Face-To-Face, Blended, Hybrid, and Online Instructional Delivery Methods: a Comparative Study of English Language Learners’ Grades in a Mathematics Course in a Higher Education Institution in the United Arab Emirates

طرق تقديم التعليم وجهًا لوجه ومختلط ومختلط وعبر الإنترنت: دراسة مقارنة لصفوف متعلمي اللغة الإنجليزية في دورة الرياضيات في إحدى مؤسسات التعليم العالي في الإمارات العربية المتحدة.

by

LISA JOSEPHA BRASHEAR

Dissertation submitted in fulfilment

of the requirements for the degree of

MASTER OF EDUCATION

at

The British University in Dubai

December 2020

DECLARATION

I warrant that the content of this research is the direct result of my own work and that any use made in it of published or unpublished copyright material falls within the limits permitted by international copyright conventions.

I understand that a copy of my research will be deposited in the University Library for permanent retention.

I hereby agree that the material mentioned above for which I am author and copyright holder may be copied and distributed by The British University in Dubai for the purposes of research, private study or education and that The British University in Dubai may recover from purchasers the costs incurred in such copying and distribution, where appropriate.

I understand that The British University in Dubai may make a digital copy available in the institutional repository.

I understand that I may apply to the University to retain the right to withhold or to restrict access to my thesis for a period which shall not normally exceed four calendar years from the congregation at which the degree is conferred, the length of the period to be specified in the application, together with the precise reasons for making that application.

Signature of the student

ABSTRACT

Advances in technology have made different instructional delivery methods possible. Moreover, due to the recent world pandemic, instruction has shifted to completely online delivery. This comparative quantitative study aims at investigating whether instructional delivery methods (such as online, hybrid, blended learning, and face-to-face delivery methods) had an effect on students' grades when teaching mathematics to English Language Learners in a Higher Education Institution (HEI) in the United Arab Emirates. Final course grades, in GPA format, of 574 students were collected over the course of three academic years. Assumptions of analysis of variance (ANOVA) were examined. The ANOVA revealed that there was a significant difference in students' average grades between the different instructional delivery methods. Six comparisons were made: (1) face-to-face versus blended, (2) face-to-face versus hybrid, (3) face-to-face versus online, (4) blended versus hybrid, (5) blended versus online, and lastly (6) hybrid versus online. Postdoc tests showed there were statistically significant mean differences between all six pairwise comparisons. Furthermore, this study used Cohen's d and Hedges' g for the pairwise comparisons of the six statistically significant results.

The results revealed that (a) students scored statistically significantly lower grades in the face-to-face group than in the blended group with a mean difference of 0.72 and an effect size of .47, (b) students in the face-to-face group scored statistically significantly lower than the students in the hybrid group, with a mean difference of 1.17 and an effect size of .79, (c) the face-to-face group displayed the largest statistically significant mean difference of 2.23 lower than the online group, with a very large effect size of 1.49, which was also the greatest effect size of the study, (d) students scored statistically significantly lower grades in the in the blended group than hybrid group, with a mean different of .45 and an effect size of 0.31, (e) students in the blended group scored lower grades than in the online delivery, with a mean difference of 1.51 and an effect size of 1.05, and (f) the hybrid group scored statistically significantly lower than the online group with a mean difference of 1.06 and an effect size of 0.87. Lastly, eta-squared (η^2), and omega squared (ω^2) revealed a medium effect size of nearly 12% of the overall group differences. The study concludes that students in the online method had a significantly higher grade with a large to very large effect size compared to other methods. Further studies can be performed to include students' grades from other faculty, and other classes in HEI's. Lastly, qualitative research is recommended to analyze students and instructors' perspectives on why students perform better online.

ABSTRACT IN ARABIC

جعلت التطورات في التكنولوجيا طرق مختلفة لتقديم التعليم ممكنة. علاوة على ذلك ، بسبب الجائحة العالمية الأخيرة ، تحولت التعليمات إلى التسليم عبر الإنترنت بالكامل. تهدف هذه الدراسة الكمية المقارنة إلى التحقق مما إذا كانت طرق تقديم التعليم (مثل طرق التدريس عبر الإنترنت والتعلم المختلط والتعلم المختلط وطرق التسليم وجهًا لوجه) لها تأثير على درجات الطلاب في دولة الإمارات العربية المتحدة. تم جمع HEIعند تدريس الرياضيات لمتعلمي اللغة الإنجليزية في مؤسسة التعليم العالي ( ) ، من 574 طالبًا على مدار ثلاث سنوات أكاديمية. تم فحص افتراضات تحليل التباين GPAدرجات الدورة النهائية ، بتنسيق عن وجود اختلاف كبير في متوسط درجات الطلاب بين طرق تقديم التدريس المختلفة. تم ANOVA). كشفت ANOVA(إجراء ستة مقارنات: (1) وجهًا لوجه مقابل مخلوط ، (2) وجهًا لوجه مقابل هجين ، (3) وجهًا لوجه مقابل عبر الإنترنت ، (4) مختلط مقابل هجين ، (5) مخلوط مقابل عبر الإنترنت ، وأخيرًا (6) هجين مقابل عبر الإنترنت. أظهرت اختبارات ما بعد الدكتوراة وجود فروق ذات دلالة إحصائية في المتوسطات بين جميع المقارنات الزوجية الستة. علاوة على ذلك ، استخدمت هذه الدراسةCohen’s d and Hedges ’g.للمقارنات الزوجية للنتائج الستة المهمة إحصائيًا

أوضحت النتائج أن (أ) سجل الطلاب درجات أقل إحصائيًا في المجموعة وجهاً لوجه مقارنةً بالمجموعة المختلطة بمتوسط فرق قدره 0.72 وحجم تأثير 0.47 ، (ب) طلابًا في المجموعة وجهاً لوجه. سجلت مجموعة الوجوه أقل إحصائيًا من الطلاب في المجموعة الهجينة ، بمتوسط فرق 1.17 وحجم تأثير .79 ، (ج) أظهرت المجموعة وجهاً لوجه أكبر فرق متوسط ذي دلالة إحصائية أقل من 2.23 المجموعة عبر الإنترنت ، بحجم تأثير كبير جدًا يبلغ 1.49 ، والذي كان أيضًا أكبر حجم تأثير للدراسة ، (د) سجل الطلاب درجات أقل إحصائيًا في المجموعة المختلطة من المجموعة المختلطة ، بمتوسط مختلف .45 و حجم تأثير 0.31 ، (هـ) سجل الطلاب في المجموعة المختلطة درجات أقل من التسليم عبر الإنترنت ، مع فارق متوسط قدره 1.51 وحجم فرق mea nتأثير 1.05 ، و (و) سجلت المجموعة المختلطة درجات أقل إحصائيًا من المجموعة عبر الإنترنت مجموعة مع) ، وأوميغا تربيع (2) عن حجم تأثير متوسط يقارب 12 ٪ من 1.06η2 وحجم تأثير 0.87. أخيرًا ، كشف إيتا - تربيع (الفروق الإجمالية للمجموعة. خلصت الدراسة إلى أن الطلاب في الطريقة عبر الإنترنت لديهم درجات أعلى بشكل ملحوظ مع حجم تأثير كبير إلى كبير جدًا مقارنة بالطرق الأخرى. يمكن إجراء مزيد من الدراسات لتشمل درجات الطلاب من أعضاء هيئة التدريس الأخرى ، والفصول الأخرى في مؤسسات التعليم العالي. أخيرًا ، يوصى بالبحث النوعي لتحليل وجهات نظر الطلاب والمعلمين حول سبب أداء الطلاب بشكل أفضل عبر الإنترنت.

DEDICATION

A special thank you goes to my family, who I haven't seen in over a year due to the pandemic. Moving to the UAE has made me realize how lucky I was to have been born in a family of doctors, successful entrepreneurs, influential lawyers and judges, leading teachers, librarians, accountants, etc. Although my father passed away when I was an infant, I grew up feeling loved and supported, predominantly by independent and strong women in my family, who praised me for hard work. My sisters; Shanti Sandoval and Joline van Steenbergen, my mother; Katrien Verniers, and my daughter; Naomi Elet, are my gasoline to my engine! I thank them for their wise advice, perpetual mindfulness, and their continuous tranquility. I reminisce about our weekends sailing, learning how to navigate on the sea, calculating distances with my step-dad, as well as focusing on getting high grades in school, talking about philosophical, political, and theological issues, cooking healthy and delicious food, and getting educated on all-things-technology by my late grandfather, Jozef Verniers. I am merely a product of the upbringing by my loving family. I hope to have made them proud of my achievements.

Furthermore, an extra special shout-out for my grandmother, Irene Verbuyst-Verniers, who has been my biggest role model and turned ninety this year. Whatever the circumstances, my Oma (grandmother in Dutch) always has my back. She made a point of understanding my thoughts and tried to influence me with unconditional love, never-ending patience, and exemplary forgiveness. She is the kindest, most integral and strongest person I know. Her hard work, outspoken personality, and life achievements inspired me throughout my life. I am fortunate to be her granddaughter.

Intense gratitude goes to my friends, Sofie Valentina Barker, Svetlana Grebenkina, Clair Madden, Sian Marie Townley, and my unparalleled best friend André Van Der Westhuizen. These people supported me during the emotional rollercoaster ride of the year 2020. From encouraging me at times when the ground underneath my feet seemingly disappeared, to cheering me on when I felt on top of the world. I'm sure everyone suffered during the pandemic, but being so far away from home exacerbated the experience. As expats, friends take on the role of family, and these people became my family in the UAE. They made me feel like I belonged by spending quality time, bringing me food, providing counseling, chatting without judgment, and unconditionally standing by me. I thank them for the numerous hours they took out of their lives to give this kind of support to me.

Last but not least, I want to say an unorthodox thank you to my ex-husband, David Elet, who provided support by video calling, messaging, and actively helping our daughter in the USA.

Yes, I feel quite lucky to have had such amazing people standing behind me, pushing me to excel, and believing in me while I attempt to graduate yet another accelerated and distinguished degree. I dedicate my dissertation to these people and by doing so, I hope that they understand precisely how grateful I am for their part in my life.

ACKNOWLEDGEMENTS

My advisor, Dr. Charles Tendai, guided me in the learning process throughout my Masters. He is the kind of professor who works tremendously hard, is well prepared, and takes pride in truly educating his students. Both of us have mathematical and statistical backgrounds, so the positivist paradigm is an effortlessly accepted worldview we share. It takes a nerd to recognize one. I feel Dr. Tendai understood that I have the drive, time, and intellect to perform summa cum laude and he matched my efforts. This is exactly what I needed to achieve even more. A vicious circle of hard work that instigated even harder work. Neither of us would be gratified with mediocrity. Innumerable emails, proofreading, and words of encouragement pushed me to read more, research more, and learn more. I think this journey with Dr. Tendai laid a foundation for future research, and that is actually a dream come true. In conclusion, he is by far the best professor I've ever worked with and I am sure other students will say exactly the same.

Lastly, I thank my colleagues at work, Dr. Gil Briones, PhD student Natalie Howard, Dr. Juan Dempere, and Dr. Ayman Moustafa Ahmed. By working closely with them for the last five years, they inspired me to continue studying and supported me during my many researches.
I hope to be working with them in years to come, not only at our institution, but also on research outside of our institution.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

LIST OF DEFINITIONS AND ABBREVIATIONS

DEFINITIONS

Face-to-face instructional delivery method or traditional instruction: the course is taught in person on campus

Blended instructional delivery method: 50% face-to-face and 50% asynchronous online learning

Online instructional delivery method: 50% synchronous (online at the same time) and 50% asynchronous (online but not at the same time) instructional delivery.

