointed out that when she needed help, she

sought it from "real" people in her life and did not consult her online classmates or even the professor of the course. Breslow et al. (2013) found a similar pattern of online students seeking help from "offline" people outside the course.

Reilley et al. (2012) also studied students' perceptions of isolation in online courses. This study consisted of telephone focus groups with 18 participants, 89% female. The study found that there were 15 themes that emerged from the focus group data, which they organized into structural, processual, and emotional factors. The emotional factors again included the themes of aloneness, anonymity, nonverbal communication, trepidations, and unknowns and over all, of the students who participated in the focus groups, those who had more technical knowledge and had a higher comfort level with computers in general were more likely to perceive the experience of an online course more positively. There is a common theme in qualitative studies with students who took an online version of a course reporting a sense of isolation or lack of connection with their instructors or their classmates (Cross, 1998; Shaw & Polovina, 1999; Chernish, DeFranco, Lindner, & Dooley, 2005; Erichsen & Bollinger, 2011).

On a similar note, Akyol and Garrison (2011) found a strong connection between collaboration in an online classroom and cognitive processes. Students in this study seemed to learn best in a social environment, which can be more difficult to attain in a distance-learning environment. In a different study on connection in an online classroom, Phelan (2012) compared post-graduate students' experiences as both online students and online teachers and evaluated how their experiences as teachers may help their students foster a sense of belonging in an online environment. The study helped to uncover some potential pitfalls that might exacerbate problems in distance learning: for example,

through interviews, it became apparent that the immediacy of lecturer responses to student questions has a tremendous impact on students and their feelings of connectedness in an online course. Pena (2004) found similar results in a study that examined best practices in online course design in a teacher preparation program.

Effectiveness of DL Courses

Many studies have examined the effectiveness of distance learning courses, with varying results. Several studies indicated that there was no statistical difference between face-to-face and distance learning delivered courses (Chernish et al, 2005; Frederickson, et al., 2005; Herman & Banister, 2007; Ashby et al., 2011). Others, however, such as Burkhardt, Kinnie, and Courhoyer (2008) found that distance learning students in an upper-level library science class actually fared better than the face-to-face students in the same course.

Herman and Banister (2007) performed a study on a graduate level curriculum-development course that had been redesigned as a distance learning offering to better accommodate more students and save money for Bowling Green State University. The intent of this research was to “explore pedagogical issues related to an online graduate course in curriculum theory and practice from the perspective of teacher” (p.319). The study involved the collection of scoring rubrics for projects and papers completed by students in both versions of the course, enrollment and cost-analysis data, archival records of online chats, and other information. This study occurred just after the course redesign and rather than attempting to explore the effectiveness of a distance learning class versus a face-to-face class, the primary focus of the study was to assess the effectiveness of this one particular class at this one particular school. Part of the

methodology involved collecting student data from the distance education class (43 students) and three sections of the traditional course (3 sections of 52 students each). Course projects were scored by two faculty members at the institution and then they “compared scores and worked to resolve any variance in scoring, clarifying the rubric and processes” (p. 323). These results were compared to the face-to-face students’ projects, which were scored with the same rubric and “preliminary results indicated no significant differences in the learning outcomes of online vs. f2f students” (p. 324). The remainder of the study focused on how well the course design held up to Bloom’s Taxonomy. Other researchers (see Ashby et al., 2011) have cited this study to demonstrate evidence of online and face-to-face learning outcomes having no statistical difference. However, this study had a) a statistically acceptable, but very small sample size; b) the researchers had a stake in the new course redesign being effective, so a potential conflict of interest in the outcome might have been present, c) there is no mention that the papers and projects were blind-graded; and d) no quantitative analyses seem to have been performed on the results, as indicated by the phrase “preliminary results” (p. 324).

In another mixed-methods study, Chernish et al. (2005) analyzed three different modes of course delivery: traditional face-to-face, online, and via instructional television. The course studied was a junior-level hospitality and human resource management course and one of the research questions asked whether or not the three different delivery methods were equally effective. They also asked if the different methods affected the learners’ feelings about being “part of the class” and whether the different methods affected the learners’ access to resources (p. 89). The methodology involved a pretest, posttest design, along with a survey instrument to compare student perceptions. The three

course sections included 34 students in the traditional classroom, 31 students in the instructional television class, and 18 students in the online via the distance-learning class. Students were administered a pretest at the beginning of the semester and at the end of the semester, the students were again administered the same set of questions as a final evaluation of the knowledge gained in the course. The results indicated that the difference between the groups was not statistically significant and thus, they found that “the delivery method did not contribute to any difference in the learners’ achievement level” (p. 91). The primary focus of this study dealt with “student perception,” with less of a focus on the effectiveness of the method. The sample size for this study consisted of 18 students for the online section, with a pretest posttest as the measure of student achievement. Attrition was not mentioned at all in this study, although it has been noted in most other studies. The result of this element of the study showed that the qualitative analysis indicated that students who had chosen the online version of the course were less satisfied with their choice throughout the course and their comfort level remained low, as compared with the other two methods.

In another study done by Frederickson et al. (2005), 16 students were randomly divided into two groups ($n=8$) and assigned to one of two teaching environments—face-to-face or online—for a graduate level statistics course. They were given a paper and pencil test to assess their math and statistics anxiety in addition to a pretest of the material meant to be covered in the course. At the end of three weeks, the students were given a posttest of the material covered in the class as well as a follow-up anxiety test, which measured “self-perceived confidence and competence with research methods and statistics” (p. 651). Then, the groups switched and for the next three weeks, the face-to-

face class experienced the online version and vice versa. The student achievement results of this study indicate “there was little difference in the scores either before or after teaching across the two types” (p. 654). The sample size for this study was small (n=8 for each group), however, and the repeated use of the same pretest and posttest for each session (each student took four of the same tests in all), might allow for students to achieve higher scores over time due to practice, familiarity, and memorization—a limitation mentioned in the study (p. 658).

Developmental Students and Attrition

Many studies that explore distance education examined higher levels of coursework, focusing primarily on university juniors, seniors, and graduate students. While it is certainly worthwhile to discover how students at higher levels perform, the results cannot necessarily be applied to all levels of students, particularly developmental-level students and or entry-level freshmen. Many studies found attrition to be a problem in distance learning courses in general, (Carr, 2000; J. Moody, 2004; Kanuka & Jugdev, 2006; Angelino, Williams, & Natvig, 2007). Other studies, however, revealed that attrition might prove to be an even greater problem when student ability or college-readiness was taken into consideration. Dutton et al. (2001) pointed out that some disadvantages of distance learning include a loss of structure provided by regular course meetings, the tendency for online students to not venture to campus to meet with professors for tutoring or help, students not being able to observe the professor deliver lectures and receive immediate responses to questions, and, importantly, less disciplined students risk falling behind in coursework (p. 131-132).

Zavarella and Ignash (2009) investigated the relationship between a student's choice of delivery mode (online or face-to-face) and retention in a developmental mathematics course. Attrition was found to be a problem in online courses, particularly in lower-level courses and one reason for this attrition might be attributed to the students' learning styles and reasons for choosing an online class in the first place (p. 3). For this study, data was collected from three groups of students: 69 students who were enrolled in three sections of a face-to-face algebra course, 67 students enrolled in three sections of a hybrid-version of the course, and 56 students enrolled in three sections of the distance learning version of the course. They then looked at the students' demographic characteristics, entry-level math test scores, a survey developed by the institution to assess why they chose the particular mode they chose, and finally, the students' scores on a learning styles test. The students enrolled in the hybrid or distance learning courses had the highest withdrawal rates, with 42% withdrawing from the hybrid and 39% withdrawing from the distance learning courses. The face-to-face students had a far lower withdrawal rate of only 20%. As a follow-up, the investigators attempted to contact the students to find out the reasons for their withdrawals, and while most of them did not respond, the 11 who did respond stated "the course presented challenges they did not expect" (p. 6).

Ashby et al. (2011) likewise found attrition to be an influential factor when examining distance-learning courses. Their study also centered on a developmental math course and compared student outcomes in terms of grades in a face-to-face course ($n=58$), a distance learning course ($n=63$), and a hybrid course ($n=46$) and also asked the question as to whether the effect of course performance depended on attrition. The students were

evaluated on their final course grades, with a grade of 70% or higher considered “successful” completion of the course. Students were compared first without taking course completion into account and again after accounting for attrition. The results indicated that before taking attrition into account, the number of students who earned passing grades was not significantly different between the three learning environments. Not all of the students took the final exam, however; some students stopped attending class without officially withdrawing, and still others attended the last week of class without taking the final exam and this obviously had a negative impact on the students’ final grades. The students who did not take the final exam were then removed from the study and these results indicated that the students in the online course performed better than either the hybrid or the face-to-face course. In this study, more students in the face-to-face class *completed* the course than the other two groups, with a completion rate of 93% while students in the hybrid course had the lowest completion rate of 70% and the online course had a completion rate of 76%.

A limitation in the study by Ashby et al. (2011), however, involved the use of grades as indicators of course success. Even in an area of math, which might exhibit less subjective grading than in other courses, the outcome measures might still be influenced by instructor bias. Likewise, a student who scores a 70 in a course would be considered “successful” in the course, but a student with similar ability who may have had an “off” exam day might earn a 69 and would then be considered “unsuccessful.” With this study, the regular exams for the online and hybrid courses (7 in all) were not proctored, but the final exam was proctored. These exams were then correlated with the students’ final course exams, which were proctored, so in theory, this should not have been a

problem in this study. Still, the differences between the final exam score average were small: 65% for the face-to-face, 66% for the hybrid, and 67% for the online version of the course. For most of the other exams (the ones that were not proctored), the face-to-face students had lower overall scores, which affected the students' overall averages.

Types of Students and Success Rates

Dutton et al. (2001) examined the differences of delivery method in two college-level introductory computer-programming courses, also using grades as a method of identifying success in the course. The total number of students involved in this study was 272, with 40 removed for not taking the final exam (and hence, not “completing the course”), 14 removed for having “pass/fail” status, and 12 auditors removed. Over all, there were 171 students in the face-to-face course and 141 students in the online course. The results of this study found that the online students actually outperformed the face-to-face students in both exam grades and final course average. The results changed slightly, however, when the students were divided into categories of “lifelong” students (graduate students, non-degree seeking students, post-baccalaureate students, and students working toward a certificate in computer programming) and regular “undergraduate” students. When the analyses were run again for these two individual groups, the results indicated that the undergraduate distance learning students outperformed the face-to-face on the final exam by a much smaller margin and the results for the course average was inconclusive with a high *p*-value. In this study, the distance-learning students were allowed to attend the on-campus lectures, although “they rarely take advantage of this extra benefit” (p. 133).

Verhoeven and Wakeling (2011) also compared the two methods of course delivery. This study examined 373 students who were enrolled in a business quantitative statistics course used final grades to determine the successful or non-successful completion of the course. They further divided the students up into “strong students” (students who had a grade of A or B in the course pre-requisite) and “weak students” (students who made a C or a D on the course pre-requisite). The results of the students were analyzed for course performance based not only on whether they were “weak” or “strong” to begin with, but also by which method of course delivery they had. The results indicated that the success rate was significantly lower for the distance-learning course across all students. The “strong” students, however, still had a significantly higher success rate than the “weak” students for both forms of delivery. While this study also relied on grades and therefore shares the limitation of using grades as a measure of success, the results comparing the “weak” students with the “strong” students resonated with the much of the evidence found in the literature: stronger students seem to be able to weather less-than-ideal delivery methods, including distance learning courses, which require students to have skills such as persistence, time-management, and self-motivation that many do not develop until later in their academic careers.

Finally, Lawrence and Singhanian (2004) conducted a study from spring of 2001 through spring of 2003, using a statistics course designed for undergraduate business students at California State University. In this study, student test results were scored from both the distance learning and face-to-face version of the course and, as with many of the studies, this one also compared tests and course averages to determine the effectiveness of the course and, in addition to this, they looked at the grades of W as well.

There was strong evidence with this study that the traditional students outperformed the distance-learning students, with an average difference of 6.001 points, with a margin of error of 4.276. The traditional face-to-face students also had a higher grade average than the distance-learning students with the average difference being .259, with a margin of error of .220.

Learning Outcome Measures

As discussed previously, much research on distance learning has been in the form of qualitative studies that examine students' perceptions of their online course experiences, examination of instructor perceptions, or examination of student perceptions of the online programs themselves. Studies examining the effectiveness of distance learning courses as compared to their face-to-face counterparts have often focused on higher levels of college-level coursework, including many that focus on graduate studies, which is important with regards to the big picture, but does not adequately address a widely-growing population of freshman-level, often "at-risk" or underprepared student population that routinely enrolls in these gateway courses. Similarly, almost all of the quantitative studies that examined student success in online courses as compared to face-to-face courses used grades as success indicators (usually with the grade of "C" equating to successful completion of the course), which had the limitation of being subject to instructor effect.

Complexities and Importance of Teaching English Composition

Composition is a complex subject to both teach and learn. Writing involves more than grammar and mechanics, but rather also involves a complex intermingling of "critical thinking, informational literacy, problem solving, quantitative reasoning, and

other skills” (Walvoord, 2014, p.1). A student of composition must not only grasp the basics of grammar and mechanics, but also be able to understand complex combinations of how to approach a writing project, determine the audience and adjusting one’s voice for that audience, understanding the purpose of a piece of writing (of which the choices are vast), understanding the connotations and denotations of words, knowing how to incorporate supporting evidence into a work, among many other skills (Lindemann, 2001). Another complex skill that must be learned in a freshman English composition course is that of properly integrating sources and correctly citing them. Moody and Bobic (2011) discuss the difficulty of teaching “digital natives” who are continually bombarded with digital information from a variety of sources how to adopt a view that intellectual property exists and how to properly cite, if only to avoid failing their classes. This task is not simply a matter of holding a brief lecture or a verbal discussion, but will necessarily involve engaging with students on a level in which their worldview is altered (Moody & Bobic, 2011). Mumford (2015) also points out that teaching some learning objectives, such as citation and the correct use of sources, can be challenging due to students’ need to completely rethink how they have been understanding intellectual property. This is another challenge involved in teaching composition.

Contributing to this complexity is the evidence that many students who enter a community college are often not academically prepared for college-level writing in the first place. According to one study, more than half of all community college students participate in a developmental course while they are enrolled in college (Bailey, Jeong, & Cho, 2010). As another more recent example, for all students in the state of Texas who had taken the Texas Success Initiatives Assessment (TSIA) test, which is now required

for all in-coming college students, only 41% of students tested as “College Ready” in the area of Reading and only 25% tested as “College Ready” in the area of Writing (Stout, 2014). Both of these areas are required for English composition mastery. While Stout (2014) points out that these scores are alarmingly low and might be due to a design flaw of the TSIA, she also mentions that the low scores could be a result of poor college preparation in the K-12 educational system. If this is the case, then it stands to reason that the student preparation has a potential to be systemic and even the students who are college ready might still need more preparation.

Composition is also an extremely important foundational course for community college students, but colleges often struggle with how to balance student ability and college success. W. Armstrong (2000) states that one major challenge faced by community colleges is that most have an open-door policy, but this is combined with a focus on increasing student success. The dilemma of community colleges involves finding a way to balance access for as many students as possible while still being able to achieve high success rates for those student (Hadden, 2000). Hughes and Scott-Clayton (2011) write that this quandary not only affects students, but also can result in faculty frustration when students are enrolled in class, yet are not academically skill-ready for the course. If students do not have a solid grasp of how to write well and express their ideas, however, then this can logically lead to difficulties successfully completing later coursework at the institution.

Summary

In this chapter, some history of distance learning was presented and discussed, along with information on the future potential of online learning in the form of MOOCs,

which many institutions have seen as a viable solution to reaching larger numbers of students. Next, this chapter reviewed literature on studies that examined student perceptions of isolation and disconnectedness, which are themes that continually emerge in qualitative studies. Following this, a review of studies done on the effectiveness of distance learning courses was presented and underscoring this was the idea that most studies that examined effectiveness tended to use grades as an outcome variable, which can potentially pose a limitation in such a study. Attrition was the next area to be discussed, especially with regards to developmental students. Much of the literature reveals that weaker students tend to not fare as well in distance learning courses. Next, this discussed the importance of using learning outcome measures in lieu of grades when investigating the difference between distance learning and face-to-face versions of a course. Lastly, the complexities and importance of English composition as a gateway course was discussed, including the difficulty of teaching complex elements of English composition.

CHAPTER 3

METHODOLOGY

This chapter includes a) a description of the population and sample, b) measurement of variables, c) the research design, d) data collection procedures, e) instrumentation, and f) data analysis procedures.

Population and Sample

The population targeted by this study was all students attending a two-year community college and who are enrolled in a freshman-level composition course. The enrollment of the community college involved in this study consisted of approximately 30,000 students distributed across three campuses and of which 40% of those students were Hispanic, 33% were White, 10% were African American, and 5% were Asian. Sixty-eight percent of the students enrolled in this college were under the age of 25 and 32% of the students were over the age of 25. Approximately 57% of the students enrolled at the college were female. Socioeconomically, 38% of students received aid in the form of grants at the time of this study, with 28% of the students receiving a Pell Grant. The sample for this study consisted of students who were randomly selected from a pool of students who had self-selected into face-to-face or distance-learning delivered freshman composition classes.

The participants for this study were selected using a convenience sample of students who self-selected into 16-week freshman-level composition courses in the fall

2014 semester and were taught by instructors who had instructed at least one face-to-face and one distance-learning version of the freshman-composition course each during the fall 2014 semester. The total number of participants in the sample pool consisted of 78 distance learning students and 117 face-to-face students, which made a total of 195 students in the sample pool. From this pool, 50 distance learning students and 50 face-to-face students were selected for this study, resulting in a total sample of $n=100$. The participants in this study consisted of 33% Hispanic students, 39% White, 19% African American, 8% Asian, with 1% (1 student) being “other” (Pacific Islander). The ages of the participants ranged from age 17 to 63, with 33% of the students being 19, 13% age 18, and 11% being age 20. Sixty-three percent of the students were female. Valid data on the students’ Pell Grant status was not available for this sample of students.

In order to limit as much as possible for instructor effect, the instructors of the courses included in this study were required to fit the following parameters: a) the instructors must have taught a minimum of six online and six face-to-face classes prior to the fall 2014 semester, b) the instructors must have each undergone a required six-week course in online instructional course design and best practices offered by the college, c) the instructors must have had their distance learning course reviewed using developed by the institution and received a score of 95% or higher using that rubric, and d) the instructors must have taught both a face-to-face and online version of freshman-level English composition during the fall 2014 semester. There were five instructors who fit these criteria and the student essay samples were selected from the students who had self-selected into these five instructors’ classes.

Measurement of Variables

In order to ensure a consistent understanding of all of the terms mentioned in this study, the following terms are defined:

Method of delivery is the independent variable. The two methods of delivery explored in this study are online and face-to-face delivery of the courses. Online courses for this study refer to courses that are taught 90% or more online. Face-to-face delivery of the courses refers to traditional course delivery of instruction in which students and the instructor met in a brick-and-mortar classroom for greater than 70% of their course instruction.

Age for this has been defined as the age of the student at the time the fall 2014 semester.

Ethnicity for this study was categorized into five variables that reflected the participants in the study. These are White, Hispanic, African American, Asian, with one student being a Pacific Islander, which was coded as “other.”

The scores on an English composition essay rubric developed for this study measured Student Outcomes.

The dependent variables for this study are the scores received on a freshman composition rubric. The independent variables are the methods of delivery: face-to-face and distance learning modes of delivery.

Research Design and Data Collection

This study used a quantitative, quasi-experimental design to compare two groups using equivalency testing. After first receiving both IRB and CPHS approvals from the institution and the University of Houston-Clear Lake, data were gathered through the

application of a rubric and equivalency testing was then used to determine whether the two course delivery methods were equivalent. Descriptive data on the students whose essays were involved in this study were also gathered.

At the institution involved in this study, all freshman composition instructors were required to have their students respond to a faculty-developed “common assignment,” which was used by the institution to determine student-learning outcomes across the college. The college implemented the common assignment requirement in an effort to create consistency to better measure the course outcomes across all disciplines, including freshman composition courses. For the freshman composition common assignment, students were required to write an essay that contained an argumentative thesis statement, consisted of roughly 3-5 double-spaced pages, incorporated at least two sources using MLA citations, must have been assigned in the last two-thirds of the semester, and must have constituted a significant grade in the class. In order to achieve consistency in the essay assignments sampled, the common assignment essays produced in each of the classes were used as the samples gathered for this study.

The common assignment essays were collected using the collection features in the institution’s course-management system. These features allow the college to then identify which essays to evaluate during district-level general education outcomes assessment. Thus, because the institution already collects these essays, it was not necessary to obtain permission from the individual instructors themselves, but rather, the individual essays could be pulled from each course’s common assignment collection site.

Upon completion of the pilot test and satisfactory inter rater reliability was confirmed, the full study was conducted during the months of June 2015 to August 2015.

As mentioned, 100 essay samples were taken from a sample pool of students who were enrolled in five different instructors' freshman-level composition courses. Fifty of these samples were randomly selected from distance-learning versions of the courses, and another 50 were taken from the face-to-face versions of the courses. In order to control for instructor effect, ten student samples were selected at random from each of the five instructors' classes for each mode of delivery, resulting in 20 student essays being selected from each instructor (ten face-to-face and ten distance learning). The essays were then downloaded and saved to a secure flash drive. To help maintain confidentiality, all identifying information for both the student and the instructor was removed and replaced with a number.

Eight different raters were recruited in order to triple-score the essays using the developed rubric. The raters selected to score the essays were full time instructors at the institution, had taught the freshman level composition course for a minimum of six years each, had used a rubric to score essays in the past, and were different instructors than the instructors whose students were selected to participate in this study. The researcher also did not participate in the scoring of the essays.

Each essay was photocopied three times in order to yield 300 hard copies of essay samples (100 essays x 3) and included a paper copy of a rubric with the corresponding number attached to each essay. Packets were then prepared for the eight raters with 37 or 38 samples in each packet. In order to avoid having one rater score all of the distance learning samples or all of the face-to-face samples only, the samples were arranged into alternating distance-learning samples and face-to-face samples, so each rater would receive an even ratio of face-to-face and distance learning samples. The packet also

included detailed instructions on how to use the rubric and was intended to supplement a norming and training session on how to use the rubric. Each of the eight raters was compensated for his or her efforts with a \$25 gift card, which was also included in the packet.

The eight raters participated in a one-hour norming session in which the raters practiced scoring two additional essays (not included in the study) using the rubric. The packets were then distributed to the raters, who had one month to score the essays using the rubrics. All essays and rubrics were collected by the end of the time limit.

Instrumentation

An analytic scoring rubric was designed for this study. An analytic rubric not only includes a total score for the essay being rated, but also includes a separate score for each evaluative criterion (Brookhart, 1999). The first step in creating a rubric is to identify which qualities would comprise a proficient outcome of student work and once this is established, specific descriptions of these outcomes should be written for the most proficient outcome expected (Arter & Chappuis, 2006). After determining and writing the strongest outcome indicators, another set of descriptions should then comprise the lowest level of proficiency in each outcome category. Once accomplished, the criteria for the middle level of performance will become apparent and these can then be written (Brookhart, 1999). While there is always a risk of ambiguity when assessing an essay using a rubric, this ambiguity can be reduced by being as descriptive as possible when writing the rubric indicators (Arter & Chappuis, 2006). It is more effective in terms of inter rater reliability to create additional levels of outcomes only if a meaningful, clear

distinction can be made between the groups. In other words, the levels of proficiency should be kept as simple as possible. A good rubric will also have very clear distinctions between the levels of outcome criteria. If no clear distinctions can be made between the levels, or if the descriptions are difficult to distinguish from one another, then this is a sign that a rubric has too many levels of proficiency outcomes and will result in lowered inter rater reliability (Brookhart, 1999). The above guidelines were followed when constructing the rubric used to evaluate the essays in this study.

The outcomes chosen for this study's rubric were chosen to reflect four specific outcomes for a freshman level English class, with another outcome score being the sum of the total score of the other four outcomes. The first outcome area was designed to measure the students' abilities to impose structure on their essays. A common weakness among freshman-level composition students, as observed over the last nine years by the researcher, was a lack of structure in freshman composition writing. Students might have an idea of what to say, but tend to have difficulty finding a logical format in which to place their thoughts. One of the outcomes of a freshman-composition class is to teach students how to write a thesis statement, how to have clear topic sentences, and how to express their arguments in a logical manner, and thus the outcome area of Structure was an important outcome to measure. The second rubric outcome area was that of content. A common problem for inexperienced writers, again observed by the researcher, was for struggling students to simply repeat the same idea over and over, or linger along the surface of an idea, rather than develop it and expand the idea with any depth. To that end, the area of Content on the rubric was developed to determine this outcome proficiency. The third outcome of Clarity helped to assess the students' abilities to write clearly and to

follow the rules of grammar and mechanics. The fourth rubric outcome was that of Sources and was designed to measure the students' correct incorporation and citation of sources in their essays. Because all students were required to use at least two sources in their district-wide assessment of the common assignment, this was an area that could also be assessed for all students. The researcher had observed that citations often pose a great deal of difficulty for students, with weaker students often not fully citing their sources, not using sources at all when asked to do so, or having different in-text sources from what was listed in the works cited page.

The rubric had three outcome score categories. A "2" would be given to a sample that demonstrated the most proficient outcome in a rubric area and evaluators were instructed, in addition to considering specific qualities listed in the rubric (see Appendix A), to consider the sample as being an A or a B for the outcome area. A "1" would be given to a sample that demonstrated acceptable proficiency, but with some errors or weaknesses, depending on the outcome area. Again, in addition to considering the specific qualities listed on the rubric, evaluators were instructed to consider this outcome as being equal to a grade of "C." Lastly, the score of "0" would be given when the student sample demonstrated the lowest levels of proficiency expected and, in addition to considering the specific qualities listed on the rubric, would have earned a grade of "D" or "F." When the individual outcomes were added together, this yielded the total score on the rubric. The highest score possible on the rubric was an 8 and the lowest possible score was that of a 0.

Pilot Test

In the spring 2015 semester, a pilot test was conducted in order to obtain inter-rater reliability tests on the rubric, as well as to allow for any necessary calibration of the zone of equivalence, to be discussed momentarily. For the pilot test, 20 student essays were randomly selected, taken from a different sample of essays from the fall 2014 semester. These essays were also the same common assignment essays used in this study, but were pulled from a different selection of instructors. Then, three raters were selected to score each essay using the developed rubric (see Appendix A). The raters for the pilot test were chosen based on the qualifications that they a) were full-time English professors at the institution, b) had more than six years of experience reading and evaluating essays at the college, and c) were not the same professors whose students had written the essays. The researcher was also not involved in the rating process. Three different raters then rated each essay and then intra-class correlation (ICC) was used to assess for inter-rater reliability. When one subject is rated by multiple coders, intra-class correlation (ICC) uses correlations to determine the agreement between raters (Hallgren, 2012). There are several ICC variants available and the variant used must be determined based on the type of agreement that the researcher wishes to determine. For this study, it was determined that because a different set of coders was randomly selected from a larger population of coders, a one-way random model was chosen. Absolute agreement between the raters was selected as well. Also, because the ICC was meant to quantify the ratings based on an average of the ratings of multiple coders, average measures was selected for this study. Because the raters in the study were randomly selected from a larger pool of raters and

the scores were meant to generalize to the larger selection of raters, the random effects model was chosen (Hallgren, 2012).