Hybrid method: a combination of the blended method and the online method

ABBREVIATIONS

ELL(s): English Language Learner(s)

HEI: Higher Education Institutions

UAE: United Arab Emirates

IDM(s): Instructional Delivery Method(s)

Chapter 1: Introduction

1.1 Introduction

1.1.1 Background

Instructional Delivery Methods (IDMs) in higher education learning have evolved from the traditional methods preferred by early institutions to the current sophisticated methods. The origin of online education is directly related to the Internet and computer technology (Allen et al., 2016; Moore, 2013; Peters, 1971) combined with pedagogical methods (Moore, 2013). By combining face-to-face instruction with online instruction, methods such as blended and hybrid IDMs are made possible (Allen et al., 2016; Mayadas, Miller & Sener, 2015; NCES, 2020). Due to the Covid-19 pandemic, students were forced to take their classes online (UAE Government, 2020). However, even before the lockdown, online education gained popularity (Bernard et al., 2014; Wavle & Ozogul, 2019).

Regardless of its popularity, the methods that range from entirely face-to-face to strictly online lack specific agreed-upon definitions in the literature (Allen et al., 2016). The face-to-face method, often called traditional instruction, is an IDM where the course is taught in person. The online IDM is an emerging system characterized by classroom materials and lessons being delivered virtually over the Internet (Johnson, Aragon & Shaik, 2000). Following the spectrum proposed by Charles R. Graham (2006), the present research defines '*blended instruction*' as 50% face-to-face and 50% asynchronous online learning; whereas the '*online method'* as 50% synchronous (online at the same time) and 50% asynchronous (online but not at the same time) instructional delivery. Amidst the spring semester of 2020, the blended instruction was constrained to online instruction, due to the Covid-19 lockdown; this is defined as the hybrid instructional method in the present research. As such, the '*hybrid*' method is a combination of the blended method and the online method.

1.1.2 Statement of the Problem

The dynamism in IDMs has been a point of numerous research works aimed at establishing the effectiveness of these methods in learning. However, the literature is divided. While some studies concluded that students are more successful in the online method (Means et al., 2010), other studies show that students are equally successful in face-to-face courses (Bernard et al., 2004; Russell, 1999), and even other studies show that students are more successful when integrating some form on online delivery (Bernard et al. 2014; Means et al., 2010) when teaching English Language Learners (ELLs) mathematics in higher education (Darling-Aduana & Heinrich, 2018; Vo, Zhu & Diep, 2017)

Additionally, blended, hybrid, and online IDMs are new methods for Higher Education Institutions (HEI) in the United Arab Emirates (UAE) (Hussein et al., 2020). Therefore, the emerging question is whether there was a difference in UAE students' grades amongst face-to-face, blended, hybrid, and online IDMs in a HEI. Furthermore, inquiries about the effects of the IDMs on students' grades as a function of ELLs in the discipline of Mathematics remain undetermined. English is the lingua franca for instruction in HEI in the UAE, yet Arabic is the students' first language; therefore, they are ELLs. Answers to these inquiries provide practical significance in the literature. Lastly, there is a noticeable variation in studies on the effects of IDMs in various disciplines. Consequently, lecturers and HEIs may be unsure which IDM works best for their respective disciplines.

The inconsistent findings in the literature and the newness of the literature in the UAE, indicate there is a need for further examination of the effects of IDMs on ELLs' grades in their mathematics courses in HEI in the UAE.

1.1.3 Purpose or Objective of the study

The main purpose of this research was to analyze whether there was a difference in the students' grades depending on online, hybrid, blended learning, and face-to-face course delivery methods when teaching mathematics to ELLs in a HEI in the UAE. Data was collected from a HEI in the UAE. This study used a comparative analysis method to understand the effect of IDMs on the students' grades in their college Mathematics course. Meta-analyses from research outside of the UAE showed dividing results. This study makes a meaningful contribution to world knowledge because there is a clear research gap in the literature, considering UAE students’ novelty in blended, hybrid, and online methods.

1.2 Research Questions

The research questions address whether there was a difference in the students' final grades between the four IDMs. That is, this study compared the differences in student grades in their college Mathematics course between online, hybrid, blended, and face-to-face IDMs. Specifically, is there a statistically significant difference among the four methods of the instructional delivery in terms ELL students' final grades in college mathematics courses in a HEI in the UAE? If so, which of the four IDMs differed from each other, and what were the effect sizes for the statistically significant results?

1.2.1 Research Question (1):

Are there differences in students' final grades among the face-to-face, blended, hybrid, and online IDMs?

1.2.2 Research Questions (2):

If there is/are statistically significant difference(s) between the four methods, exactly which one(s) is/are?

1.2.3 Research Questions (3):

What is/are the effect size(s) of the statistically significant different methods?

1.3 Relevance and Importance for the Study

This research aspired to construe the effects IDMs have on ELL students' grades in their mathematics course. Data was collected from ELL students' grades in a UAE-based HEI where different IDMs are used such as face-to-face, the hybrid method, online learning, and blended delivery methods. The study focused on mathematics grades in GPA format. The findings of the present study may be adopted by HEIs where they may inform their policies on which learning method to use. Especially during the pandemic, structural changes in education might be required. With the statistical comparison of results in this study, specific inferences may be made for future planning in higher education and the UAE's education sector in general. Other researchers may also find this research useful in providing points of criticism and further research on the topic. This may include research on students in other disciplines such as liberal arts, linguistics, or humanities.

Chapter 2: Literature Review

2.1 Instructional Delivery Methods

Higher Education has widely been connected with a) the advancement and acquisition of learning in its ultimate form and b) the distribution of diplomas guiding entry to studied professions (Mgqwashu, 2000). HEIs offer courses and content to their students through a variety of IDMs. An IDM, or mode of instruction, is defined as how instruction is delivered to the learner (Margulieux, McCracken & Catrambone, 2016). There is a continuum of instruction modes, from face-to-face methods to fully online learning, based on integrating technology to deliver instruction (Allen et al., 2016; NCES, 2020).

2.1.1 Historical Background and Definitions of Instructional Delivery Methods in the Literature

The traditional classroom modalities were characterized by face-to-face instructional methods where the course content and supporting learning materials were passed to students in a live interaction between learners and their instructors (Snart, 2010). Instructor and students hold live interactions. Many modern educational systems have shifted away from fully face-to-face methods and integrated some type of online delivery (Bernard et al., 2014; Wavle & Ozogul, 2019). In contrast to face-to-face instruction is fully online instruction, commonly known as distance learning, an IDM where the course content is delivered over the Internet (Johnson, Aragon & Shaik, 2000; Smith & Brame, 2020). It is a form of distance learning conducted electronically over an Internet network between the learners and teachers in what is sometimes referred to as e-learning. The delivery method of online learning can either be synchronous or asynchronous (Almansoori & Akre, 2016). In the synchronous method, online learning occurs simultaneously, and there is live interaction between learners and instructors, such as in live chats and video conferencing. On the other hand, the asynchronous method incorporates electronic media facilitation of online learning through discussion boards, emails, or text messages (Hrastinski, 2008) and does not occur at the same time.

Historically, the definition of distance education included the following essential and unique features: 1) industrialized form of teaching and learning (Peters, 1971) where learning is made possible through the application of modern technology, 2) independence of the learner (Moore, 1973; Wedemeyer, 1977) where the responsibility of learning lies with the student (Wedemeyer, 1981), 3) nontraditional learning, where the distance learner is physically separated from the instructor (Rumble, 1986), and 4) noncontiguous communication (Garrison & Shale, 1987; Holmberg, 1989) where two-way interaction and communication takes place anywhere and anytime through the use of technology.

Employing the most recent definition by the Online Learning Consortium, online courses are courses where all the activities are done online and there are no face-to-face classes nor on-campus activity requirements (Mayadas, Miller & Sener, 2015). Au contraire, in the 13th annual report by the Babson Survey Research Group, an online course is defined as

> "One in which at least 80% of the course content is delivered online. Face-to-face instruction includes courses in which zero to 29% of the content is delivered online; this category includes both traditional and webfacilitated courses. The remaining alternative, blended (or hybrid) instruction, has between 30% and 80% of course content delivered online." (Allen et al., 2016)."

The Babson Survey Research Group is in partnership with the Online Learning Consortium, Pearson, WCET, StudyPortals, and Tyton Partners, and the National Center for Education Statistics' Integrated Postsecondary Education Data System and has been using this definition for thirteen years (Allen et al., 2016).

Furthermore, renewed instructional program classifications and characteristics are produced in the 2020 Edition of the Classification of Instructional Programs by the National Center for Education Statistics, the primary US federal entity for collecting and analyzing data related to education (NCES, 2020). They provided a new definition for Online Educator/Online Teaching in their 2020 Edition: "A program that prepares individuals to teach students at various grade levels through online instructional technologies." (NCES, 2020). While in 2015, the ‘Every Student Succeeds Act’ (ESSA) defined the term "digital learning" as "any instructional practice that effectively uses technology to strengthen a student's learning experience and encompasses a wide spectrum of tools and practices" (“Congress.gov”, 2015, p. 1969), including a variety of technological tools and practices (NCES, 2020).

By introducing a mix of online instruction with face-to-face instruction, new IDMs are formed. As such, blended, hybrid, webfacilitated, synchronous, flexible modes of instruction and more are current practices (Mayadas, Miller & Sener, 2015). However, noticeable divergence exists amongst instructors when defining each IDM. The report by the Babson Survey Research Group on the state of online learning in U.S. higher education uses the following classifications (Figure 2.1) depending on the proportion of content that is delivered online.

Proportion of Content Delivered Online	Type of Course	Typical Description
0%	Traditional	Course where no online technology used — content is delivered in writing or orally.
1 to 29%	Web Facilitated	Course that uses web-based technology to facilitate what is essentially a face-to-face course. May use a learning management system (LMS) or web pages to post the syllabus and assignments.
30 to 79%	Blended/Hybrid	Course that blends online and face-to-face delivery. Substantial proportion of the content is delivered online, typically uses online discussions, and typically has a reduced number of face-to-face meetings.
80+%	Online	A course where most or all of the content is delivered online. Typically have no face-to-face meetings.

Figure 2.1: Prototypical IDM classification by the Babson Survey Research Group

According to Bonk and Graham (2012), blended learning refers to an instructional learning method that combines the traditional face-to-face method and digital-based distributed learner engagement methods and was first introduced around the start of the twenty-first century (Vo, Zhu & Diep, 2017). Researchers emphasize computer usage in the blended instructional delivery system (Graham, 2006). More specifically, the report by the Babson Survey Research Group (Allen et al., 2016) classified online courses as courses that utilize a minimum of 80% of online technology for the content materials, assessments, exams, and communication. While courses instructed in a face-to-face setting, possibly including 0-29% instructional materials and assessments online, are classified as face-to-face (Allen et al., 2016). Lastly, the report classifies blended instruction as courses that combine face-to-face with 30-79% online technology to deliver content, assessments, or electronic communication (Allen et al., 2016).

Note that the report of classification by the Babson Survey Research Group makes no distinguishment between blended and hybrid learning, as if they are interchangeable. Similarly, the ESSA of 2015 does not differentiate between hybrid or blended learning and states that hybrid or blended learning occurs under transpires under direct instructor guidance ("Congress.gov", 2015; NCES, 2020). However, the ESSA of 2015 does not put a range on the amount of the content that is delivered online, yet states that instruction, in part, takes place through online delivery of instruction ("Congress.gov", 2015; NCES, 2020). Hybrid instruction, just like blended instruction, is concerned with combining traditional personal interactions in class with technology-based IDMs. The main difference with blended instruction is that whereas in the blended method, the online learning materials are meant to complement and enrich the traditional learning experience, in the hybrid method, the virtual materials are considered part of the central lesson plan (Castle & McGuire, 2010).