When using intra-class correlation (ICC) to determine reliability, higher ICC values indicate greater inter-rater reliability. An ICC of 1 would indicate complete agreement between the raters and a 0 would indicate only random agreement (Hallgren, 2012). This study used the cut offs for agreement for ICC as being poor if the reliability statistic (Cronbach's Alpha) was less than .40, fair for values that fell between .40 and .59, good for values that fell between .60 and .74, and excellent for values that fell between .75 and 1.0 (Cicchetti, 1994).

For the pilot test conducted, the ICC value using Cronbach's Alpha as the reliability statistic fell into the excellent cut off category for each of the rubric outcomes.

Table 1 presents the ICC values for the pilot test.

Table 1

Intra-Class Correlation (ICC) Values for Pilot Study

Outcome	Cronbach's Alpha	Cut-Off Value
Structure	0.903	>.75 (Excellent)
Content	0.815	>.75 (Excellent)
Clarity	0.857	>.75 (Excellent)
Sources	0.863	>.75 (Excellent)

After the pilot test was conducted, ICC tests were also performed for the main research study on the rubric scores. Again, the Cronbach's Alpha value fell within the excellent cut off for each outcome area. The results of the ICC values for the main study are presented in Table 2.

Table 2
Intraclass Correlation (ICC) Values for Main Study

Outcome	Cronbach's Alpha	Cut-Off Value
Structure	0.905	>.75 (Excellent)
Content	0.873	>.75 (Excellent)
Clarity	0.905	>.75 (Excellent)
Sources	0.959	>.75 (Excellent)
Total Scores	0.961	>.75 (Excellent)

Hypotheses and Data Analysis

The data analysis for this study used the two one-sided t test (TOST) technique to determine equivalency between two groups (Rogers, Howard, & Vessey, 1993).

Equivalency testing differs from traditional hypothesis tests—such as traditional t tests—that seek to determine whether or not two groups are different from one another.

Equivalency tests are more appropriate when the goal of a research study is not to determine whether one treatment is superior over another, but to determine if the two treatments are equally effective (Rogers et al., 1993). Equivalency testing involves specifying a small, nonzero difference between the two treatments that would serve to define a zone of equivalence. If the difference between the two groups is small enough to fall within this defined zone of equivalence, then that difference can then be considered sufficiently small or inconsequential enough for the groups to be considered equivalent.

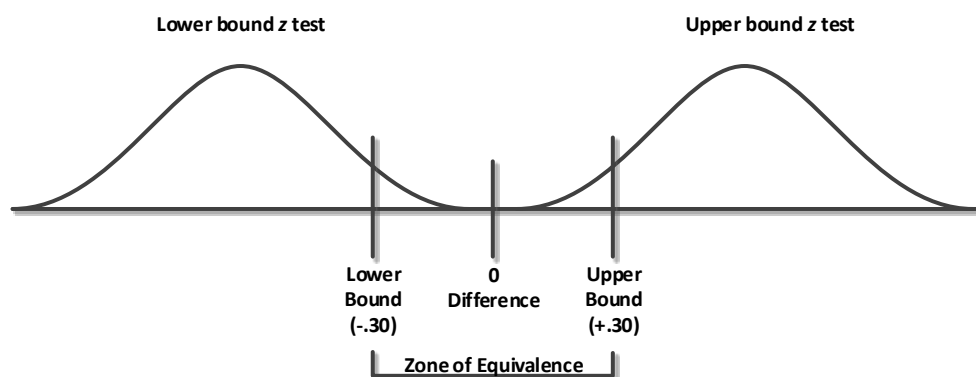
For equivalency testing, the null hypothesis and the alternative hypothesis are somewhat different than with some traditional tests; the null hypothesis would assert that the difference between the two groups will be at least as large as the zone of equivalence defined by the researcher—in other words, the null hypothesis will assert that the groups will be “different.” The alternative hypothesis would then assert that the difference between the groups will be smaller than the zone of equivalence as defined by the

researcher—in other words, the alternative hypothesis will assert that the groups will be “the same” (Rogers et al., 1993).

According to Rogers et al. (1993), the first step in equivalency testing is to determine the aforementioned zone of equivalence. The zone of equivalence must be a specification of the minimum difference between the two groups which would make the two groups different. If either of the two groups’ scores (in this case, the rubric scores for the face-to-face classes and the distance learning classes) fall outside the zone of equivalence, then these groups will be considered “different,” and these two groups would not be considered equivalent. If, on the other hand, both of the scores on the rubric fall *within* the zone of equivalence, then the difference between the two groups is considered too negligible or inconsequential to be considered different and these two groups would be considered equivalent. The first step, however, was to determine the zone of equivalence.

Figure 1

Zone of Equivalence for TOST Tests



Adapted from Rogers et al., (1993).

Rogers et al. (1993) stated that investigators who are experts in their respective fields will need to “define meaningful equivalence intervals [zone of equivalence] relative to the substantive issues at hand” (p. 555). The zone of equivalence might be a difference of less than a specified difference from the mean, as deemed appropriate by the researchers. They cited an example of a study in which the zone of equivalence for scores on a Minnesota Multiphasic Personality Inventory (MMPI) fell within 10% of the mean scores of the subjects studied. For that particular study, the difference 10% or less was considered to be clinically trivial. It is up to the researcher, therefore, to determine what constitutes equivalency for a specific study, based on various considerations unique to that study. For this study, the zone of equivalence was determined to fall within 30% of the mean of the rubric scores. This decision was based on the following logic: If one of the rubric outcome measures—for example, the structure measure on the rubric—is scored by itself as a variable, the highest score that outcome can receive will be a 2, and the lowest score will be a 0. A series of hypothetical possible scores were then considered at random and the means determined for each of the different sets of mock scores. Then, hypothetical “grades” were assigned to each of the different scores based on the actual grades earned on five sets of student essays in the fall 2013 semester of the same freshman level composition course that was the focus of this study. This study did not attempt to assign actual “grades” to any of the essays being examined and the rubric used in this study was not designed to necessarily translate into numerical grades, but this method allowed for a potential emulation of an actual distribution of rubric scores. The method also allowed for the opportunity to hypothesize whether or not the zone of equivalence of less than $\pm 30\%$ of the mean would be appropriate. For each scenario, it

was determined that when using the rubric for one learning objective, an zone of equivalence of less than $\pm 30\%$ was appropriate, especially due to the small scale in which the student essays would be assessed (scores of 0-2 for a single outcome). This logic was reflected in the normal distribution of the fall 2013 student essays. The same zone of equivalence also appeared to be ideal when the learning objective scores were added up and the highest score was thus an 8 for the rubric and the lowest score would, again, be a zero (in this case, the essay would have to score a 0 on each of the four measured learning outcomes).

Another method for performing equivalency tests, as opposed to conducting two one-sided t tests, is to instead use confidence intervals to determine equivalence. Using this method, if the confidence interval is contained within the zone of equivalence, then equivalency can be confirmed (Rogers et al., 1993). This method was also used in this study in order to confirm the TOST results.

Each essay received three separate ratings from three different raters randomly assigned from a pool of 8 possible raters and these rating scores were entered into an excel spreadsheet and imported into SPSS for ICC tests for inter rater reliability. Once the reliability was determined to be in the excellent category, the means of the three ratings were computed for each of the groups being examined and these means were used for the equivalency tests. The rubric scores were then entered into an Excel spreadsheet and then imported into SPSS. Intra-class correlation (ICC) tests were then conducted for reliability of the raters.

Two simultaneous one-sided t tests were then performed for each variable. The first test sought to reject a null hypothesis that asserts that the difference between the two

means is less than or equal to the smaller delta (∂_1). The second test sought to reject a null hypothesis asserting that the difference between the two means is greater than or equal to the larger delta (∂_2). The goal was to demonstrate that an observed difference between the two means is too large to have come from a distribution with a mean of ∂_1 (Test 1) and simultaneously too small to have come from a distribution with a mean of ∂_2 (Test 2) (Rogers et al., 1993).

Additionally, if the equivalency testing revealed that the two samples were not equivalent, then independent t tests would also be performed to determine whether the differences between the means were statistically significant.

Because the descriptive data of the students participating in the study were also gathered for each of the students, student data was then sorted and new equivalency tests were run for each new group. For example, in order to assess whether the courses were equivalent for a specific variable group, the groups of Hispanic students enrolled in the online courses were compared to other Hispanic students enrolled in the face-to-face classes. This replication of equivalency tests was done for each of the variables.

As noted, the student demographic data and essays were protected via the use of a non-identifying number and the data gathered for this study was kept on a flash drive and a password-protected computer, both in a locked and secure environment, for three years. After this time, the data will be destroyed.

Summary

This chapter discussed and described the methodology of this study, including the population and sample examined and the measurement of variables. This chapter also

discussed and explained the two one-sided t test (TOST) method for determining equivalence. This chapter outlined the design for this study, which is as follows:

- Student artifact samples were taken from a “common assignment” essay assigned to all freshman composition students.
- The students had self-selected into a freshman composition course taught by professors who were teaching both face-to-face and distance-learning version of the course during the fall 2014 semester and had similar experience and training teaching both modes of delivery.
- A rubric was designed to assess the essay samples and a pilot test was performed to determine inter rater reliability.
- Intra class correlations were performed on both the pilot test and the main study rubric scores in order to determine inter rater reliability.
- The essays were removed of all identifying information and eight raters scores was used to conduct the TOST tests.
- A zone of equivalence of $\pm 30\%$ of the means for the rubric scores was determined and using this zone of equivalence, TOST tests were performed for each of the variables.
- The confidence interval method of determining equivalence was also performed to determine equivalence.
- Then, t tests were conducted to determine whether the differences in the means were significantly different.

This chapter also discussed the design of the rubric as well as the rationale for the student outcomes measured by the rubric. This chapter also discussed the process for training the faculty raters and the time-frame of the data-gathering phase of this study.

CHAPTER 4

RESULTS

Presented in this chapter are the results of the equivalency tests performed on the rubric scores of the freshman level composition course essay samples for classes taught via distance learning delivery and face-to-face delivery. This chapter also presents the results of the equivalency tests of the essay rubric scores as grouped by ethnicity and gender. The results of the t tests conducted following the equivalency tests are also presented in this chapter. A description of the sample as a whole was presented in Chapter Three.

Research Question and Hypothesis One

Do freshman-level English composition courses taught online have equivalent outcomes as English composition courses taught via face-to-face delivery?

H_0 : The outcomes for the freshman-level English composition courses will not be equivalent to English composition courses taught via face-to-face delivery.

H_1 : The outcomes for the freshman-level English composition courses will be equivalent to the English composition courses taught via face-to-face delivery.

In order to determine whether the course outcomes are equivalent, equivalency tests were performed using the rubric scores for the face-to-face ($n=50$) and distance learning ($n=50$) essay samples. The means for each rubric outcome score for all 100 of

the essays sampled were calculated and $\pm 30\%$ of the means was used for the zone of equivalence for each test.

The equivalency tests for the over all scores of the face-to-face students and the distance-learning students revealed that the course outcomes were not equivalent. The difference between the face-to-face scores ($M=5.09$, $SD=1.87$) and the distance learning scores ($M=3.23$, $SD=1.95$) was 1.87, which fell outside the zone of equivalence of $\pm 30\%$ of the mean, which would be a difference between the means of ± 1.20 . These results thus required the acceptance of the null hypothesis, which states that the two groups were not equivalent.

Equivalency tests were also performed on the four individual outcomes that comprised the rubric, which were Structure, Content, Clarity, and Sources. The equivalency tests performed on the individual outcome areas revealed that the difference between the outcome scores of the face-to-face and distance learning courses fell outside the zone of equivalency for all four of the rubric outcomes. For the rubric outcome of Sources, for example, the scores for the face-to-face students ($M=1.26$, $SD=.63$) were not equivalent to the distance learning students ($M=.69$, $SD=.67$) because the difference between the means was .57, which was not contained within the $\pm .30$ zone of equivalency of the means for the simultaneous independent t tests. It was thus necessary to accept the null hypothesis, which asserts that the groups are not the same. This was the case with every test done for the four rubric outcome areas. The results of these equivalency tests are presented in Table 3.

Table 3

Results of Equivalency Tests for Distance Learning and F2F Rubric Scores

Rubric Outcome	F2F			DL			ZE	Diff	<i>p</i>	90% CI	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>				LCL	UCL
Structure	50	1.31	.59	50	.85	.66	±.30	.46	.897	.252	.668
Content	50	1.25	.53	50	.83	.57	±.30	.41	.846	.230	.597
Clarity	50	1.27	.57	50	.85	.47	±.30	.43	.884	.252	.601
Sources	50	1.26	.63	50	.69	.67	±.30	.57	.979	.352	.781

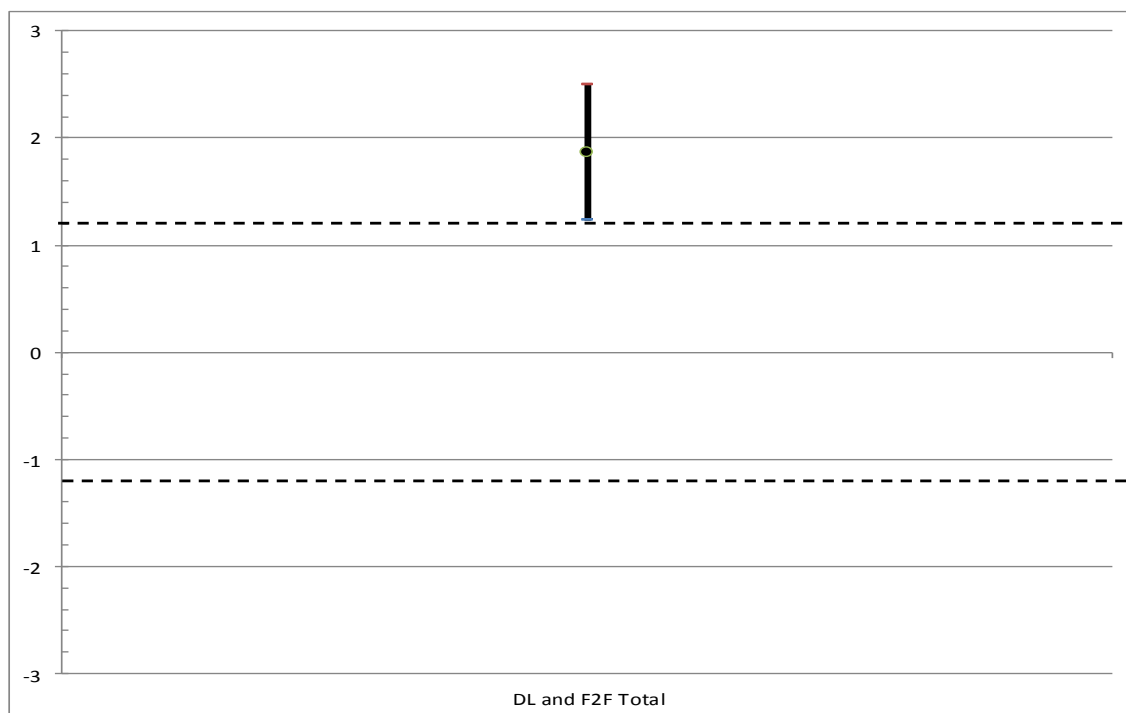
Note: The highest *p* value of the two one-sided tests has been reported.