The history of distance learning stretches back to the beginning of the digital age, where personal computer accessibility and Internet connectivity became widespread in the late twentieth century (Moore, 2013). Graham (2006) points out that the concept of distance learning has its origins in British higher learning institutions where correspondence courses were delivered through email exchanges between learners and instructors. Most of the early online courses were conducted through intranet networks developed within the learning institutions (Moore, 2013). With time, HEIs adopted blended learning methods to cater to the growing demand for flexibility in IDMs from students (Paul & Jefferson, 2019). In the UAE, the face-to-face delivery method of learning had dominated the early stages of higher education. In the wake of the new millennium, tertiary institutions in the UAE have invested heavily in online digital learning to cope with the complexity and dynamism of contemporary learners (Hussein et al., 2020). Most higher learning institutions have been forced to close down to prevent the spread of the virus among teachers and students (UAE Government, 2020). Therefore, the outbreak of COVID-19 has made most institutions to rethink their method of instruction.

2.1.2 Definitions used in the Present Research

When defining each IDM, a wide variety of definitions exists in the literature. Therefore, in this section, the researcher defines face-to-face, blended, hybrid and online IDMs.

The face-to-face course in the present research is similar to the "web facilitated" classification by the Babson Survey Research Group (Allen et al., 2016) as web-based technology is used to assist when teaching students in a brick-and-mortar setting. BlackBoard Learn is used as a learning management system. Students take the course in a face-to-face setting. The blended instructional method is defined as an instructional method that combines online and face-to-face delivery. Students take the course 50% in a face-to-face setting and 50% in an online setting where they watch videos, assignments, and assessments in BlackBoard Learn. Hence the course is blended 50/50 online/face-to-face.

As the pandemic entered the UAE, the blended instructional method was replaced by fully online instruction and is referred to as the hybrid instructional method in the present research. The hybrid method was put in place due to Covid-19 after only a few weeks of blended instruction. Therefore, in this study, the hybrid instructional method is different from the blended instructional method, contrary to some definitions in the literature where they are considered the same. A vital element of the hybrid method in this research is the instruction started out as blended, but then moved to the online method from March 2020 to keep students safe while continuing education during the pandemic (UAE Government, 2020). Numerous hours were spent on converting the current curriculum content to a format more suitable for online delivery. In the summer semester of 2020, the IDM was completely online. The students first learned 50% face-to-face and 50% asynchronously, but the face-to-face lessons were replaced by online classes, hence the course ended up being instructed 50% synchronously and 50% asynchronously. The research named 50% synchronously and 50% asynchronously the online instructional method.

2.2 Mathematics Online

Prior to the digital era, mathematics stood out as the discipline that would usher in a new edge of computerization. This meant that mathematics was the stepping stone towards fostering a digitalized education system (Graham, 2006). As the world rapidly advances into the fourth industrial revolution, online mathematics lessons are becoming mainstream with the rapid spread calling for innovative ways to combat inefficiencies that may result.

The teaching approach for STEM disciplines is different from other disciplines because STEM subjects give more attention to a) applying and testing hypotheses using a linear line of reasoning in the pure disciplines and b) developing problem-solving skills in the applied disciplines (Neumann et al., 2002). Research has shown that blended learning methods would be suitable for handling the complex nature of delivering mathematical lessons in higher institutions (Gürsul & Keser, 2009; Vo, Zhu & Diep, 2017), because mathematics instruction is based on grounded theories and adopts a direct approach of teaching (Annand, 2011).

Some studies note no significant difference in students using the traditional face-to-face method compared to those using blended methods (Bernard et al., 2004; Russell, 1999), while other studies find that more students fail in online IDMs (Shea and Bidjeran, 2017), and even other studies show that blended, hybrid, or online methods outperform the traditional methods (Bernard et al., 2014; Darling-Aduana & Heinrich, 2018; Mahmud, 2018; Means et al., 2010; Vo, Zhu & Diep, 2017; Xu et al., 2018)

For the discipline of mathematics, studies have recommended an online-based curriculum that is entirely problem-based (Ortiz, 2006). Moreover, other researchers argue that due to the need to develop complex thinking skills in problem-solving, Mathematics learners require a collaborative online learning environment (Gürsul & Keser, 2009). Problem-based applications and the use of instructional technologies or learning objects for illustrations in the mathematics classrooms have seen an improvement in the way instructors deliver lessons and the overall understanding of mathematical concepts by students (Baki & Çakıroğlu, 2010). This merge of mathematics and arts has been made possible by the multimodal communication enabled by the Internet.

Most tertiary institutions in the UAE were already offering distance learning classes in liberal arts disciplines; the Coronavirus pandemic disruptions have necessitated structural changes that have accommodated large-scale online classes (Hussein et al., 2020). These changes have included core sciences and mathematics courses in fully online programs. Even though online mathematics programs gained traction in the country before the Covid-19 pandemic, studies showed numerous challenges specific to online delivery methods in mathematics and sciences. Abdulla Aldarayseh (2020) summarized these challenges as a decry of the lack of hands-on learning activities and interactions in labs and classrooms and difficulties in managing learners' behaviors during online classes. The study also recommended more specific training to science and mathematics instructors for a smooth transition.

Through the Internet, there has been a good transformation in the learning practice for both mathematics students and teachers. However, this mode of instructional delivery has been criticized in ways researchers believe the method could impair some student's cognitive abilities in solving mathematical problems. For instance, the dependence on calculators for simple calculations and the different ways in which expressions are written in online platforms have been attributed to low mathematics grades in some institutions (Borba, 2012). This is why some instructors in an online mathematics class will prefer to do long mathematical calculations using the traditional pen and paper and turn to computers for simulations. Thus, it is imperative to note that the effectiveness of online mathematics classes will be greatly influenced by students' and teachers' ability to combine traditional and digital ways of learning actively.

2.3 English Language Learners in Mathematics

Numerous studies have shown that most English Language learners experience unique challenges when learning mathematical concepts (Borgioli, 2008). Unlike their English speaking counterparts, these group of learners face the difficult task of simultaneously learning a second language and learning mathematical concepts written in an unfamiliar language. A report published by the Pew Hispanic Centre (USA) revealed that almost 50% of ELL learners in lower grade curricula recorded scores below basic in mathematics (Fry, 2007), which is in contrast with their English native speaking counterparts who exhibited higher scores in the same subject.

The disparity in mathematics performance of ELL students and native English-speaking students continue to be seen in HEIs. Globalization has resulted in an increase in the number of learners seeking higher education in tertiary institutions that are linguistically diverse (Driver & Powell, 2017). Most school programs and materials in these HEIs remain organized and delivered in English despite an influx of non-English speaking students. The institutions also insist on a uniform grading system for students with disregard to the lingual differences (Kersaint, Thompson & Petkova, 2014).

Most institutions of higher learning also insist on ELLs getting their English instructions through mainstreaming, which means the students will have to depend on their regular curriculum to learn English and use it in other subjects. This approach has been found to be least effective in helping ELLs access equal education with their peers (Whiting, 2017). ELLs in mathematics classes have thus had to adapt and modify their ways of learning to enable them to be at par with other students who are fluent in English.

2.4 Conceptual Framework

The proposed study will be anchored on four main concepts and theories that will be used to give more insights into the topic under study. These theories and ideas will form a framework that will be significant in understanding concepts that will help in answering research questions posed in this study. The concepts are (1) Ebbinghaus' studies on memory, (2) self-paced learning: study anytime, anywhere concept, (3) the social cognitive theory, and (4) the theory of transactional distance.

2.4.1 Ebbinghaus’ Studies on Memory

German Psychologist and philosopher Hermann Ebbinghaus pioneered studies on human memory through a series of experiments he conducted on himself. His theory focused on exploration of the nature of perception and sensation. According to Ebbinghaus (1913), memory decay occurred rapidly at first but the proportion of decay would level off with time. He reckoned that the same effect of forgetting after a few days also occurs on longer and shorter periods of time. Figure 2.2 demonstrates the ramification of time and memory retention. Memory retention decreases at the greatest rate of 20% within the first day. Furthermore, the Ebbinghaus' forgetting curve shows that if no review takes place, no memory will be retained after 3 days. Additionally, when reviewing the material after one day, 100% of memory is retained and retention decreases at a slower rate of 20% after two days. It will take 7 days to lose all memory. The more material is reviewed, the longer memory is retained. Therefore, the more students review, the more information they retain over time, and the slower they will forget the information.

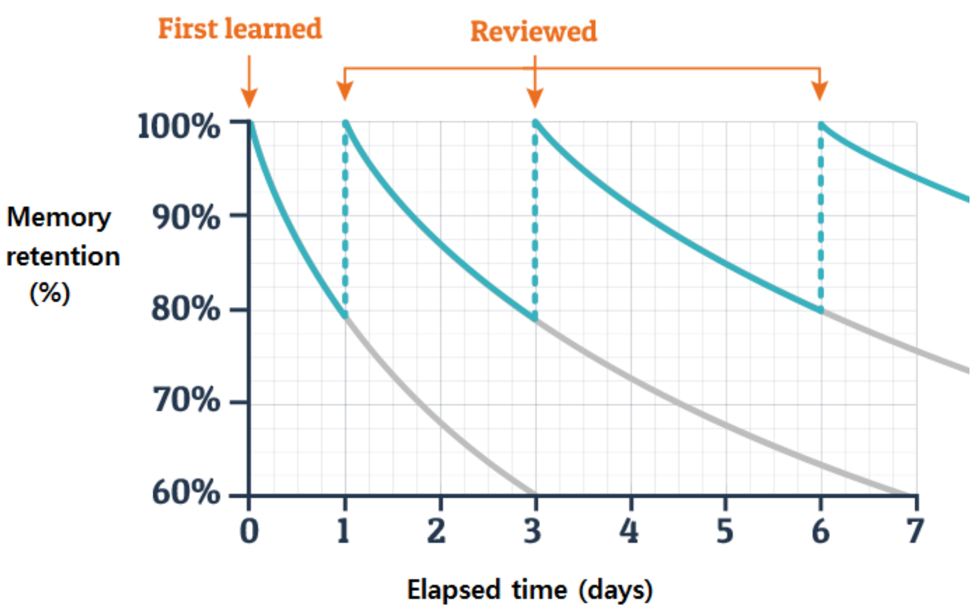


Figure 2.2: Ebbinghaus' forgetting curve and review cycle by Ebbinghaus (1913)

A similar research, later on, confirmed that students taking a new language tend to forget about half of the vocabulary learned within three years. Their memories would then remain constant in subsequent years (Bahrick et al., 1994). Ebbinghaus' experiments also birthed a learning principle he referred to as the spacing effect. Under this principle, he opines that effective learning will occur when a similar study is spread out over a period of time than when it is consolidated at the same time (Ebbinghaus, 1913). It thus suggests that students with limited time to study tend to learn less than those with more time to study a similar subject. He emphasizes on the effects of overlearning which he believes makes our memories vulnerable. His experiments proved the capacity of the human brain to be trained to make sense and meaning out of meaningless items (Schwartz, 2020).

The principles put forward by Ebbinghaus will be significant in this research since the study will be engrossed in cognitive abilities of ELL students in the respective learning methods being investigated. His findings will help the researcher to apply the concepts in coming up with findings from the research participants.

2.4.2 Self-paced Learning: Study Anytime, Anywhere Concept

This concept is synonymous with distance learning or online learning techniques where education extends beyond the traditional brick and mortar classrooms or regular school days. This concept focuses on the interaction between humans and technological devices such as computers, mobile phones, tablets, and television in their access to education remotely (Muir, 2017). It is also referred to as mobile learning and comprises pedagogical approaches where instructors' direct instructions are given to individual learners in a dynamic and interactive space where students work at their own pace (Muir, 2017).

The concept of M-learning allows learners to be flexible and take advantage of digital mobile devices that allow them to learn while attending to other duties. This concept is essential in this study since the study emphasizes current technological learning approaches, especially during the COVID-19 pandemic, where human mobility has been hampered.

2.4.3 The Social Cognitive Theory

This theory was fronted by psychologist Albert Bandura in 1986. It is a modification of previous Bandura's works on social learning theory. The theory, illustrated in Figure 2.3, states that learning takes place in a societal environment where the learner interacts with other members of the society in a dynamic and reciprocal interaction of the person, environment, and behavior (Bandura, 1999). According to Bandura (1999), individual experience is a critical element that will influence a learner's ability to adapt to certain dynamic learning behaviors.

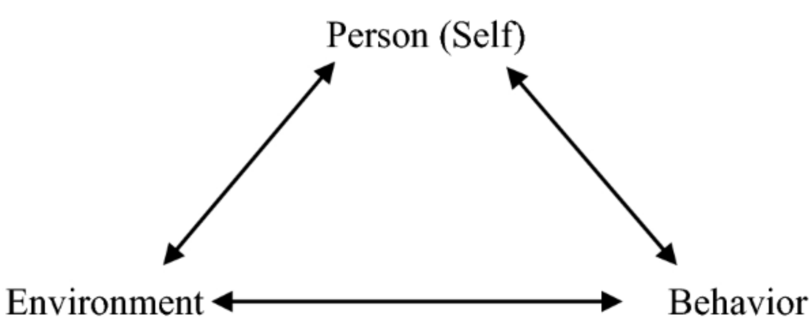


Figure 2.3: Social Cognitive Theory by Bandura (1986)

This concept provides an understanding of learners' motivations since it provides that motivation is specific to a certain context (Wigfield, Cambria & Eccles, 2012). Students' motivation in this context is pegged on motivational concepts such as goals, values, and motives that are varied across learners (Martin, 2004). Through the social cognitive theory, it will be easy to understand how different students are socially constructed to effectively handle self-regulated learning.

2.4.4 The Theory of Transactional Distance

This theory is the work of Michael Moore who argued that distance education was more of a pedagogical concept as opposed to simply a geographical separation of students and their instructors (Moore, 1997). Moore believed that the teacher-student relationship could be systematized into a typology that is modelled around basic constructs of 1) their interactions or dialogue, 2) the instructional programs or structures used, and 3) the learner's nature and degree of self-directedness. Figure 2.4 illustrates the influence of dialogue, structure, and learners' independence on transactional distance. The figure shows how these three variables influence how much the learner feels connected or removed from the lecturer, and how they affect each other. For example, an independent learner with prior knowledge and experience, requires less dialogue and can cope with a more rigid course structure. Inversely, a learner with no prior knowledge or experiences, requires more dialogue and a more flexibility structure. Furthermore, Figure 2.4 shows that when dialogue between the instructor and student isn't effective, students will need to be more independent to feel connected to the instructor and the material. Lastly, as the transactional distance increases, due to highly structured courses, and low dialogue, the higher the level of autonomy is required of the student.

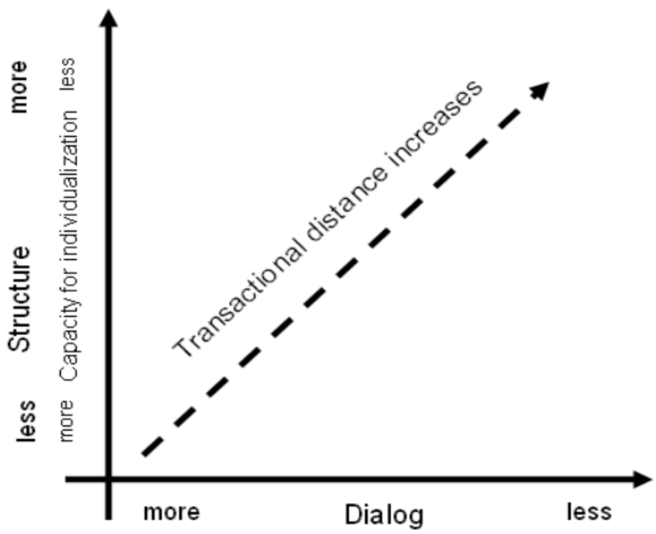


Figure 2.4: Moore's Theory of Transactional Distance (1997) adapted by EDEN (2006)

The adoption of this theory in the field of distance education was based on the fact that special patterns of behaviors were observed due to the geographical separation of students and instructors in distance learning (Moore, 1997). The geographical separation of instructors and students create a psychological space with the potential to cause misunderstandings between the instructor's inputs and the learner's output (Foloon, 2011). This theory will be vital in this study since it will help in understanding how online learning has created a structural communication distance and how student's autonomy in this setup has affected the learning experience.

2.5 Significance and Effect Sizes of Similar Studies

2.5.1 Positivism

The research is a comparative analysis of students' grades in different instructional methods. In essence, 1) cause and effect relationships, 2) research based on deductive reasoning, 3) hypotheses testing, 4) mathematical calculations, and then 5) forming conclusions, have a positivist paradigm as worldview for research (Kivunja & Kuyini 2017).

A positivist paradigm: 1) seeks to explain and predict based on measurable results, 2) accepts that cause and effect are palpable and analytically distinct, and 3) uses a deductive research approach, with an objective epistemology, a naïve realistic ontology, an experimental methodology, and a beneficence axiology (Cohen, Manion & Morrison, 2011; Kivunja & Kuyini, 2017). The epistemology is objective because it is believed that human understanding comes from reasoning, and that we acquire knowledge through research which will make us "more objective in understanding the world around us" (Kivunja & Kuyini, 2017). This research will provide knowledge on whether students have better scores in certain instructional methods. The ontology is naïve realist as the research assumes the belief that truths about students' grades and delivery methods exist and knowledge of their properties can be justified. The methodology is called experimental as the independent variables are manipulated to determine the effect on the dependent variable. Lastly, the axiology is beneficence as the research pursues to optimize positive outcomes for future decisions on delivery methods in HEIs.

2.5.2 No significant differences

One of the earliest leading comparative meta-analysis found a no significant difference between the different IDMs by analyzing 355 studies, reports, and dissertations (Russell, 1999). Similarly, in a 2004 meta-analysis of empirical literature, Bernard et al. (2004) concluded that the online instructional method was as effective as traditional instruction for student achievement. 40,495 students' achievements were comprised in the meta-analysis. A very small (close to zero) yet significantly positive mean effect size for the online method compared with the face-to-face method on student achievement by measuring Hedges' $g+ = 0.0551$.

2.5.3 "Some-form-of-online" > Traditional

In the following headings, literature that claims some form of online learning is more effective than traditional learning will be reviewed.

2.5.3.1 Vo, Zhu, and Diep, Study (2017)

This study was a meta-analysis research on the effects of blended learning delivery methods and traditional learning methods on students' performance in HEIs. In this study, 51 effects from 40 studies were assessed with moderating variables being disciplines and teacher's end-of-course evaluation method. The researchers used Cohen's d method to compute effect sizes. The data was analyzed using the Comprehensive Meta-Analysis (CMA) Software and the findings showed that:

1. Blended learning had a small size effect of ($g+ =.385$, $p < .001$) in comparison to the traditional face-to-face teaching method.

2. STEM disciplines exhibited a significantly higher mean effect size of $g+ =.496$, $k = 30$ in comparison to the non-STEM disciplines which scored $g+ = .210$, $k = 20$.

2.5.3.2 Bernard et al. Study (2014)

In a 2014 meta-analysis of comparative studies between the blended method and the face-to-face method, Bernard et al. (2014) found that 1) technology has a positive impact in higher education learning with Hedges' $g = 0.35$, $p < .01$, and 2) students' achievements were higher in the blended learning method with Hedges' $g = 0.334$, $k = 117$, $p < .001$. The researchers analyzed 96 studies that yielded 117 effect sizes. Cohen's *d* method was used as the basic unit for comparing the experimental conditions in this study. The data was analyzed using Comprehensive Meta-Analysis.

2.5.3.3 Means et al. Study (2010)

A meta-analysis study was conducted by Means et al. (2010) in the United States where the researchers evaluated patterns in online learning practices. In this research, 99 different studies comparing online learning and face-to-face learning were analyzed. 45 of these provided sufficient data for estimating 50 independent effects. From the 50 set of contrasts, 11 effects gave significant positive results in favor of online or blended teaching methods while 3 gave significant effects in favor of face-to-face learning. The study used Cohens' *d* and *Hedges' g+* methods in comparing the pairs of data by computing and weighting the effect size for the studies in contrast to the inverse of variance.

Key findings, relevant to the present study, revealed that:

1. Learning outcomes for online students exceeded the outcomes of students taking face-to-face lessons with a size effect averaging +0.20. The mean difference was statistically significant, and the level was at $p < .001$.

2. Instructions delivered in blended learning methods had an advantage over the face-to-face method with a mean size effect of +0.35, $p < .001$ compared to +0.05, $p = 0.46$ for the face-to-face method.

2.5.4 Blended Learning for ELLs

Mahmud (2018) performed a similar study as the study of Bernard et al. (2014), measuring technology's effectiveness in blended learning. His focus was on students learning a language. Applying the Cohen's d method to estimate effect sizes on the effectiveness of technology in blended learning of language, 59 samples showed that:

1. There is an overall effectiveness of technology to blended learning of language with an effect size of ($d = .8$).

2. Blended learning can potentially improve learner's performance for language acquisition (most variables recorded medium effect size of ($d = 0.5$).

Both findings in the study of Mahmud (2018) have larger effect sizes than the study of Bernard et al. (2014), indicating that students learning languages seem to benefit even greater from using technology and blended instructional methods.

Furthermore, Darling-Aduana & Heinrich (2018) analyzed the use of technology on the outcomes of ELLs. Cohen's d method was also used in the study. Findings revealed that:

1. Extensive use of technology caused higher effect sizes in ELLs' outcomes by ($d = 0.50$).

2. Blended instructional methods are more effective in yielding better outcomes for ELLs in Mathematics classes; size effect ($d = 0.29$)

A meta-analysis by Xu et al. (2018) reported the largest effect sizes found in literature similar to the present study. The study used the Hedge's g method to estimate effect sizes regarding technology applications on adult ELLs. 21 effects from 16 studies were analyzed and yielded even larger effect size results: Technology applications produced large effects than traditional teaching methods of $g = .93$ and $g = 1.28$ in fixed effects and random effect model respectively on adult ELLs' quality of education. The large effect sizes are similar to previous findings. However, the mean effect size when analyzing language development in ELLs is substantially larger than in prior meta-analyses that analyzed language development in native students (Xu et al., 2018).