ZE=Zone of Equivalence

The confidence interval method was also used to test for equivalency for the total scores on the rubric. Equivalency can be concluded if the confidence interval is contained within the zone of equivalence (Rogers et al., 1993). For the face-to-face and distance learning groups, however, the confidence interval CI [1.232, 2.501] was not entirely contained within the zone of equivalence of a difference of $\pm 30\%$ of the mean (± 1.20), and thus, the two groups could not be considered equivalent. The confidence interval method is depicted in Figure 2.

Figure 2

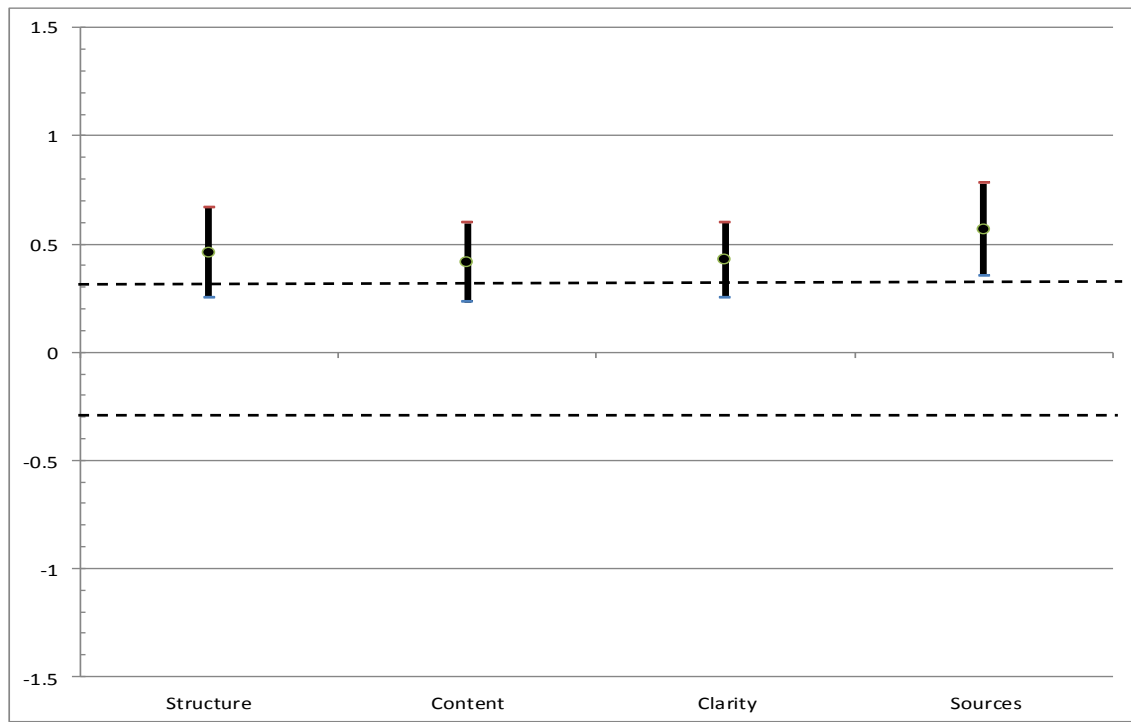
Confidence Intervals for Face-to-Face and Distance Learning Total Scores



Likewise, the confidence intervals for the four individual rubric outcome areas of Structure, Content, Clarity, and Sources were not contained within the zone of equivalency. Using the area of Sources once again as an example, 90% CI [.352, .781] fell outside the zone of equivalency of $\pm 30\%$ of the mean ($\pm .30$). Therefore, equivalency cannot be concluded and it is necessary to accept the null hypothesis that the groups were not equivalent. Figure 3 illustrates the results of the confidence intervals for the four main rubric outcomes of Structure, Content, Clarity, and Sources.

Figure 3

Confidence Intervals for Face-to-Face and Distance Learning by Outcome



Because each of the face-to-face means was higher, t tests were then conducted to see if this difference was statistically significant. For Structure, there was a significant difference in the scores of the distance learning samples ($M=.85$, $SD=.66$) and the face-to-face samples ($M=1.31$, $SD=.59$); $t(98)=-3.67$, $p<.001$. In terms of Content, there was a significant difference in the scores of the distance learning samples ($M=.83$, $SD=.57$) and the face-to-face samples ($M=1.25$, $SD=.53$), $t(98)=-3.73$, $p<.001$. These results suggest that the face-to-face students had a better performance outcome than the distance-learning students with regard to the content of their essays. Similarly, the t tests conducted for the area of Clarity also revealed a significant difference between the distance learning samples ($M=.85$, $SD=.47$) and the face-to-face samples ($M=1.27$, $SD=.57$); $t(98)=-4.06$, $p<.0001$. With regards to Sources, the t tests revealed significant

differences between the distance learning group ($M=.69$, $SD=.67$) and the face-to-face group ($M=1.26$, $SD=.63$); $t(98)=-4.39$, $p<.0001$. Finally, the tests conducted on the entire rubric scores also revealed statistically significant differences between the distance learning samples ($M=3.23$; $SD=1.95$) and the face-to-face samples ($M=5.10$, $SD=1.87$), $t(98)=-4.89$, $p<.0001$.

Research Question and Hypothesis Two

Do freshman-level English composition courses taught online have equivalent outcomes as English composition courses taught via face-to-face delivery for the ethnicities of White, Hispanic, and Black students?

H_0 : The outcomes for the freshman-level English composition courses will not be equivalent to the English composition courses taught via face-to-face delivery for White, Hispanic, and Black students.

H_1 : The outcomes for the freshman-level English composition courses will be equivalent to the English composition courses taught via face-to-face delivery for White, Hispanic, and Black students.

The student samples were sorted by both mode of delivery and ethnicity and equivalency tests were then conducted using a zone of equivalency of 30% of the mean for the total number of face-to-face and distance learning students. No tests were done for the Asian students in the study due to a very small sample size for the distance-learning students ($n=2$) or for the single Pacific Islander student ($n=1$).

The equivalency tests for scores of the face-to-face students and the distance-learning students when separated by ethnicity revealed that the course outcomes were not equivalent for the White, Hispanic, and Black students. For each ethnicity, the difference

between the means fell outside the zone of equivalence of $\pm 30\%$ of the mean and thus it was necessary to accept the null hypothesis that the two groups were not equivalent. For White students, for example, the outcome scores for the face-to-face students ($M=5.67$, $SD=1.74$) were not equivalent to those of the distance-learning students ($M=3.56$, $SD=1.78$) because the difference between the means was 2.11, which was not contained within the zone of equivalency of $\pm 30\%$ of the mean (a difference of ± 1.20). In this case, it was necessary to accept the null hypothesis that the courses were not equivalent. Likewise, the outcome scores for Hispanic face-to-face students ($M=4.83$, $SD=2.00$) and distance-learning students ($M=2.89$, $SD=2.10$) had a difference of 1.94, which was not contained within the zone of equivalence (± 1.20). It was thus necessary to accept the null hypothesis that these two outcomes were not equivalent. Table 4 presents the results of the equivalency tests performed for the ethnicities of White, Hispanic, and Black students.

Table 4
Results Equivalency Tests for DL and F2F by Ethnicity

Ethnicity	F2F			DL			ZE	Diff	<i>p</i>	90% CI	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>				LCL	UCL
White	18	5.66	1.74	21	3.56	1.78	± 1.20	2.11	.942	1.158	3.064
Hispanic	21	4.83	2.00	12	2.89	2.10	± 1.20	1.94	.838	.690	3.183
Black	5	5.53	1.91	14	3.12	2.02	± 1.20	2.41	.871	.610	4.219

Note: The highest *p* value of the two one-sided tests has been reported.

ZE=Zone of Equivalence

Equivalency tests were also performed for the individual rubric outcomes for the face-to-face and distance learning White, Hispanic and Black students for each rubric outcome area of Structure, Content, Clarity, and Sources. The equivalency tests performed revealed the differences between the means of the face-to-face and distance

learning outcome scores for White, Hispanic, and Black students fell outside the zone of equivalence of 30% of the means for each of the rubric outcomes. For example, for the White students, the outcomes scores for Sources for White face-to-face students ($M=1.56$, $SD=.63$) was not equivalent to the distance learning students ($M=.76$, $SD=.63$) because the difference between the means was .79, which fell outside the zone of equivalency of $\pm 30\%$ of the mean ($\pm .30$) and it was thus necessary to accept the null hypothesis that the two groups were not equivalent. This was the case for every test conducted for the face-to-face and distance learning students when grouped by ethnicity. The results of the equivalency tests for each of the rubric outcomes by ethnicity is presented in Table 5.

Table 5

Results Equivalency Tests for DL and F2F by Ethnicity by Outcome

Rubric Outcome	F2F			DL			ZE	Diff	<i>p</i>	90% CI	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>				LCL	UCL
White Structure	18	1.26	.60	21	.83	.60	$\pm .30$.43	.754	.108	.760
White Content	18	1.32	.52	21	1.00	.57	$\pm .30$.32	.533	.019	.610
White Clarity	18	1.54	.59	21	.97	.46	$\pm .30$.57	.942	.287	.851
White Sources	18	1.56	.63	21	.76	.63	$\pm .30$.79	.991	.455	1.133
Hisp Structure	21	1.41	.60	12	.89	.67	$\pm .30$.52	.836	.141	.906
Hisp Content	21	1.22	.59	12	.64	.48	$\pm .30$.58	.916	.244	.923
Hisp Clarity	21	1.18	.53	12	.67	.49	$\pm .30$.51	.861	.189	.827
Hisp Sources	21	1.02	.59	12	.69	.73	$\pm .30$.32	.536	-.074	.717
Black Structure	5	1.47	.61	14	.88	.74	$\pm .30$.59	.776	-.055	1.226
Black Content	5	1.40	.55	14	.71	.54	$\pm .30$.69	.906	.197	1.175
Black Clarity	5	1.33	.47	14	.86	.45	$\pm .30$.48	.767	.065	.887
Black Sources	5	1.33	.67	14	.67	.73	$\pm .30$.67	.831	.020	1.314

Note: The highest *p* value of the two one-sided tests has been reported.

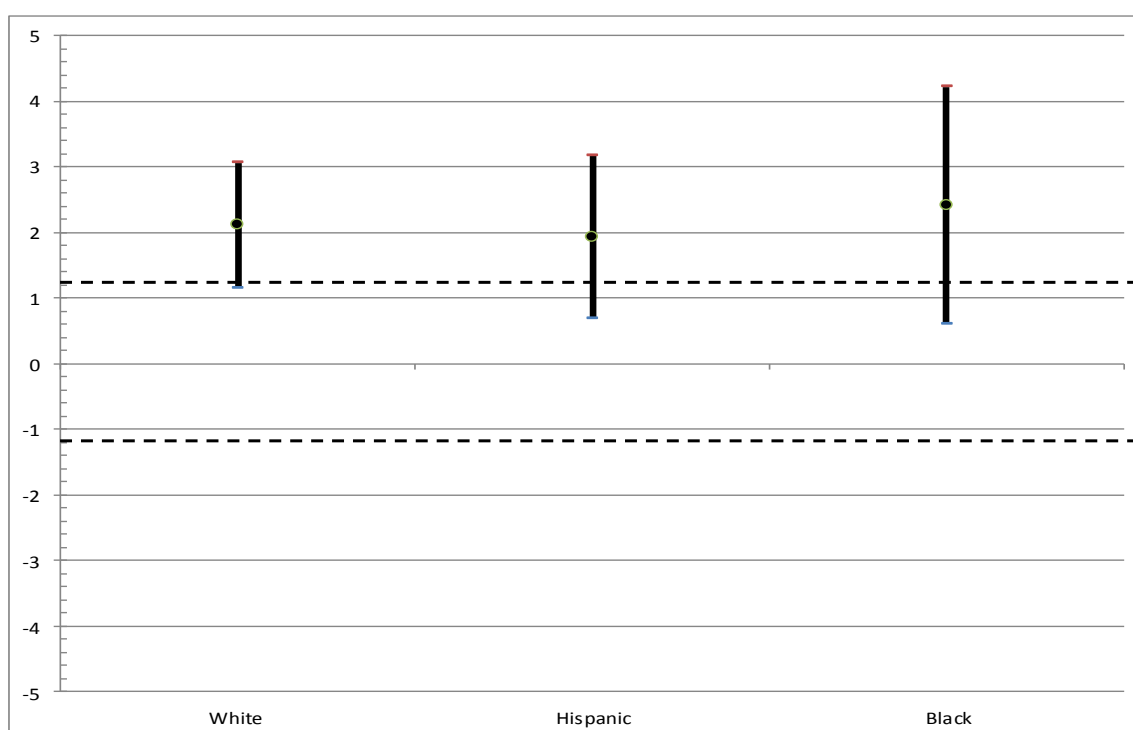
ZE=Zone of Equivalence

The confidence interval method was also used to test for equivalency and this method also confirmed that the outcome scores for the White, Hispanic, and Black

student groups were not equivalent. For the total scores on the rubric, the confidence intervals were not contained within the zone of equivalence and thus, the face-to-face and distance-learning groups could not be considered equivalent. Figure 4 illustrates the confidence intervals and the zone of equivalence of $\pm 30\%$ of the mean (± 1.20) for the total scores on the rubric for all students.

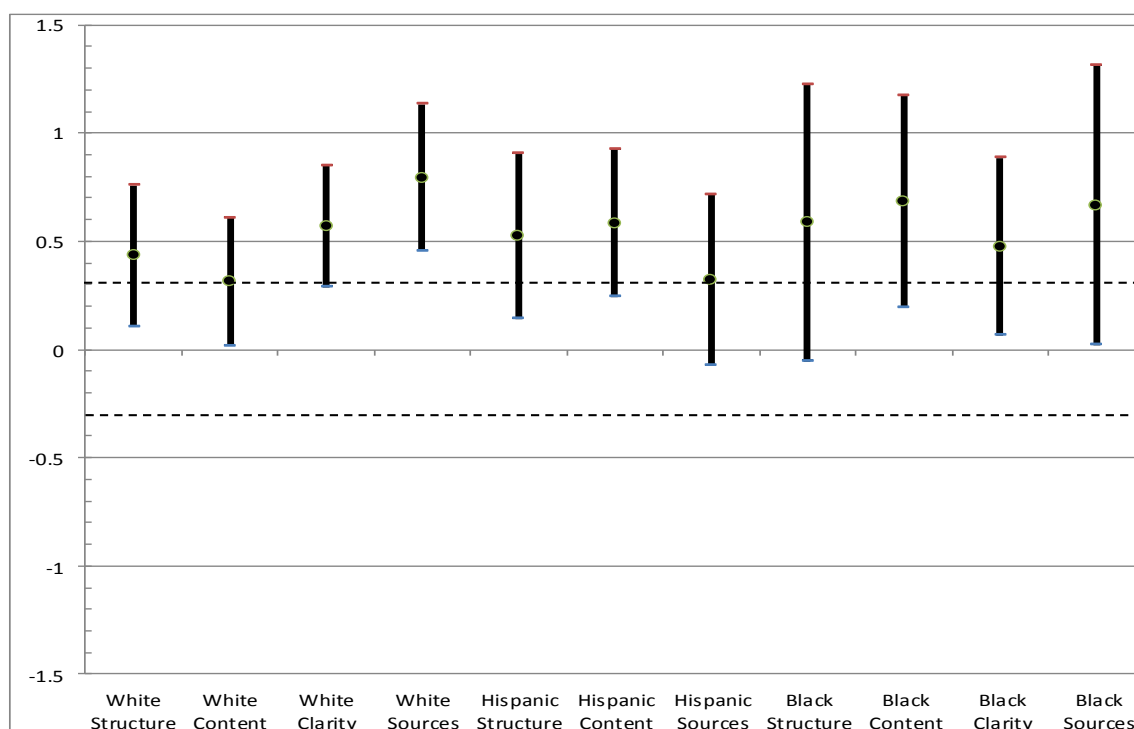
Figure 4

Confidence Intervals for White, Hispanic, and Black Total Scores



Likewise, when the confidence interval method was used for the individual outcome scores on the rubrics by ethnicity, the confidence intervals were not contained within the zone of equivalence of $\pm 30\%$ of the mean and thus, it was necessary to accept the null hypothesis that the groups were not equivalent. Figure 5 illustrates the confidence intervals for each of the rubric outcome areas for all ethnicities tested.

Figure 5

Confidence Intervals for Outcome Scores by Ethnicity

Because the means of the face-to-face student groups were higher than the distance-learning groups, t tests were conducted to determine these differences were statistically significant for the White, Hispanic, and Black student samples.

For the White students, the t tests conducted for the outcome areas of Structure, Clarity, Sources, and Total Scores showed a significant difference between the face-to-face and distance learning outcomes. The areas of Content, the face-to-face to face samples ($M=1.32$, $SD=.52$) and the distance learning samples ($M=1.00$, $SD=.57$) had a p value of .080, so the difference between the groups was not significant for the outcome of Content. The results of these tests are presented in Table 6.

Table 6
Results of Independent t Tests Comparing F2F and DL of White students

Rubric Outcome	<i>n</i>	Mean	SD	df	<i>t</i>	Difference	<i>p</i> value (two tailed)
F2F Structure	18	1.26	.60	37	-2.25	-.434	.031*
DL Structure	21	.83	.60				
F2F Content	18	1.32	.52	37	-1.8	-.315	.080
DL Content	21	1.00	.57				
F2F Clarity	18	1.54	.59	37	-3.4	-.569	.002*
DL Clarity	21	.97	.46				
F2F Sources	18	1.56	.63	37	-3.95	-.794	<.001*
DL Sources	21	.76	.63				
F2F Total	18	5.67	1.74	37	-3.74	-2.111	.001*
DL Total	21	3.56	1.78				

* $p < .05$

For the Hispanic student samples, the *t* tests conducted revealed a significant difference between Structure, Content, Clarity, and Total Scores. The Sources rubric outcome did not reveal a significant difference because the face-to-face to face samples ($M=1.02$, $SD=.59$) and the distance learning samples ($M=.69$, $SD=.73$) had a *p* value of .178, and therefore, the difference between the groups was not significant. The results of the Hispanic student outcomes are presented in Table 7.

Table 7
Results of Independent t Tests Comparing Outcomes of Hispanic students

Rubric Outcome	<i>n</i>	Mean	SD	df	<i>t</i>	Difference	<i>p</i> value (two tailed)
F2F Structure	21	1.41	.60	31	-2.321	-.52	.027*
DL Structure	12	.89	.67				
F2F Content	21	1.22	.59	31	-2.910	-.58	.007*
DL Content	12	.64	.481				
F2F Clarity	21	1.18	.53	31	-2.703	-.51	.011*
DL Clarity	12	.67	.49				
F2F Sources	21	1.02	.59	31	-1.378	-.32	.178
DL Sources	12	.69	.73				
F2F Total	21	4.83	2.00	31	-2.634	-1.94	.013*
DL Total	12	2.89	2.10				

* $p < .05$

For the Black student outcomes sampled, the t tests conducted revealed significant differences for the areas of Content ($p=.026$) and Total Scores ($p=.033$). The other three rubric area outcomes of Structure, Clarity, and Sources were not significant. The complete results for the t tests for the outcomes for Black students are presented in Table 8.

Table 8
Results of Independent t Tests Comparing Outcomes of Black Students

Rubric Outcome	n	Mean	SD	df	t	Difference	p value (two tailed)
F2F Structure	5	1.47	.61	17	-1.591	-.586	.130
DL Structure	14	.88	.74				
F2F Content	5	1.40	.548	17	-2.440	-.686	.026*
DL Content	14	.71	.537				
F2F Clarity	5	1.33	.47	17	-2.017	-.476	.060
DL Clarity	14	.86	.45				
F2F Sources	5	1.33	.67	17	-1.792	-.667	.091
DL Sources	14	.67	.73				
F2F Total	5	5.53	1.91	17	-2.328	-2.414	.033*
DL Total	14	3.12	2.02				

* $p<.05$

Research Question and Hypothesis Three

Do freshman-level English composition courses taught online have equivalent outcomes as English composition courses taught via face-to-face delivery for male and female students?

H_0 : The outcomes for the freshman-level English composition courses will not be equivalent to the English composition courses taught via face-to-face delivery for male and female students.

H_1 : The outcomes for the freshman-level English composition courses will be equivalent to the English composition courses taught via face-to-face delivery for male and female students.

The student samples were also sorted by sex and equivalency tests were then conducted using a zone of equivalency of 30% of the mean for the total number of face-to-face and distance learning students.

The equivalency tests performed for the female and male students revealed that the two groups were not equivalent. For the total scores on the rubric, the female face-to-face students ($M=5.23$, $SD=1.93$) and the female distance-learning students ($M=3.34$, $SD=1.99$) had a difference of 1.89, which fell outside the zone of equivalence of 30% of the mean (± 1.20) and it was thus necessary to accept the null hypothesis that the two groups were not equivalent. Likewise, for the total scores of the male students, the male face-to-face students ($M=4.88$, $SD=1.81$) and the distance-learning students ($M=3.00$, $SD=1.89$) had a difference of 1.88, which fell outside the zone of equivalency of 30% of the mean, or ± 1.20 . It was thus necessary to accept the null hypothesis that the two groups were not equivalent.

When the total scores on the rubric were broken down by individual outcomes of Structure, Content, Clarity, and Sources for the female and male students, the results also revealed that the face-to-face and distance learning groups were not equivalent. For the female students, the greatest difference was in the outcome area of Sources, which had a difference of .65. This difference, as with all of the differences, fell outside the zone of equivalence of $\pm 30\%$ of the mean. The results of the female equivalency tests for each rubric outcome area are presented in Table 9.

Table 9
Results of Equivalency Tests for Female Rubric Scores

Rubric Outcome	F2F			DL			ZE	Diff	<i>p</i>	90% CI	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>				LCL	UCL
Structure	30	1.31	.61	33	.94	.64	±.30	.37	.674	.108	.635
Content	30	1.21	.56	33	.83	.55	±.30	.38	.721	.148	.618
Clarity	30	1.37	.54	33	.88	.53	±.30	.49	.916	.263	.713
Sources	30	1.34	.60	33	.70	.70	±.30	.65	.980	.370	.925

Note: The highest *p* value of the two one-sided tests has been reported.

ZE=Zone of Equivalence

The equivalency tests performed for the male students also did not reveal equivalence with any of the rubric outcome areas. The largest difference for the male students was the area of Structure. This outcome had a difference between the face-to-face and distance-learning means of .63, which fell outside the zone of equivalency of ±30% of the mean and it was thus necessary to accept the null hypothesis that the groups were not equivalent. This was the case for all of the male rubric outcome areas. The results of the equivalence tests for the male students are presented in Table 10.

Table 10
Results of Equivalency Tests for Male Rubric Scores

Rubric Outcome	F2F			DL			ZE	Diff	<i>p</i>	90% CI	
	<i>n</i>	<i>M</i>	<i>SD</i>	<i>n</i>	<i>M</i>	<i>SD</i>				LCL	UCL
Structure	20	1.32	.57	17	.69	.70	±.30	.63	.939	.278	.983
Content	20	1.30	.49	17	.84	.63	±.30	.46	.800	.146	.768
Clarity	20	1.13	.61	17	.78	.35	±.30	.35	.615	.067	.631
Sources	20	1.13	.65	17	.69	.61	±.30	.45	.757	.095	.799

Note: The highest *p* value of the two one-sided tests has been reported.

ZE=Zone of Equivalence

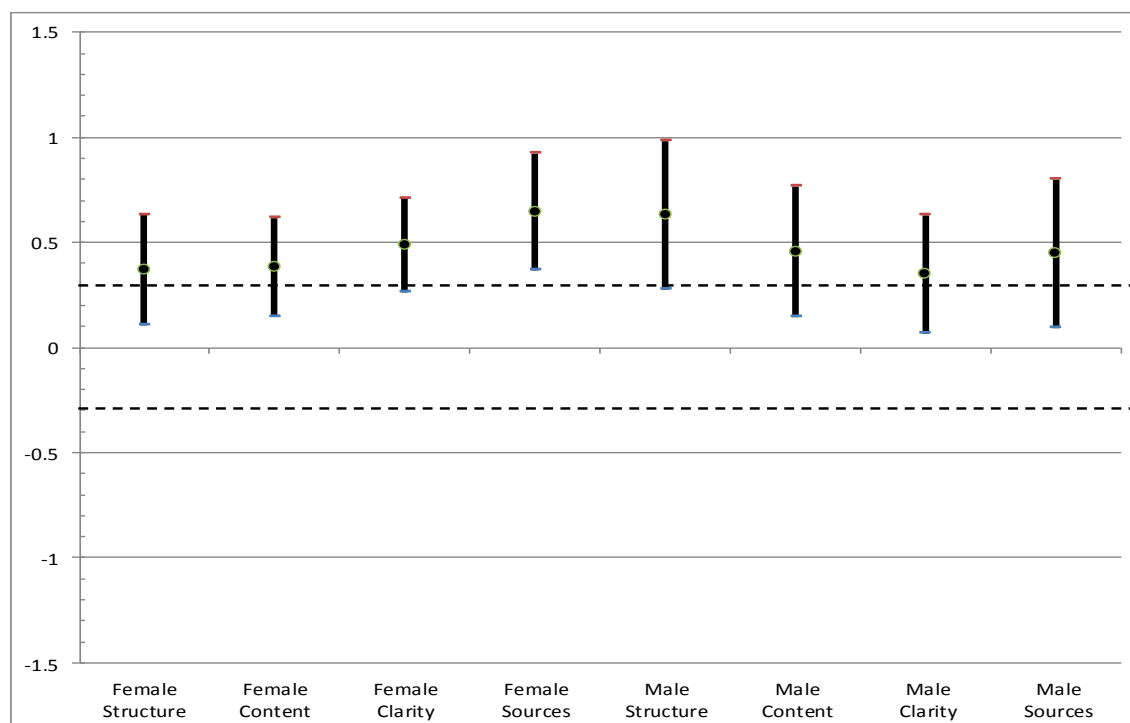
Using the confidence-interval method, the confidence intervals were not contained within the zone of equivalency for any of the rubric outcome areas and thus, it

was necessary to accept the null hypothesis that the groups were not equivalent. For the Total Scores for the female students, the 90% CI [1.062, 2.718] was not contained within the zone of equivalence of $\pm 30\%$ of the mean (± 1.20) and it was thus necessary to conclude that the groups were not equivalent. Likewise, for the Total Scores for the male students, the 90% CI [.855, 2.912] was not contained within the zone of equivalence either, and thus, it was necessary to accept the null hypothesis that the groups were not equivalent.