The present research will use the effect sizes of similar studies to compare the effects of this study to other studies, using a standard metric (Maxwell and Delaney, 2004, p. 548), which will increase the relevance and practical importance of the statistical hypothesis tests' results (Cohen, Manion & Morrison, 2011; Olejnik & Algina, 2003)

2.6 Summary

In this chapter, there has been an extensive review of secondary data relevant to the topic under study. The study has incorporated peer-reviewed journal articles and books in a bid to understand what has been previously done in the area of study. The first part delved into the historical background and definition of the four methods discussed under this study. The second part highlighted online teaching methods with a focus on mathematics amid the current situation of the Covid-19 pandemic. The next part defined and explained four concepts and theories that the research was based on. The final part of the chapter looked into similar research conducted in the same area.

Chapter 3: Methodology

3.1 Introduction

The progress of the World Wide Web and educational technology have shaped how course content can be delivered. Course content evolved from being taught face-to-face, to completely online. During the 2020 pandemic, course content was forced to be delivered online worldwide. The question emerging was whether the IDM has an effect on student grades in their classes? Based on this assumption, the purpose of this research was to examine whether there was a statistical significant difference in the students' grades depending on online, hybrid, blended learning, and face-to-face course delivery methods when teaching mathematics to ELL in a HEI in the UAE? Based on the research questions, this study used a comparative analysis method to gain a better perspective of the explanatory variable, delivery method, and the response variable, students' grades.

This chapter describes the methodology of this research. Respectively, research design, descriptive statistics, inferential statistics, reliability, validity and ethical considerations are explained.

3.2 Research Design

Data was collected from Academic Years (AY) 2017-2018, 2018-2019, and 2019-2020. Student grades for each group were collected. This study compared students' grades between the different instructional methods. Therefore, the quantitative comparative analysis utilized descriptive and inferential statistics, to explain the causal inference about the relations between the IDM and students' grades (Kish, 1959; Pickvance, 2001). Comparative research design permits the “study of variation in variables that are controlled in the case of a particular society (Pickvance, 2001).

The averages of each delivery method were compared with the other methods of delivery to observe the significance. Means, frequencies, modes, and standard deviations were computed for each IDM. Moreover, ANOVA’s were calculated for each delivery method to analyze whether there is any statistically significant difference between the different IDMs. Furthermore, post hoc tests were performed to reveal exactly which delivery methods were different. Lastly, effect sizes were calculated to illustrate the practical importance of the statistically significant results.

3.2.1 Participants

A little over 20,000 students are enrolled in the HEI spread out over seven Emirates, with over 1,100 faculty. At the researcher’s campus, approximately 2,480 total students are enrolled, which is about 12% of all students enrolled in this HEI. An average of 588 new students enrolled each year for the last three AYs, which is about 14% of all new enrollments. An average of 270 students were enrolled in Mathematics classes each trimester on campus the last three AYs.

Table 3.1 Overview Participants in each Mode of Instruction during Three Consecutive Academic Years

AY	**2017-2018**			**2018-2019**			**2019-2020**		
Amount of Students	70	79	26	66	73	41	76	119	24
Trimester	Fall	Spring	Summer	Fall	Spring	Summer	Fall	Spring	Summer
Mode	F2F	F2F	F2F	Blended	Blended	Blended	Blended	Hybrid	Online
Total Students per mode	175			256				119	24
Total Students	574								

Study participants included 574 students who took a mathematics course required for any student pursuing a bachelor's degree from the HEI during 3 consecutive academic years. As illustrated in Table 3.1, of the 574 students, 175 were automatically enrolled in a F2F delivery method during AY 2017-2018. The following year, from fall 2018 until fall 2019, 256 students participated in the blended learning method. In spring 2020, 119 participated in the hybrid delivery method, and lastly, during the summer of 2020, 24 students participated in the online course. Students registered for their mathematics course, and their counselor randomly assigned them to different faculty teaching the course. Each trimester the HEI chose the delivery method, for example, students in the spring semester of AY 2019-2020 were enrolled in Hybrid Courses.

3.2.2 Data Collection Instruments

Employing a new taxonomy to define the four IDMs, the online method is where all instructions are delivered strictly online, two hours synchronously and two hours asynchronously per week. Furthermore, the face-to-face method is an IDM where students were taught on campus, facilitated by web-based technology for four hours per week. BlackBoard Learn is used as a learning management system. Additionally, the blended method has two hours face-to-face and two hours asynchronous online learning per week. Lastly, hybrid learning is a combination of the blended method and the online method, and only came into existence during the spring semester of 2020 when students started out in the blended method but then were moved to the online method due to the Covid-19 lockdown.

Student grades were collected from a secure and archived database containing the students' final letter grades. The data of each class was combined with other data for each of the four IDMs. Data of the four delivery methods (online, hybrid, blended, and face-to-face) were analyzed.

3.2.3 Variables

In this comparative analysis, an attempt is made to explain the causal effect of the explanatory variable on the response variable. According to Kish (1959) there are four types of variables in comparative analysis research design: A) dependent variables, often called explanatory or predictor variable, which are the variables the research wants to explain, B) independent variables which are variables that the research presumes explain the variation in the predictor variables, C) uncontrolled variables, which are the uncontrolled variables that influence the dependent variables, and lastly, D) confounding variables, which are independent variables that influence both the independent and dependent variables (Kish, 1959; Pickvance, 2001)

The independent variable consisted of four IDMs. These four groups, or factors, are (1) face to face, (2) blended, (3) hybrid, and (4) online IDM and are a nominal measure in SPSS.

The outcome or dependent or response variable is the students' final grades, measured on a 4.0 grading scale and are these are continuous (interval/ratio), hence a scale measure in SPSS. Table 3.2 shows the grade points with their corresponding percentage ranges and letter grade.

Table 3.2: Letter grade in percentage ranges and corresponding grade points

Letter grade	F	D	D+	C-	C	C+	B-	B	B+	A-	A
%	0-59	60-63	64-66	67-69	70-73	74-76	77-79	80-83	84-86	87-89	90-100
Grade points	0	1	1.3	1.7	2	2.3	2.7	3	3.3	3.7	4

Confounding variables, such as different instructors, qualifications of instructors, years of experiences, major of students, age of students, gender, and more, can plausibly affect the outcome variable (WWC, 2020). A variety of variables can influence students' grades. To reduce the effect of confounding variables and uncontrolled variables, the researcher attempted to control as many variables as possible by keeping them constant. Variables such as faculty member, homogeneous demographics, and learning objectives were unchanged to ensure the internal validity of the research. One instructor taught all the students enrolled in the different delivery methods over three AYs. The study included only one trimester of online instruction, including 24 students, and one trimester of hybrid instruction, including 119 students, while the other two methods included multiple semesters. A large enough sample and randomization of students secure "the same average value between different groups" (Thomas, 2020) and hence will not confound the study (Sternstein, 2010). This study did not systematically match characteristics between the different groups to minimize the effect of confounding. It can be assumed that students were assigned to the faculty's course randomly. Each trimester, students register for mathematics courses, and the counselor assigns them to a faculty member. However, the delivery method was determined by the HEI. For instance, during the summer semester of 2020, all students were automatically enrolled in online classes due to the Covid-19 pandemic, which made the use of single-blinding impossible (Sternstein, 2010), as students are aware of the method they are being instructed in.

3.2.4 Data Collection Procedures

Approval for this research was requested and given by the HEI director (see Appendix B), where the study was conducted. Permission from the manager was given to use the grades of students enrolled in the researcher’s college mathematics courses from Fall 2017 - Summer 2020.
The course numbers of the classes that the researcher had taught in the given time frame were looked up and classified as one of the four IDMs. Then student grades for each course were

collected and combined with other courses for each of the accumulated IDMs in Microsoft Excel and copied to SPSS. Personal or identifying data between students and grades was deleted.

3.2.5 Data Analysis Procedures

The data for this quantitative study were 574 students' final course grades enrolled in a Bachelor mathematics course, which provided empirical data for three consecutive academic years. Final course grades were retrieved from the researcher's courses. Excel and SPSS was used for the data analysis.

3.3 Descriptive Statistics

Descriptive statistics including measures of location, measures of central tendencies, and measures of variation, often called measures of dispersion. As such, mean, median, mode, range, skewness, kurtosis, standard deviation, variance, frequencies, percentages, were calculated. A relative frequency table is made which shows the count in percentage of each grade in each delivery methods. Furthermore, the relative frequency table shows the success rate of each delivery method. The success rate of the delivery method is the percentage of the students that passed in that delivery method. Boxplots and bar graphs of the four groups were chosen to display the measures graphically. The side-by-side bar graph displays the relative frequencies (in percent) of each grade in each delivery method, showing the percentage of the students who achieved a certain grade. A boxplot is a visual representation of variation that displays the minimum, maximum, median, quartile 1, quartile 3, and the mean (Sternstein, 2010, p. 62). Furthermore, descriptive statistics provides practical data before the start of the statistical inference (Verma & Abdel-Salam, 2019).

3.4 Inferential Statistics

Causal relations are a matter of inference (Pickvance, 2001). The samples descriptions are used to infer about the population. That is, are any mean differences found between the four delivery methods genuine, or are they occurring by chance? If these mean differences are genuine, which method is superior to the other delivery methods? The Null Hypothesis (H_0) conjectures that there are no differences among the four delivery methods. The research hypotheses are tested with a 5% level of significance and hence a 95% confidence interval estimate as "range of population values with which a sample statistic is consistent at a given level of confidence" (Sim & Reid, 1999). As such, the inferential statistics is regarded as scientific method (Bluman, 2007; Sim & Reid, 1999), and included in this study are test for 1) normality; skewness and kurtosis, 2) homogeneity; Levene Statistics, 3) *F*-tests; ANOVA, Welch and Brown-Forsythe, Kruskal-Walis and 4) Post Hoc test; Games-Howell, and 5) effect sizes of the comparisons that showed statistically significant differences. Both parametric tests and nonparametric tests are performed depending on assumptions about the data for each test.

The ultimate question of this research is whether and to what extend students' final grades are impacted by the IDM. Meaning, is there a statistically significant difference between the means of the students' final grades across the four different IDMs and what are the effect sizes?

When comparing two means for statistically significant difference, t-tests are used (Abbot, 2011; Bluman, 2007; Huizingh, 2007). To compare all four instructional methods, six t-tests should be used as shown in Table 3.3.

Table 3.3 Group Comparisons

Comparison 1	Face-to-face versus blended IDM
Comparison 2	Face-to-face versus hybrid IDM
Comparison 3	Face-to-face versus online IDM
Comparison 4	Blended versus Hybrid IDM
Comparison 5	Blended versus online IDM
Comparison 6	Hybrid versus online IDM

However, the type I error (α) for t-tests accumulates when performing multiple t-tests on the same data and is referred to as familywise error (Abbott, 2011). The alpha error for a single t-test comparison is 0.05. With six comparisons, the type I error is increased to

and the confidence level , which compromised the overall t-test of more than one comparisons (Abbott, 2011). Therefore one omnibus test is needed, which performs these comparisons without the familywise error by calculating the variation (1) within each sample group, (2) between each sample group and the overall mean, and (3) the total variance from all the data. Such an appropriate statistical technique for testing differences among three or more means, is the Analysis of Variance (ANOVA) or its alternative. (Abbott, 2011; Bluman, 2007; Verma & Abdel-Salam).

The ANOVA answers whether the sample means of the four instructional methods are different enough to reject the Null Hypothesis that the mean students' grades in our population are all equal (SPSS Tutorials, n.d.).

3.4.1 Assumptions for the ANOVA

Running the ANOVA requires checking its assumptions to prevent invalid significant results (Abbott, 2011; Verma & Abdel-Salam, 2019; Lix, Keselman & Keselman, 1996).

The assumptions of ANOVA are (1) independent samples, (2) normally distributed populations, and (3) equal variances within the groups.

The following assumptions of ANOVA were examined:

3.4.1.1 Assumption 1. Independent samples

The four samples are the four - independently chosen from each other - IDMs. There is no relationship between the students in each group or between groups (delivery methods). Each observation was independent of each other.

The dependent variable, students' final course grade, is an interval or ratio level of measurement.

3.4.1.2 Assumption 2. Normal distribution

Researchers evaluate whether the dependent variables are univariate normal by either

1) reporting skewness and kurtosis or 2) perform a test of normality, since the non-normality of the distribution of sample populations effects the Type I error performance. From the SPSS Menu, click **Analyze**, **Descriptive Statistics**, and then **Explore…** In the Explore box, move the Final Grade to the Dependent List and the Instructional Delivery Method to the Factor List as shown in Figure 3.1. In the Plots… tab, select Normality plots with tests as shown in Figure 3.1: Screen showing commands for test of normality.

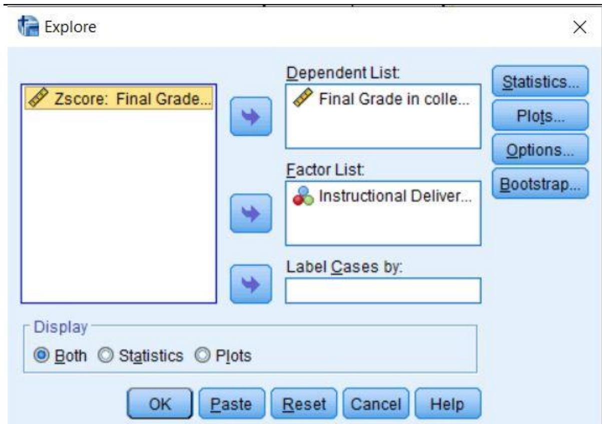


Figure 3.1: Screen showing commands for test of normality

In the Plots… tab, select Normality plots with tests as shown in Figure 3.2.

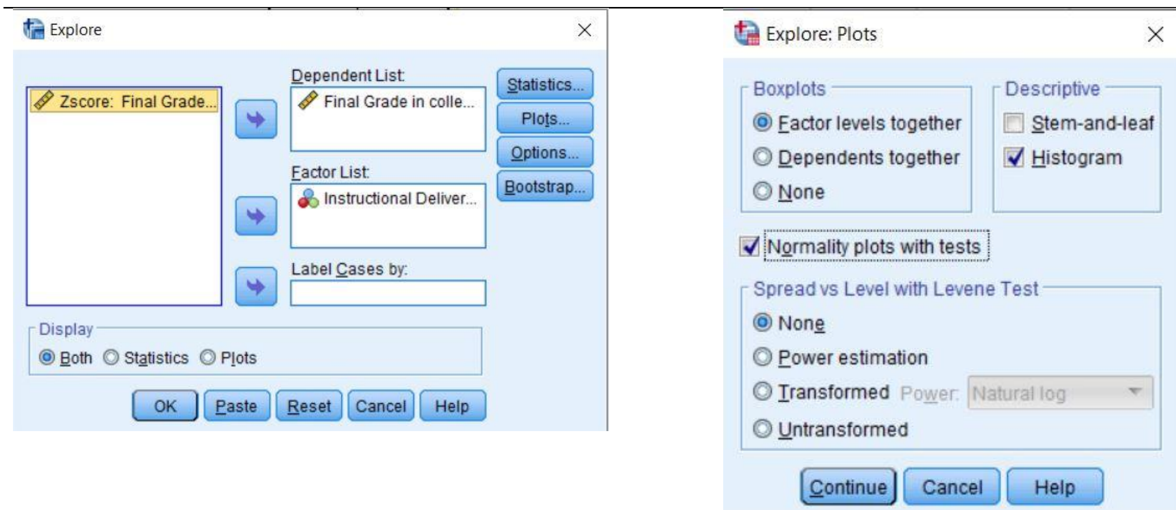


Figure 3.2: Screen showing commands for test of normality

The output of the Kolmogorov-Smirnov and Shapiro-Wilk normality tests are shown in Table 3.4.

Table 3.4: Output of Tests of Normality in SPSS

Tests of Normality

	Instructional Delivery Method	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Final Grade in college Math course	Face-to-Face	.328	175	.000	.767	175	.000
	Blended	.186	256	.000	.847	256	.000
	Hybrid	.183	119	.000	.848	119	.000
	Online	.342	24	.000	.761	24	.000

a. Lilliefors Significance Correction

The Null Hypothesis of normal distribution is rejected for each of the four methods as Sig. < .05 and hence the Alternative Hypothesis of not normal distribution for the four methods is accepted.

However, ANOVA is fairly robust to violations of a normal distribution with large sample sizes due to the Central Limit Theorem (Lix, Keselman & Keselman, 1996; Bluman, 2007; Verma & Abdel-Salam, 2019). Furthermore, t-Test and ANOVA are robust on non-normally distributed data with skewness and kurtosis less than and less than respectively (Boneau, 1960; Posten, 1992; Schmider et al, 2010). Hence prior to conducting the hypothesis testing of equal means (ANOVA or its alternative), the assumptions of normality were evaluated to satisfy all four groups' distributions with skew less than and kurtosis less than .

3.4.1.3 Assumption 3. The homogeneity of the variance.

This assumption of ANOVA assesses "sameness" in the four methods' variances and was performed with the Levene's test in SPSS. An equal variance assumption across the four instructional methods is stated as Null Hypothesis. The Null Hypothesis is retained (or failed to be rejected) if the Levene's test is not significant. If the variance of one of the four methods is substantially different than the variance of the other group, then the zero hypothesis is rejected. So, if the p-value (Sig.) for this test of equality of error variance is less than the .05 significance value, then the hypothesis of equal variances is rejected.

If the Levene's test showed the four methods had similar variances, and the other three assumptions were met, then a regular ANOVA can be performed. If the test revealed that the variance across the four IDMs did not have homogeneity, then the regular ANOVA cannot be employed (Abbott, 2011; Huizingh, 2007; Verma & Abdel-Salam, 2019).

The Null Hypothesis for the homogeneity of variance is that the variances across the four teaching methods are equal.

The Alternative Hypothesis is that the variance across the four teaching methods are not equal.

Table 3.5: Levene's test for the homogeneity of variance assumption in SPSS

Test of Homogeneity of Variances

Final Grade in college Math course

Levene Statistic	df1	df2	Sig.
27.590	3	570	.000

As shown in Table 3.5, the homogeneity of variance, Levene *F* test, $p < .001$ while the alpha level is .05, and p (000) < α (.05) , violating the assumption of homogeneity, hence the test is statistically significant, meaning the variances are significantly different. Therefore, the equal variance assumption, or the Null Hypothesis of the Levene's test is rejected and its Alternative Hypothesis is assumed. Concluding that the variances across the four teaching methods are not equal.

3.4.2 ANOVA F-Test for Research Question 1

An F-Test (ANOVA, or its alternative test) is used to answer research question if there are differences in students' final grades among the face-to-face, blended, hybrid, and online IDMs?"

The Null Hypothesis, , is that the means of the students' grades across the four instructional methods are equal.

The Alternative Hypothesis, , meaning that at least one mean is different.

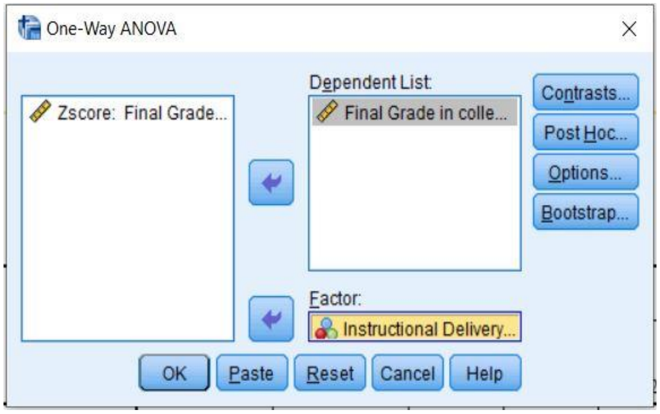
If *P* ("Sig.") < 0.05, then the Null Hypothesis is rejected, and the Alternative Hypothesis is accepted, which implies that the means are not all equal across the four instructional methods.

When assumptions underlying the ANOVA are violated, i.e. homogeneity and normality, the use of the ANOVA *F* test could be invalidated as the probability of mistakenly rejecting a correct Null Hypothesis increases. Alternative *F*-tests and nonparametric techniques are called upon. Nonparametric techniques are less conservative. A popular nonparametric test that compares two or more independent samples is the Kruskal-Wallis Test. While some scholars state that the Kruskal-Wallis Test is used only when the normality assumption is violated, others suggest to use the Kruskal-Wallis Test when the homogeneity condition does not hold also. So, scholars are not in agreement on "the exact null and alternative hypotheses, as well as the assumptions of this test" (Vargha and Delaney, 1998). Furthermore, Vargha and Delaney (1998), clarified in their study on "The Kruskal-Wallis Test and Stochastic Homogeneity" that when the variance homogeneity condition does not hold, robust ANOVA alternatives need to be applied, and therefore also suggest to use the Kruskal-Wallis Test. Contrary, Feir-Walsh and Toothaker (1974) support the use of the ANOVA *F*-test even under violation of assumptions when testing hypotheses about means, and their empirical study concluded that the Kruskal-Wallis test was found to be competitive to the ANOVA *F*-test. In light of the inconsistency, and since this research has a variance homogeneity violation, this research not only performed the robust ANOVA Welch and Brown-Forsythe tests, but also the nonparametric Kruskal-Wallis test, and the more powerful parametric ANOVA test.

3.4.2.1 Regular and Robust ANOVA

In this study, the homogeneity assumption of equal variance for the regular ANOVA was violated, and so a more robust ANOVA test of equality of means must be performed to compare the means across the four methods. Glantz, Slinker, and Neilands (2016) suggest to use the Welch test when the homogeneity assumption is violates, as the Welch test maintains alpha at 5% and has more power, and to use the Brown-Forsythe when the data are skewed. Lix, Keselman & Keselman (1996) reported that the "James and Welch tests performed best under violations of the variance homogeneity assumption".

In the SPSS Menu, click **Analyze**, then **Compare Means**, then **One-Way ANOVA**. Move the Final Grade to the **Dependent List** and the Instructional Delivery Method as **Factor**, then click the **Options…** tab and select the Statistics **Brown-Forsythe** and **Welch** and click on **Means Plot** as shown in Figure 3.3. Click **Continue**, then **OK**.