The same results were revealed when the scores for the smaller rubric outcomes were examined. Figure 6 depicts the confidence intervals for the female and male scores for each of the rubric outcomes.

Figure 6

Confidence Intervals of Rubric Outcomes by Sex



Because each of the face-to-face means was higher, t tests were then conducted to see if the differences were statistically significant. The results of the t tests comparing the outcome results for the female students are presented in Table 11. For the female students, all of the rubric outcome areas revealed a significant difference between the face-to-face and distance learning students. The greatest difference revealed by the t tests was in the rubric area of Sources, in which the mean for the face-to-face mode ($M=1.34$, $SD=.60$) had a significant difference than that of distance learning mode ($M=.70$, $SD=.70$); $t(61)=-3.93$, ($p<.0001$).

Table 11

Results of Independent t Tests Comparing Outcomes of Female Students

Rubric Outcome	n	Mean	SD	df	t	Difference	p value (two tailed)
F2F Struct.	30	1.31	.61	61	-2.36	-.372	.022*
DL Struct.	33	.94	.64				
F2F Content	30	1.21	.564	61	-2.72	-.383	.009*
DL Content	33	.83	.554				
F2F Clarity	30	1.37	.54	61	-3.62	-.488	.001*
DL Clarity	33	.88	.53				
F2F Sources	30	1.34	.60	61	-3.93	-.647	<.0001*
DL Sources	33	.70	.70				
F2F Total	30	5.23	1.93	61	-3.81	-1.89	<.001*
DL Total	33	3.34	1.99				

* $p<.05$

Similarly the male student scores also revealed a significant difference for each of the rubric areas. The results for the independent t tests conducted for the male students are presented in Table 12. As with the female scores, the male scores for all rubric outcome areas revealed a significant difference between the face-to-face and distance-learning delivery modes. One of the greatest differences for the rubric outcome areas for males was that of Structure, in which the face-to-face students ($M=1.37$, $SD=.57$) was

significantly different than the distance-learning students ($M=.69$, $SD=.70$) with a difference of $-.63$ ($p=.005$). Likewise, the Total Scores for the males also showed a significant difference of -1.88 ($p=.004$), with the face-to-face students having higher rubric outcome scores than the distance-learning students. This significant difference was the result for tests done on all of the rubric outcomes.

Table 12

Results of Independent t Tests Comparing Outcomes of Male Students

Rubric Outcome	n	Mean	SD	df	<i>t</i>	Difference	<i>p</i> value (two tailed)
F2F Struct.	20	1.37	.57	35	-3.023	-.63	.005*
DL Structure	17	.69	.70				
F2F Content	20	1.30	.95	35	-2.482	-.457	.018*
DL Content	17	.84	.63				
F2F Clarity	20	1.13	.61	35	-2.091	-.349	.044*
DL Clarity	17	.78	.35				
F2F Sources	20	1.13	.65	35	-2.145	-.447	.039*
DL Sources	17	.69	.61				
F2F Total	20	4.88	1.81	35	-3.094	-1.883	.004*
DL Total	17	3.00	1.89				

* $p<.05$

Summary

This chapter presented the findings of the data analysis conducted when examining the research questions and hypotheses. The results of the equivalency tests were presented and these indicated that the distance learning courses were not equivalent to the face-to-face courses for all three of the research questions. In each case, the difference between the means of the face-to-face outcome scores and the distance learning outcome scores fell outside of the zone of equivalence of $\pm 30\%$ of the mean and thus it was necessary to accept the null hypothesis that the groups were not equivalent. The confidence-interval method was also used to test for equivalency and when these

tests were conducted, in every case, the confidence intervals were not contained within the equivalency intervals and it was thus necessary to accept the null hypothesis that the face-to-face and distance-learning groups were not equivalent. When follow-up t tests were conducted to determine whether the observable differences between the means for the face-to-face and distance learning groups were significant, most of the results revealed significant differences, with the face-to-face students having a higher means than the distance learning students.

CHAPTER 5

FINDINGS, IMPLICATIONS, AND CONCLUSIONS

Presented in this chapter is a discussion of the findings of the analyses presented in Chapter Four. This chapter offers a discussion of the outcomes of freshman composition classes taught via online and face-to-face delivery and examines the differences between the two delivery modes, as revealed in the equivalency tests and independent t tests. This chapter also discusses the differences in the outcomes when the groups were divided into various ethnicities and also by sex. Finally, this chapter discusses the implications of the findings, the limitations of the study, along with areas of future research, and the conclusions reached by the researcher.

Summary of Findings

The results of the equivalency tests conducted revealed that freshman level composition courses taught online do not have equivalent outcomes to courses taught via face-to-face delivery. This was the case for all of the tests conducted, including tests conducted on the various sub-categories based on ethnicity and sex. In addition, the t tests that were conducted after the equivalency tests revealed that there were significant differences between the means of the face-to-face and distance learning students over all, with significant differences in many of the sub categories based on ethnicity and sex.

Face-to-Face and Distance Learning Equivalency for Freshman Composition

The face-to-face and distance learning courses did not have equivalent outcomes and the post hoc *t* tests revealed that all outcome areas were significantly different for the two groups with the face-to-face means consistently higher. These results differ from many other studies that concluded no statistical difference between face-to-face and distance learning courses (Chernish et al., 2005; Frederickson et al., 2005; Herman & Banister, 2007; Ashby et al., 2011). These results also differ from Burkhardt et al. (2008), who found, through comparative analysis, that the distance learning students in an upper-division library science course fared “as well, and perhaps a bit better” than the face-to-face students (p. 387). The results are, however, in line with Verhoeven and Wakeling (2011), who found that online students has a significantly lower success rate in an undergraduate business course. Lawrence and Singhania (2004) also had similar results with a study comparing face-to-face and distance learning business courses.

The differences between the means of the face-to-face and distance learning classes was substantial for all outcome areas, but the rubric area of Sources had the lowest scores for the distance-learning students, which is in line with Moody and Bobic (2011) and Mumford (2015), who point out that this particular outcome area can be challenging to teach due to many factors, including a fundamental difference in understanding of how students and instructors understand the concept of intellectual property, let alone the mechanics of how to cite sources in general. Citing sources might be one of the more difficult concepts for students to master and this area did indeed seem to cause the distance learning group of students the most difficulty. This provides a bit of additional evidence that the face-to-face students grasped the course material better than

the distance-learning students, especially when learning a new or challenging concept, such as incorporating sources or using citations.

There are many possible reasons why the distance learning freshman composition students did not perform as well as their face-to-face counterparts. Students who do not feel connected to an institution, may not spend time on a campus or identifying with an institution, or are not able to visualize the long-term benefits of obtaining a degree are more likely to have poorer grades or abandon their academic efforts all together (Tinto, 1997; Schuetz, 2005; Bers and Schuetz, 2014). By definition, distance-learning students do not physically come to the campus very often, if ever, and thus they may have a more difficult time connecting to the campus, instructors, and peers, and thus have more difficulty identifying with their institution. Additionally, because there is often no instructor physically present to ensure students are paying attention to the delivery of material, distance-learning students must be able to be more self-motivated than the face-to-face students.

Along a similar note, many studies have shown that feelings of isolation and a lack of connection to the other participants in the course, including a lack of connection to the instructor and other classmates have shown up again and again in the literature (Chernish et al. 2005; Erichsen & Bollinger, 2011; Shaw & Polovina, 1999; Cross, 1998). This lack of connection might account for the disparity in the outcome scores between the face-to-face and distance-learning courses. Akyol and Garrison (2011) found that there is a strong connection between collaboration in an online classroom and cognitive processes, but this quality is difficult to attain in a distant learning course, especially a

course in which the main skill learned, writing, has the potential to be a very solitary endeavor in the first place.

Another difficulty for distance learning students has to do with the added freedom of being able to log into the course and do the coursework completely autonomously. In addition to being able to plan sufficient time to read critically, write essays, and revise assignments, which all freshman composition students must do, distance learning students must also have the self-discipline necessary to log into the course regularly without being “forced” to do so at a specific time. They must also view instructional material on their own, read and understand assignment instructions, as opposed to having the material delivered verbally, which may also require reliance on stronger reading skills. Distance-learning students in general must be more self-motivated with regards to taking ownership of their learning and completion of their courses.

Said another way, distance-learning students require a higher level of grit than their face-to-face counterparts, which Duckworth, Peterson, Matthews, and Kelly (2007) define as “perseverance and passion for long-term goals” and that it “entails working strenuously toward challenges, maintaining effort and interest....despite failure, adversity, and plateaus in progress: (p. 1087-1088). Similarly, Smith and Zhang (2009) examined the trait of “academic ethic” in college freshman and described this trait as being a learned behavior that included placing their studies above other activities, studying on a daily or near-daily basis, and otherwise being able to place academics above other competing priorities in their lives. They found that students who possessed academic ethic had higher grade point averages than those who did not. In order to be successful, it makes logical sense that distance learning students must have higher levels

of “academic ethic” or “grit” than the face-to-face population, which might explain why studies that explore distance-learning for upper college course levels and graduate courses have better success rates than developmental and gateway courses such as freshman composition (Chernish, et al., 2005; Herman & Banister, 2007).

Several studies have pointed out the challenges of offering distance learning courses to developmental and lower-level students (Dutton et al., 2001; Zavarella & Ignash, 2009; Ashby et al., 2011). Distance learning delivery can present additional challenges to students because, in addition to learning the material and adapting to college-level courses in general, distance learning courses by nature contain a certain amount of loss in structure, they usually have no regular meeting times, and, by nature, there is a lack of instant personal interaction with instructors who might otherwise be able to answer students’ questions immediately (Dutton et al., 2001). Additionally, for instructors, there is no way to easily assess whether or not students are paying attention or absorbing the material and much more responsibility is placed upon the student. In a face-to-face class, when a student is not paying attention or otherwise not engaged, for example, the instructor can see this and immediately intervene and take measures to redirect the student. This is rarely the case in a distance-learning course.

While it is possible for distance-learning instructors to gather basic data such as how long students are logging into the course site, how long they are remaining in different areas of the course, and a multitude of other pieces of data, there are still limitations to how well an instructor can determine whether or not students are engaging with the material and take immediate steps to correct those who are not. By the time a student fails his or her first quiz or essay, often it can be too late to help the student self

correct. Likewise, as Phelan (2010) discovered, students who had a more immediate response from a lecturer experienced increased feelings of connectedness in a course. It stands to reason that not having that immediacy may result in poor success for distance learning students over all.

Attrition is another challenge with distance learning courses and is a topic that reappears frequently in the literature (Dutton et al., 2001; Zavarella & Ignash, 2009; Kwon et al, 2010; Breslow et al., 2013). While this study did not specifically examine attrition or the reasons behind it, attrition was nevertheless a factor when gathering the samples. Although the distance learning classes originally had 25 students enrolled in each section, the number of distance learning students remaining in the class long enough to write the common assignment essay was 14 or less, with some classes having as few as 7 students remaining in a section of the course. When that happened, a different distance-learning section from the same instructor was used to gather the samples for the essays. While it is impossible to tell in this study why many students did not complete the course, it is logical to consider that the students who did complete the course long enough to complete the common assignment—the persisters, in other words—represented the strongest students enrolled. Thus, the results of this study does not even take the high level of attrition in the distance learning courses into account. In theory, the weaker students may have conceivably dropped out of the course before their outcome level could be assessed.

Equivalency of Courses and Ethnicity

As with the face-to-face and distance learning courses in general, the equivalency tests performed on the student essays when separated by ethnicity also showed no

equivalence between the face-to-face and distance learning groups. In fact, the means of the distance learning groups were consistently lower across the board, with the distance-learning students performing worse than the face-to-face students when the students were separated by ethnicity. When the independent t tests were conducted, there were, however, a few outcome areas where there were no significant differences found when the independent t tests were conducted. One of these areas was for White students in the rubric outcome area of Content, where the face-to-face rubric score ($M=1.32$) and the distance learning rubric score ($M=1.00$) had a difference of $-.315$ and a $p=.080$, which was not significant. This example illustrates, however, the usefulness of conducting equivalency tests, because as noted, while the independent t tests revealed no significant difference between the face-to-face and distance-learning students with regard to Content, this does not mean that the courses were equivalent. As the equivalency tests revealed, the rubric outcome for Content for the White students was not equivalent. For the White student scores for Content, the difference between the means was $.32$, which fell outside the zone of equivalence of $\pm 30\%$ of the mean and the 90% CI $[-.019, .610]$ was not contained within the zone of equivalence either.

When the equivalency tests were conducted for the Hispanic students, none of the rubric outcome areas revealed equivalency between the face-to-face and distance-learning courses. When follow up independent t tests were performed, the results indicated that the areas of Structure, Content, Clarity, and Total Scores, the courses were significantly different. The exception was for the area of Sources. For this area, the difference was $-.32$ ($p=.178$), which was not significant. Still, the equivalency tests

revealed that the difference between the means of .32, because it fell outside the zone of equivalence of $\pm 30\%$ of the mean, did not reveal equivalence.

In addition to this, while none of the rubric outcome areas for Black students were equivalent, the independent t tests conducted revealed that there was no significant difference between the face-to-face and distance-learning groups for some of the outcome areas. The sample size for the Black distance-learning students was very small, however, ($n=5$), so this may have played a role in the results.