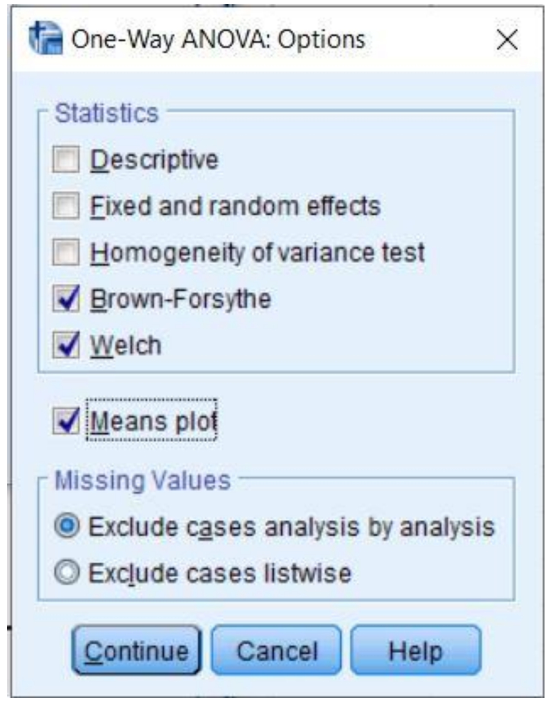


Figure 3.3: Screen showing commands for Robust ANOVA test in SPSS

The results of the regular and robust ANOVA, as well as the Mean Plots are shown in Table 3.6: Result of ANOVA, Welch and Brown-Forsythe tests, and Means Plot

Table 3.6: Result of ANOVA, Welch and Brown-Forsythe tests, and Means Plot

ANOVA

Final Grade in college Math course

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	167.435	3	55.812	26.125	.000
Within Groups	1217.693	570	2.136		
Total	1385.128	573			

Robust Tests of Equality of Means

Final Grade in college Math course

	Statistic[a]	df1	df2	Sig.
Welch	103.503	3	185.424	.000
Brown-Forsythe	37.179	3	498.856	.000

a. Asymptotically F distributed.

Means Plots

3.4.2.2 Nonparametric Kruskal-Walis Test

This is a nonparametric alternative for a One-Way ANOVA and uses a similar Null and Alternative Hypothesis as the ANOVA. Instead of comparing the means, the medians are used when testing for differences among three or more treatment groups (Spurrier, 2003). The Kruskal-Walis Test is an analog to the ANOVA test (Chan & Walmsley, 1997)

From the SPSS Menu, click **Analyze**, then **Nonparametric Tests**, then **Independent Samples…**, and choosing customize analysis. Dragging "Final Grade" and "Method" in the test fields and groups box respectively, as shown in Figure 3.4.

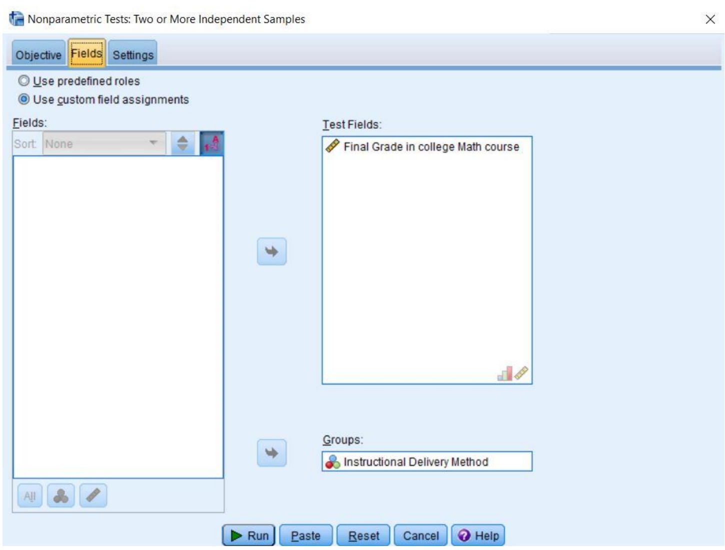


Figure 3.4: Screen showing options chosen for non-parametric test in the Fields Tab

In the tab **Settings**, select Kruskal-Walis and opt for all pairwise comparisons, as shown in Figure 3.5, then click **Run**.

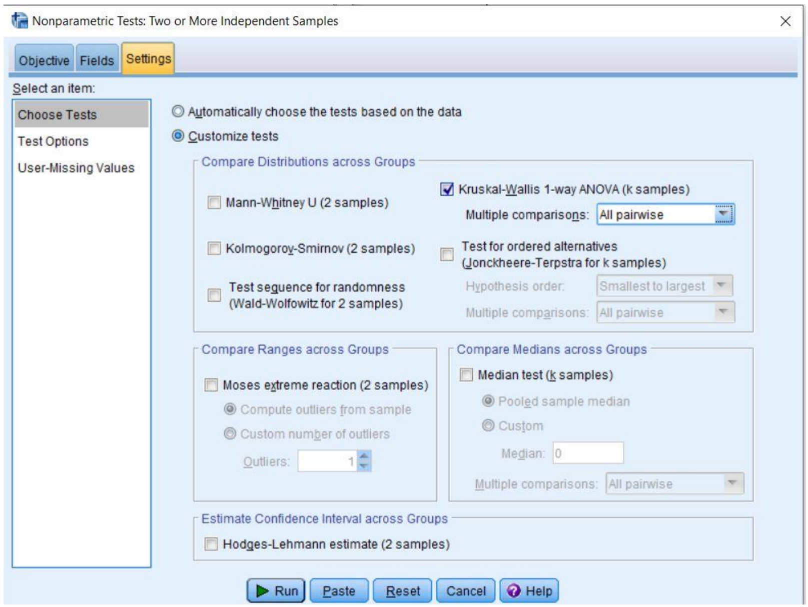


Figure 3.5: Screen showing options chosen for non-parametric test in the Settings Tab

SPSS now shows the Null Hypothesis of equal distributions across the four delivery methods is rejected, see Figure 3.6.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Final Grade in college Math course is the same across categories of Instructional Delivery Method.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

Figure 3.6: Output of Kruskal-Wallis test

3.4.2.3 Conclusion

Since p < .001, for both (1) the Robust test of equality of means, and (2) the nonparametric Kruskal-Wallis, the Null Hypothesis, , that the means of the students' grades across the four instructional methods are equal, is rejected. Therefore, the Alternative Hypothesis, , that at least one mean is different, is accepted.

3.4.3 Post Hoc Analyses for Research Question 2

If the *F*-test or its alternatives show a statistically significant difference; that not all means across the four IDMs are equal, then a post hoc test is performed to determine exactly which means are not equal. The post-hoc comparison of each pair of groups is performed to identify which groups have different population means, and are often referred to as contrast or pairwise comparisons between group means. (Verma & Abdel-Salam, 2019) and answers research question 2:

If there is/are statistically significant difference(s) between the four methods, exactly which one(s) is/are?

This analysis is comparable to individual *t*-tests of pairs of samples of six different *t*-tests to conduct comparisons between the four IDMs:

- Face-to-face versus blended IDM
- Face-to-face versus hybrid IDM
- Face-to-face versus online IDM
- Blended versus Hybrid IDM
- Blended versus online IDM
- Hybrid versus online IDM

The homogeneity of variance assumption was not met, hence the Games-Howell (Kohr & Games, 1974) post hoc procedure in SPSS is used, using a priori alpha level of .05.

In the SPSS menu, click **Analyze**, then **Compare Means**, then **One-Way ANOVA**, list the final grade as dependent, and the method as factor, then opt for **Post Hoc**… Select the Games-Howell comparison as shown in the Figure 3.7.

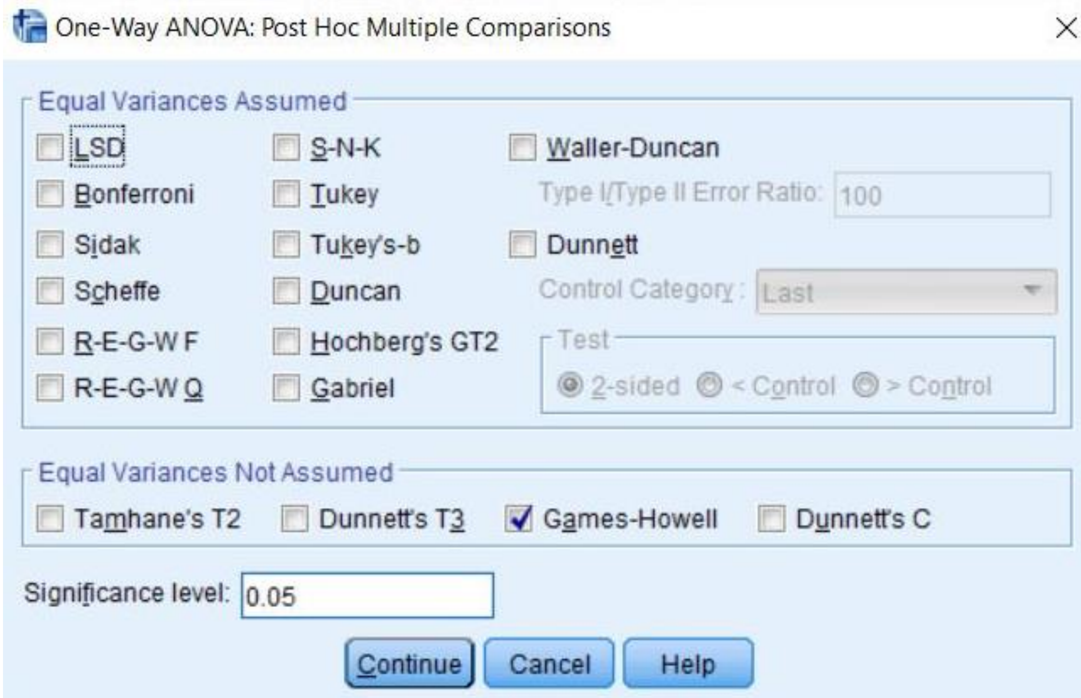


Figure 3.7: Screen showing options chosen for Game-Howell Post Hoc test

Table 3.7: Output of Game-Howell Post Hoc test shows the output in SPSS of Post Hoc test, as well as the standard error (Std. Error) and the 95% confidence interval of the differences between the means. The confidence intervals of 95% means that if the study is performed 100 times, then 95 times the actual (population) difference will fall between the lower and upper bound. Confidence intervals that contain the value zero, result in accepting the Null Hypothesis (Huizingh, 2007). In other words, if zero is included in the interval between the lower and upper bound, then there could be no difference.

Table 3.7: Output of Game-Howell Post Hoc test

Multiple Comparisons

Dependent Variable: Final Grade in college Math course

Games-Howell

(I) Instructional Delivery Method	(J) Instructional Delivery Method	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Face-to-Face	Blended	-.72089*	.15230	.000	-1.1140	-.3278
	Hybrid	-1.17042*	.17062	.000	-1.6114	-.7295
	Online	-2.23131*	.14156	.000	-2.5989	-1.8637
Blended	Face-to-Face	.72089*	.15230	.000	.3278	1.1140
	Hybrid	-.44953*	.15270	.018	-.8444	-.0547
	Online	-1.51042*	.11935	.000	-1.8212	-1.1996
Hybrid	Face-to-Face	1.17042*	.17062	.000	.7295	1.6114
	Blended	.44953*	.15270	.018	.0547	.8444
	Online	-1.06089*	.14198	.000	-1.4304	-.6913
Online	Face-to-Face	2.23131*	.14156	.000	1.8637	2.5989
	Blended	1.51042*	.11935	.000	1.1996	1.8212
	Hybrid	1.06089*	.14198	.000	.6913	1.4304

*. The mean difference is significant at the 0.05 level.

Furthermore, a contrast test was performed to triangulate the results of the post hoc test.

A contrast test also provided the corresponding *t* value to calculate Cohen's *d* = (2t) / sqrt (df). In SPSS Menu, click **Analyze**, then **Compare Means**, then **One-Way ANOVA**, adding the final grades in the dependent list and the IDM as the factor. Then open Options… and enter the following coefficients for each of the 6 contrasts as shown in Table 3.8.

Table 3.8: Coefficient of contrast comparisons

Contrast	Coefficient	Coefficient	Coefficient	Coefficient	Compares
1	1	-1	0	0	F2F vs Blended
2	1	0	-1	0	F2F vs Hybrid
3	1	0	0	-1	F2F vs Online
4	0	1	-1	0	Blended vs Hybrid
5	0	1	0	-1	Blended vs Online
6	0	0	1	-1	Hybrid vs Online

After clicking **continue** and then **OK**, SPSS showed the following output (Table 3.9).