Over all, no clear pattern emerged that might reveal how ethnicity might have influenced the students' outcomes in their courses. Due to the limited small sample sizes, it is not possible to make generalized comments about the outcomes of students as influenced by ethnicity, other than to reveal that the face-to-face and distance learning courses did not have equivalent outcomes for any of the ethnicities studied.

Equivalency of Courses and Sex

The equivalency tests performed on the outcome scores for the students when separated by sex likewise revealed no equivalency between the face-to-face and distance learning classes. In addition to this, when independent t tests were conducted to determine the extent and the direction of the difference between the means, a significant difference was found for both the female students and male students for the face-to-face and distance learning students, with the distance learning students having significantly lower outcome scores for both females and males for every rubric outcome area.

There were observational differences between the female and male outcome areas for both face-to-face and distance-learning students. For the face-to-face students, males had higher rubric outcome means than the females in the areas of Structure and Content,

but females had higher rubric outcome means than males for the other two outcome areas of Clarity and Sources. For the distance-learning students, however, the males had lower means than females in the areas of Structure, Clarity, and Sources. These differences were not large, however, and there was no clear pattern observed between the four rubric outcome areas that would lead to any conclusions about the differences between the male and female student outcomes.

Limitations

Because the students involved in this study self-selected themselves into their respective courses in the first place, one limitation points to that self-selection process. It is certainly true that students self-select for many different reasons and some of those reasons might make them predisposed to self-select into a certain type of course. These underlying reasons or traits might be the very traits that contribute to the student's ultimate performance in the course. This self-selection is a limitation because the students' intentions behind taking certain courses can never be known. There was no satisfactory way of ameliorating this limitation. Equivalency testing does not effectively allow for a pre-test post-test design, which would then require determining equivalency between the differences in the pre-test and the post-test rather than equivalencies in the student outcomes as demonstrated by the rubric scores. Another method to help control for this limitation that was considered was to obtain GPA data for each of the students enrolled in the courses and perform *t* tests to ascertain whether differences existed between the students' GPAs prior to their enrollment in either the face-to-face or distance learning courses. This method was not feasible, however, because since students typically take the freshman composition class during their first semester of college, the

majority of students enrolled in the course did not have GPA information available yet. Thus, any data gathered for GPAs would not have been reliable to use in this study.

On the other hand, however, students usually self-select into courses, so one may never be able to extricate whether success or failure in a course is due to some part of that self-selection process, or whether something life-changing has happened to the student, or whether the student has simply had academic troubles all along—these are difficult elements to ferret out in any class. Students might self-select into online courses as opposed to face-to-face courses for many reasons, but they also might choose various times of the day or week to take a face-to-face class (ie: they might take the class in the evening or the morning), or different lengths of classes (8 week classes, 10 week classes), or they might seek out a specific professor, or even a specific campus, all of which can have potentially predict performance in the class if those factors were able to be known and quantified. Colleges do, however, offer—and market—online classes as being equivalent to face-to-face classes. Therefore, determining whether or not the outcomes for these courses are, indeed, equivalent, (no matter how students self-select, since they will self-select regardless) is the main focus of this study.

This study took place in a large metropolitan area in Texas, with a very specific demographic, and the results of this study are not necessarily generalizable to other populations. This study also explored a very specific type of student: A community college student taking a freshman-level English class. The results of this study are therefore not generalizable to other types of students or students studying other subjects, such as math.

The total sample size for this study was $n=100$, with the distance learning sample being $n=50$, and face to face sample being $n=50$. This is a sufficient sample size to represent the overall population. When looking at the demographic subgroups of males and females, White, Hispanic, and Black students, some of the sample sizes were quite small and therefore, caution should be used when interpreting the results. The results do indicate an area for future research, however.

Attrition of the distance-learning students was another limitation in this study. Many researchers have concluded that attrition is a prevalent problem among distance learning students (Carr, 2000; Moody, 2004; Kanuka & Jugdev, 2006; Angelino et al., 2007). Because the essay samples used in this study were written during the second half of the semester, attrition in the examined distance learning classes created a situation in which the samples were obtained only from the student persisters in the courses and would not necessarily reflect the abilities of all of the distance learning students originally enrolled in the courses. In some cases, only 8 to 12 students remained in each of an instructor's two distance learning sections, out of an original total of 50 students (25 in each section). In this situation, when a distance learning section did not have enough student samples, samples were then taken from another distance learning section from the same instructor. This attrition factor, however, potentially created a situation in which only the strongest students in the distance learning classes (or at least the students who persisted in the courses) were the only students examined. One might argue, however, that if all of the student samples could have been collected, the differences between the two courses would have been even greater.

Future Research

This study did not examine how different ethnicities are affected by distance learning platforms, especially with regards to gateway courses such as freshman composition classes. While this study had sufficient sample sizes for the larger samples of students, some of the ethnicities, particularly the Black face-to-face student sample, were small and revealed no statistical difference in some of the independent t tests. Therefore, one area of expanded study of those particular ethnicities would be a good area in which to do further research.

Another area of future research would be to more fully examine the students who self-select into distance learning courses and perhaps determine a predictive model for how those students might fare in their distance learning courses, specifically in freshman composition courses. A larger study might be done, using logistic regression, or some other appropriate method to predict how various ethnicities, traditional vs. non-traditional, full-time vs. part-time, etc., may perform in a gateway course taught online versus face-to-face.

This study examined the specific gateway course of freshman composition courses taught via face-to-face and distance learning, but another area of research might be to examine the equivalence of other gateway courses, including college algebra or statistics, both of which are taught online and both of which few studies are done. Distance-learning modes of delivery are offered for many courses in the core curriculum and they are all assumed to have equivalent outcomes for students.

Implications and Conclusions

The reasons why students may not perform as well in the distance-learning courses are complex and need further study, although many theories abound. This study sought to simply determine whether or not the course delivery methods were equivalent and the results indicated that they are not. The reasons for this are merely conjecture, but worth considering, especially in light of distance-learning courses still being offered to a wide range of students, including developmental and freshman students.

Course design might be an obvious reason for why students may not learn the material as well in a distance-learning course. The course sites for this study had undergone a fairly-extensive review for best practice, which requires that faculty include some collaborative elements of course design, including the requirement that students introduce themselves to the class on a discussion board. There were also some other requirements with regards to the general organization of the course material, but these elements alone may not be enough to fundamentally adapt and change a face-to-face course into a distance-learning course. Most of the course sites at this institution—as is the case with most institutions—were still designed for students to engage in “learner-material” or “learner-instructor” modes, in which the student had to read, then partake of an online lecture, and then write a paper. This might be improved on quite a bit. It may be necessary for distance-learning instructors to fundamentally change how they teach their classes and incorporate many more collaborative learning and constructivist techniques into their courses, rather than patchwork a few discussion boards into an otherwise “learner-material” or “learner-instructor” course design. The more engaged the students are with their classmates and instructors, the more likely they are to continue in

the course, which is what makes a collaborative, constructivist approach so important. Likewise, all students need to see the currency of what they are learning and recognize its relevance, so that message must also come through as well in the design of the course. All of these best-practices might not be in use to the fullest extent possible.

Using technology as a scapegoat is also a common problem in distance-learning courses. Although in the year 2015, most distance-learning technology is no longer cutting-edge or new, instructors and students still struggle with the technology of online courses, which can also impact learning. Students can become frustrated with problems such as a Flash Player not updating or not being able to convert files and upload them to the correct link. Even online instructors can become stymied with technology, many needing several semesters to fully navigate, learn, and deploy some of the many tools that can be used in an online class. For some instructors who find technology challenging, it is arguably easier for them to put up a Power Point or a Word file with some directions and expect the students to glean their instruction on their own, without learning some of the new technology—using wiki features, creating smaller discussion groups, using live voice chat—necessary for more collaborative learning. Online professors are also more likely to accept late work due to perceived technological problems (Dutton et al., 2001). This can cause instructors to then enable students and allow them to turn work in late, which can ultimately hurt their learning because they can fall behind in the course and thus fail to grasp new learning outcomes.

There are many reasons why colleges offer distance-learning courses and will continue to do so in the future. For one thing, distance-learning courses are cost-effective solutions for institutions that want to increase enrollment numbers without having to also

increase brick-and-mortar class space. Colleges and universities must also offer distance-learning courses in order to stay competitive with one another, as well as to remain competitive with for-profit institutions. Distance learning courses also afford a great deal of flexibility for students and therefore, colleges must offer these courses so as to meet the demand of students who require that flexibility. That said, freshman composition is a gateway course and therefore, it is critical for students' future college success to be able to absorb and apply the learning outcomes and pass these courses the first time around. Students who are unable to pass a freshman composition class the first time around risk finding themselves road-blocked and being unable to ultimately receive a certificate or a degree (Clery & Achieving the Dream, 2011). Similarly, because the skills obtained in freshman composition courses are widely used in other college courses, from biology, to history, to psychology, and so on, failure to grasp the learning outcomes can prove devastating to students, even if they do pass the course without learning the material.

For awhile, Massive Open Online Courses (MOOCS) seemed to offer a great deal of potential in terms of allowing colleges to reach larger and larger amounts of students, as well promote learning for those people who did not have the finances to pay for a degree. In spring of 2013, San Jose State University (SJSU) collaborated with the Silicon Valley-based MOOC, Udacity, to offer courses such as Entry Level Mathematics, College Algebra, and Elementary Statistics. While non-SJSU could also enroll in the courses, seats were reserved for SJSU students, including students from under-resourced high schools, community college students, and veterans. The hope of this project was to increase access for students and in this respect, the venture was successful because general enrollment in the courses through Udacity exceeded 15,000 students (Firmin et

al., 2014). The study revealed that the pass rates for the matriculating students was 30% for the Entry Level Mathematics class, 50% for the College Algebra class, and 54% for the Elementary Statistics class. When compared to the face-to-face counterparts for these courses, the distance-learning versions of the courses had lower pass rates. All of the MOOC versions of the classes had pass rates that fell below the historic face-to-face pass rate ranges, with the statistics class being the most notable (Firmin et al., 2014). While these results are not as poor as EdX's initial MOOC, where less than 5% of the students completed a course (Breslow et al., 2013), they are still nothing to rave about. In fact, in a 2013 interview for *Fast Company Magazine*, Udacity's creator, Sebastian Thrun himself expressed dismay at how well his product of mass online education worked. In that interview, he stated, "I'd aspired to give people a profound education—to teach them something substantial....but the data was at odds with this idea" (Chafkin, 2013).

It would seem as though a certain sub-set of students would benefit from a distance-learning course, to be sure. Students who have already have mastered time-management and other success skills would not be as affected by potential disadvantages of distance learning as inexperienced learners or struggling students who are still in the process of learning how to navigate college. Higher education presents its own set of challenges for students as it is: students must adapt to higher levels of rigor, must be able transition into a college culture, must know how to navigate new financial pressures, and must otherwise adapt to a host other pressures. Distance-learning courses compound these pressures further by isolating weaker students from the support systems that brick-and-mortar classrooms and campuses can provide. One can logically conclude that students with less "grit" may simply stop doing the work or drop these courses, which

may account for the high level of attrition seen in distance learning courses. This attrition, while not measured in this study, still played a role at the institution where this study was conducted. Similarly, weaker students may not be able to self-regulate their behavior and spend the time necessary to participate in the courses as fully as they would for a face-to-face course and thus, do not learn as much, even if they do pass the course.

Institutions do not differentiate between face-to-face and distance learning courses when they are marketed to students, other than to tell students the mode of delivery. These courses appear on transcripts as equivalent to any other freshman composition course. There is no separate syllabus for face-to-face or distance learning versions of the courses. When a transfer institution or an employer looks at a transcript, they expect the two modes of delivery will have equivalent outcomes. Students are expected to leave the courses with an equivalent set of knowledge and skills, yet as this study revealed, they apparently do not necessarily have these same skills when they complete their distance-learning course. Face-to-face and distance learning courses do not have equivalent outcomes and in fact, the distance learning students performed worse than the face-to-face students across the board.

In order to ensure equivalence in face-to-face and distance-learning courses, colleges might consider adopting an assessment process similar to this study that attempts to measure student artifacts to determine whether or not the student outcomes are, indeed equivalent. Most colleges—especially community colleges—already use an assessment procedure for core courses, so assessing distance-learning courses should be relatively easy to implement. If differences are then detected between the two methods of delivery,

changes can be made to address any weaknesses in the distance learning courses so as to achieve a process of continued improvement.

If institutions opt to continue to offer distance-learning courses to undergraduate students, they should do so using a great deal of care and caution as they balance the fiscal needs of the institution with the success of their students. First, if at all possible, most freshman students should not take freshman composition online at all. These students are often still “learning to do school” and studies have shown that being more connected to the campus, belonging to student clubs, having the chance to meet with professors and engage in the college by visiting a writing center or tutoring center can help them be more successful in the long run (Tinto, 1997; Schuetz, 2005). Second, if distance learning composition classes are offered to freshmen students, then great care should be taken when placing students in such courses. Students should first receive an assessment of their ability to manage time, to be self-motivated, to understand the technology involved, and otherwise under go an assessment as to whether or not they would make good candidates online courses in the first place. Weaker students with a poor academic success record or low test scores should be encouraged (or required) to take a face-to-face version of the course that fits their schedules. Third, to combat the sense of isolation and disconnectedness associated with distance learning, the course design itself should include elements of a social constructivist model, which allows students to learn cooperatively and from one another instead of being expected to passively absorb information from the instructor in isolation (Pena, 2004).

Fiscal responsibility must be balanced with student success and nowhere is this more evident than the offering of distance-learning composition courses to college

freshmen. Great care needs to be taken when deciding which courses to allow students to enroll, especially when the course can potentially have so much added value to the students' future college careers.

REFERENCES

- Akyol, Z., & Garrison, D. (2011). Understanding cognitive presence in an online and blended community of inquiry: Assessing outcomes and processes for deep approaches to learning. *British Journal of Education Technology*, 42(2), 233-250
- Allen, E. & Seaman, J. (2013). *Changing Course: Ten Years of Tracking Online Education in the United States*. Babson Survey Research Group: USA.
- Angelino, L.M., Williams, F.K., & Natvig, D. (2007). Strategies to engage online students and reduce attrition rates. *Journal of Educators Online*, 4(2), 1-14.
- Armstrong, D. A. (2011). Students' perceptions of online learning and instructional tools: A qualitative study of undergraduate students use of online tools. *Turkish Online Journal Of Educational Technology - TOJET*, 10(3), 222-226.
- Armstrong, W. (2000). The association among student success in courses, placement test scores, student background data, and instructor grading practices. *Community College Journal of Research & Practice*, 24(8), 681–695. Taylor & Francis.
- Retrieved from <http://www.tandfonline.com/doi/abs/10.1080/10668920050140837>
- Arter, J.A., & Chappuis, J. (2006). *Creating and Recognizing Quality Rubrics*. Portland, Oregon: Pearson.
- Ashby, J., Sadera, W. A., & McNary, S. W. (2011). Comparing student success between developmental math courses offered online, blended, and face-to-face. *Journal of Interactive Online Learning*, 10(3), 128-140.

- Bailey, T., Jeong, D., & Cho, S. (2010). Referral, enrollment, and completion in developmental education sequences in community colleges. *Economics of Education Review*, 29(2), 255-270.
- Bers, T., & Schuetz, P. (2014). Nearbies: A missing piece of the college completion conundrum. *Community College Review*, 42(3), 167-183.
doi:10.1177/0091552114525834
- Breslow, L., Pritchard, D.E., DeBoer, J., Stump, G.S., Ho, A.D., and Seaton, D.T. (2013). Studying learning in the worldwide classroom: research into edX's first MOOC. *Research and Practice in Assessment*, 8 (Summer), 13-25.
- Brookhart, S. M. (1999). The art and science of classroom assessment: The missing part of pedagogy. *ASHE-ERIC Higher Education Report* 27(1).
- Burkhardt, J.M., Kinnie, J., & Cournoyer, C.M. (2008). Information literacy successes compared: Online vs. face to face. *Journal of Library Administration*, 48(3), 379-389.
- Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *Chronicle of Higher Education*, 46(23), A39.
- Chafkin, M. (2013). Udacity's Sebastian Thrun, godfather of free online education, changes course. *Distance Education Report*, 17(24), 5-8.
- Chernish, W. N., DeFranco, A. L., Lindner, J. R., & Dooley, K. E. (2005). Does it matter? Analyzing the results of three different learning delivery methods. *Quarterly Review of Distance Education*, 6(2), 87-95.

- Cicchetti D.V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, 6(4), 284-290.
- Clery, S. & Achieving the Dream, I. (2011). Gateway coursework: Time to completion. *Data Notes, Achieving the Dream*. 6(3), 1-5.
- Cross, K. (1998). Why learning communities? Why now?. *About Campus*, 3(3), 4.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal Of Personality & Social Psychology*, 92(6), 1087-1101.
- Dutton, J., Dutton, M., & Perry, J. (2001). Do online students perform as well as lecture students?. *Journal Of Engineering Education*, 90(1), 131-136.
- Equivalence (1988). In *Webster's ninth new collegiate dictionary*. Springfield, MA: Merriam-Webster, Inc.
- Erichsen, E., & Bolliger, D. (2011). Towards understanding international graduate student isolation in traditional and online environments. *Educational Technology Research & Development*, 59(3), 309-326.
- Firmin, R., Schiorring, E., Whitmer, J., Willett, T., Collins, E. D., & Sujitparapitaya, S. (2014). Case Study: Using MOOCs for conventional college coursework. *Distance Education*, 35(2), 178-201.
- Frederickson, N., Reed, P., & Clifford, V. (2005). Evaluating web-supported learning versus lecture-based teaching: Quantitative and qualitative perspectives. *Higher Education*, 50(4), 645-664. doi:10.1007/s10734-004-6370-0

- Garrison, D.R. & Arbaugh, J.B. (2007). Researching the community of inquiry framework: review, issues, and future directions. *The Internet and Higher Education*, 10(3), 157-172.
- Gerlich, R., & Sollosy, M. (2009). Comparing outcomes between a traditional F2F course and a blended ITV course. *Journal Of Case Studies In Education*, 11-9.
- Hallgren, K.A. (2012). Computing inter-rater reliability for observational data: An overview and tutorial. *Tutor Quant Methods Psychology*, 81(1), 23-34.
- Hadden, C. (2000). The ironies of mandatory placement. *Community College Journal of Research & Practice*, 24(10), 823–838.
- Harris, L. (2013). MOOCs - An answer to teacher shortages worldwide. *E-Learning Update*, (38), 1-4.
- Herman, T., & Banister, S. (2007). Face-to-face versus online coursework: A comparison of costs and learning outcomes. *Contemporary Issues In Technology & Teacher Education*, 7(4), 1.
- Hughes, K., & Scott- Clayton, J. (2011). Assessing developmental assessment in community colleges. *Community College Review*, 39(4), 327-351.
- Lawrence, J. A., & Singhania, R. P. (2004). A study of teaching and testing strategies for a required statistics course for undergraduate business students. *Journal of Education For Business*, 79(6), 333-338.
- Lindemann, E. (2001). *A Rhetoric For Writing Teachers*, 4th Edition. New York: Oxford UP.

- Kanuka, H., & Jugdev, K. (2006). Distance education MBA students: An investigation into the use of an orientation course to address academic and social integration issues. *Open Learning*, 21(2), 153-166.
- Kolowich, S. (2012). How will MOOCS make money?. *Inside Higher Ed*. Retrieved from <http://www.insidehighered.com/news/2012/06/11/experts-speculate-possible-business-models-mooc-providers>.
- Kwon, K., Han, D., Bang, E., Armstrong, S., (2010). Feelings of isolation and coping mechanism in online learning environments: A case study of asian international students. *The International Journal of Learning*, 17(1), 343-355.
- Moody, J. (2004). Distance education: Why are attrition rates so high?. *Quarterly Review Of Distance Education*, 5(3), 205-210.
- Moody, R. H., & Bobic, M. P. (2011). Teaching the net generation without leaving the rest of us behind: How technology in the classroom influences student composition. *Politics & Policy*, 39(2), 169-194. doi:10.1111/j.1747-1346.2011.00287.x
- Moore, M.G., & Kearsley, G. (1996). *Distance Education: A systems view*. Belmont: Wadsworth.
- Moskal, Barbara M. (2000). Scoring rubrics: what, when and how?. *Practical Assessment, Research & Evaluation*, 7(3). Retrieved from <http://PAREonline.net/getvn.asp?v=7&n=3>.
- Mumford, S. S. (2015). Practical ideas for teaching citation. *Modern English Teacher*, 24(3), 57-59.

- Pena, C.M. (2004). The design and development of an online, case-based course in a teacher preparation program. *Journal of Interactive Online Learning*, 3(2), 1-18.
- Phelan, L. (2012). Interrogating students' perceptions of their online learning experiences with Brookfield's critical incident questionnaire. *Distance Education*, 33(1): 31-44.
- Reilly, J., Gallagher-Lepak, S., and Killion, C., (2012). "Me and my computer": Emotional factors in online learning. *Nursing Education Perspectives*, 33(2), 100-105.
- Rogers, J. L., Howard, K. I. & Vessey, J. T. (1993). Using significance tests to evaluate equivalence between two experimental groups. *Psychological Bulletin*, 113(3), 553.
- Schuetz, P. (2005). UCLA community college review: Campus environment: A missing link in studies of community college retention. *Community College Review*, 32(4), 60-80.
- Shaw, S. S., & Polovina, S. S. (1999). Practical experiences of, and lessons learnt from, Internet technologies in higher education. *Educational Technology & Society*, 2(3), 16-24.
- Shoikova, E., & Krumova, M. (2010). O*NET and competency based education. *Annual Journal Of Electronics*, 4(1), 145-151.
- Smith, W.L., & Zhang, P. (2009). The academic ethic and the transition to college. *College Student Journal*, 43(1), 86-98.
- Stout, R. H. (2014). *Influence of race, ethnicity, sex, time away from secondary education and parental education on college readiness as measured by the texas success*

- initiative assessment* (Doctoral Dissertation). Retrieved from Dissertations & Theses @ University of Houston - Clear Lake. (1655431129).
- Texas Higher Education Coordinating Board (2013). *ACGM Lower Division Academic Course Guide Manual*. Retrieved from <http://www.thecb.state.tx.us/AAR/UndergraduateEd/WorkforceEd/acgm.html>
- Tinto, V. (1997). Classrooms as communities: Exploring the educational character of student persistence. *Journal of Higher Education*, 68, 599-623.
- Verhoeven, P., & Wakeling, V. (2011). Student Performance In A Quantitative Methods Course Under Online And Face-To-Face Delivery. *American Journal Of Business Education*, 4(11), 61-66.
- Walvoord, B.E. (2014). *Assessing and Improving Student Writing in College*. San Francisco: Josey-Bass.
- Warnock, S., Bingham, K., Driscoll, D., Fromal, J., & Rouse, N. (2012). Early participation in asynchronous writing environments and course success. *Journal Of Asynchronous Learning Networks*, 16(1), 35-47.
- Yi, Y., & Durrington, V. A. (2010). Investigation of Students' Perceptions of Online Course Quality. *International Journal On E-Learning*, 9(3), 341-361.
- Zavarella, C. A., & Ignash, J. M. (2009). Instructional Delivery in Developmental Mathematics: Impact on Retention. *Journal of Developmental Education*, 32(3), 2-13.

APPENDIX A

FRESHMAN COMPOSITION RUBRIC USED IN STUDY

APPENDIX A

FRESHMAN COMPOSITION RUBRIC USED IN STUDY

Outcome	Excellent/Good A/B (2)	Average C (1)	Poor D/F (0)
Structure: How well the essay is structured, whether or not the material is presented in a logical manner. This covers the basic structure of essay (Introduction, thesis, body, conclusion) using transitions throughout	Has a clear thesis; paragraphs have strong topic sentences and paragraphs are focused and have a purpose. Essay has an introduction that “leads into” the thesis statement and a clear conclusion. This level of essay will also make use of transitions to guide the reader throughout. One error or less with regards to structure.	The thesis statement is present, but might need some work with the wording or phrasing. Topic sentences are present, but there might be one paragraph (or two minor parts of two paragraphs) that lose focus or seem to have a debatable purpose, but the over all essay still has basic focus and structure. Conclusion is present. Transitions used throughout, but there might be one area in need of improvement. Three errors or less with regards to structure.	Essay either lacks a thesis statement entirely or else makes a statement that is not debatable or arguable. Topic sentences are weak in more than two paragraphs, or these paragraphs are otherwise unfocused. This essay might just “repeat” or “retell” the same points over and over again in different ways. Essay might ramble and jump around. Might be missing the major essay parts. No transitions. Four or more errors with regard to structure.
Content: This area includes the development of ideas through the essay and how the contribute to the whole. This area also includes the student’s logical approach to his or her argument.	Not only argues the thesis clearly and logically throughout, but also expands on those ideas using logical (and other) evidence. Clearly ties the evidence into the main argument. Does not just stay on the surface, but delves below to make connections for the reader beyond the obvious.	Attempts to argue that thesis, but might miss the mark in some paragraphs. Ideas might be a bit “vanilla” or mundane. Some of the paragraphs might be underdeveloped, or simply repeat ideas (or state the obvious) rather than develop deeper connections for the reader.	Essay reveals multiple problems with regard to substance. Ideas <i>extremely</i> simplistic or ill-formed. A poor essay may have extremely short paragraphs, underdeveloped paragraphs throughout, or have a “draft-like” or “free-writing-ish” feel to it.
Clarity: This area assesses the basic grammar and mechanics of an essay and how well the student followed the rules for Edited American English. This also includes spelling.	A good essay is written clearly, eloquently, with near-perfect grammar, mechanics, and spelling. Correct use of punctuation throughout. Might have very minor errors (1-3) that do not detract from the whole, but otherwise, this essay is error-free with regards to grammar, mechanics, and spelling.	An average essay will have over all competent use of grammar, mechanics, and spelling, but there might be several (4-5) minor errors or one larger error. Readability is not affected. This essay is acceptable in terms of grammar and mechanics, but essay would have benefitted from further editing to fine-tune it.	A poor essay will reveal many substantial errors with regard to grammar, mechanics, or spelling. This essay is extremely “draft-y.” There will be multiple problems throughout that can affect the readability of the document.
Sources —this area assesses the student’s correct use of MLA citations	MLA citations are nearly 100% correct in both the in-text citations and the works cited page. Might have one or two very minor errors in MLA, but otherwise, the citations are error-free.	Essay has both in-text citations and a works cited page. There might be 2-3 minor errors in either the in-text or works cited page (ie: not alphabetizing the works cited page, no hanging indents, adding the URL), but essay exhibits understanding of the process.	Essay is missing one of the two components of the citation (in-text or works cited); may be completely missing works cited; no sources used; cited in APA or other method; or might contain a variety of other errors. This essay does not demonstrate a mastery of the citation process.

Resume

EDUCATION

Dec 2015 Ed.D Educational Leadership/Higher Ed, University of Houston-Clear
Lake

May 2006 M.A. Humanities/Aesthetic Studies, University of Texas at Dallas

Aug 1993 B.A. English Literature, Colorado State University

EMPLOYMENT HISTORY

2014-Present: San Jacinto College South Campus

Department Chair, English, Modern Languages, and Speech

2007-2014: San Jacinto College Central Campus

English Professor

2006-2006 San Jacinto College Central Campus

Adjunct English Professor

Dissertation

- DeLauro, K, Determining Equivalence in Learning Outcomes For Freshman-Level Composition Courses Taught Online and Via Face-To-Face Delivery, (Defended December 1, 2015)