Table 3.9: Output of Contrast Tests in SPSS

Contrast Coefficients

Contrast	Instructional Delivery Method			
	Face-to-Face	Blended	Hybrid	Online
1	1	-1	0	0
2	1	0	-1	0
3	1	0	0	-1
4	0	1	-1	0
5	0	1	0	-1
6	0	0	1	-1

Contrast Tests

		Contrast	Value of Contrast	Std. Error	t	df	Sig. (2-tailed)
Final Grade in college Math course	Assume equal variances	1	-.72	.143	-5.029	570	.000
		2	-1.17	.174	-6.740	570	.000
		3	-2.23	.318	-7.013	570	.000
		4	-.45	.162	-2.772	570	.006
		5	-1.51	.312	-4.841	570	.000
		6	-1.06	.327	-3.244	570	.001
	Does not assume equal variances	1	-.72	.152	-4.733	357.653	.000
		2	-1.17	.171	-6.860	280.811	.000
		3	-2.23	.142	-15.763	157.877	.000
		4	-.45	.153	-2.944	258.004	.004
		5	-1.51	.119	-12.656	124.270	.000
		6	-1.06	.142	-7.472	129.166	.000

The data of this study does not assume equal variance; hence the bottom part is used for the pairwise comparisons. The results of the post-hoc test and the contrast tests are the same.

For all significant pairwise comparisons, the effect size are calculated.

3.4.4 Effect Sizes for Research Question 3

In the previous sections, statistical significance revealed that the effect of the delivery method on students' grades was not due to chance, but that the difference was real. However, with a sample size large enough, nearly all difference will be statistically significant. According to Lakens (2013), the most important outcome of empirical studies are effect sizes. Furthermore, nowadays stipulating an effect size measure is mandatory by many leading journals to "enhance the meaningfulness of the results when statistical hypothesis tests are used" (Olejnik & Algina,

2003) and illustrate the practical importance of results by quantifying the difference between the groups (Cohen, Manion & Morrison, 2011). Moreover, effect sizes make it possible for researchers to compare effects across studies, as effect sizes provide a standard metric (Maxwell and Delaney, 2004, p. 548). Therefore, after finding statistically significant mean differences, effect sizes are calculated. This section will answer Research Questions (3):

What is/are the effect size(s) of the statistically significant different methods?

Although effect sizes have many names, and each its own calculators, yet ultimately all reporting the same entity (Lakens, 2013). As such, coefficient of determination (r^2), eta-squared (η^2), partial eta-squared (η_p^2), Omega squared (ω^2), Cohen's *d*, Hedge's *g*, etc. are often used measures of effect size. Coefficient of determination (r^2), eta-squared (η^2), partial eta-squared (η_p^2), and Omega squared (ω^2) are measures of strength of association, or the ratio (proportion) of the variation in the dependent variable that is associated with or explained by the predictor (independent) variable. While Cohen's *d*, and Hedge's *g* are standardized mean differences, and also calculate the proportion of variance explained by an effect. Standardized mean difference are used to calculate the effect size between two groups. While using the association measures of effect sizes, such as eta-squared, partial-eta squared, or omega-squared represent the overall group differences.

3.4.4.1 Effect Sizes for each Pairwise Comparison

Cohen's d and Hedges' *g* is the most commonly used to report effect size (Lakens, 2013; McGrath & Meyer, 2006). The focus of this study is on group differences, where means and standard deviations are used, therefore it appropriate to express effect size in standard deviation units. Furthermore, Cohen's d is used when determining the effect size for differences between two groups that have similar standard deviations and have the same sample sizes, while Hedges' *g* is used for small sample sizes of less than 20, or when the groups have different sample sizes. Glass's *delta* uses only the standard deviations of the control group (Hedges, 1981).

Table 3.10 shows the formulas used in this study to calculate the effect size for each of the pairwise comparisons that showed statistically significant differences.

Table 3.10: Formulas used in Excel to calculate the effect size for each statistically significant difference

Lakens (2013) Cohen's de formula is	Hedges (1981) Hedges' g formula is
$d_s = \dfrac{M_1 - M_2}{\sqrt{\dfrac{(n_1 - 1)SD_1^2 + (n_2 - 1)SD_2^2}{n_1 + n_2 - 2}}}$	Hedges' $g = d\left(1 - \dfrac{3}{4(n_1 + n_2) - 9}\right)$

Using Excel, Cohen's d and Hedges' g are calculated for the following statistically significant differences:

- Face-to-face vs blended IDM
- Face-to-face vs hybrid IDM
- Face-to-face vs online IDM
- Blended vs Hybrid IDM
- Blended vs online IDM
- Hybrid vs online IDM

Calculations employing Excel can be found in Appendix A.

Cohen (1988) set up the .2, .5, and .8 benchmarks to interpret d as small, medium, and large effects respectively. Moreover, Sawilowsky (2009) classifies 1.2 and 2 as very large, and huge respectively. However, Thompson (2007) emphasized to interpret effects in “direct and explicit comparison against the effects in the related prior literature” and thus not interpret effect sizes using Cohen’s benchmarks.

3.4.4.2 Effect size for overall Group Differences

Lakers (2013) suggests finding the effect size for the overall group differences to perform a power analyses, and for comparisons of effect sizes across studies with the same experimental design. Similarly, to ANOVA, the effect size of the overall group differences illustrates the magnitude of the differences among the four groups, however, it will not explain differences between the specific groups. As explained in the above section 3.4.4.1 Effect Sizes for each pairwise comparison, to identify the differences between specific groups, Cohen’s d and Hedge’s g were performed.

In this study, the effect size of the overall group differences, partial eta-squared (η_p^2), eta-squared (η^2), and omega squared (ω^2) were calculated. They estimate how much variance in the students’ grades are accounted for by the IDM. It does not, however, identify differences between the specific IDMs.

As shown in Figure 3.8, to calculate partial eta-squared (η_p^2), click **Analyze** in the SPSS Menu, then **General Linear Models**, choose **Univariate…** After selecting the final grades as dependent variable, and IDM as fixed factors, click the button **Options**, then select **Estimates of Effect Size**.

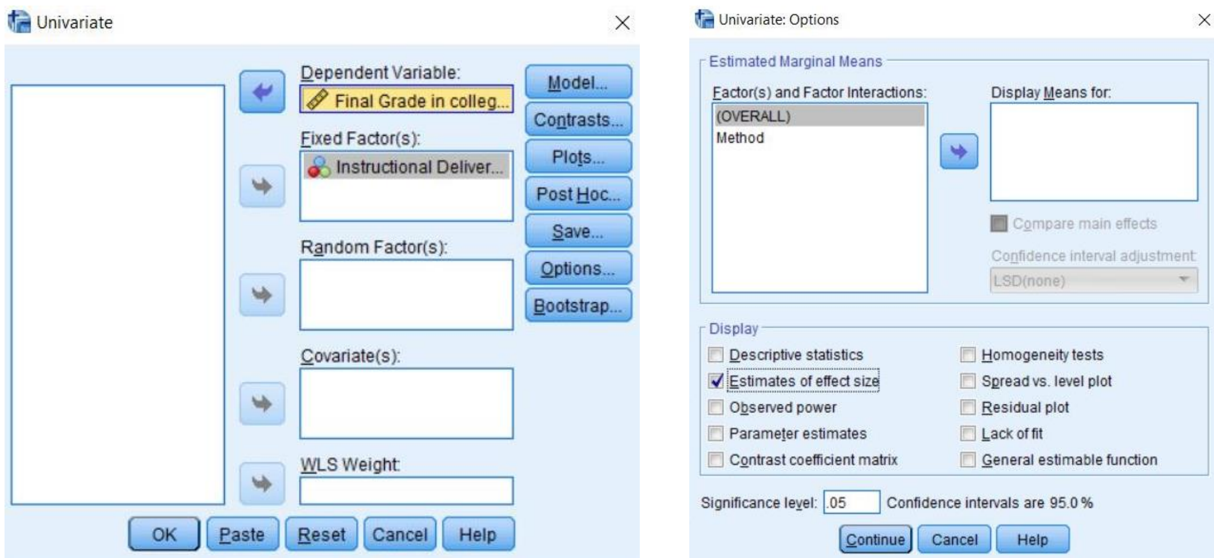


Figure 3.8: Screen showing commands for partial eta squared (ηp2) in SPSS

SPSS calculates partial eta-squared (η_p^2) to be .121, see Table 3.11.

Table 3.11: Output of Partial Eta Squared in SPSS

Tests of Between-Subjects Effects

Dependent Variable: Final Grade in college Math course

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Corrected Model	167.435[a]	3	55.812	26.125	.000	.121
Intercept	1613.582	1	1613.582	755.315	.000	.570
Method	167.435	3	55.812	26.125	.000	.121
Error	1217.693	570	2.136			
Total	3869.230	574				
Corrected Total	1385.128	573				

a. R Squared = .121 (Adjusted R Squared = .116)

When double clicking on the Partial Eta Squared value, the precise value shows as .120880.

As triangulation, Eta Squared was calculated using SPSS > Analyze > Descriptive Statistics > Crosstabs… Rows: Instructional Delivery Method and Column: Final Grade. Click Statistics > Select Eta (Figure 3.9).

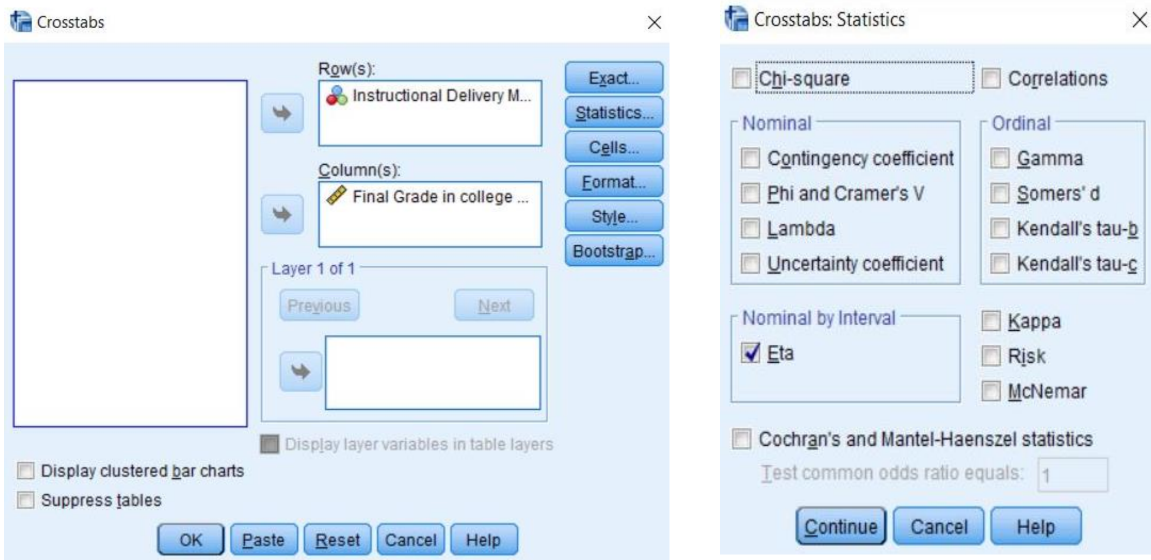


Figure 3.9: Screen showing commands for eta squared (η2) in SPSS

Table 3.12: Output of Eta in SPSS

Directional Measures

			Value
Nominal by Interval	Eta	Instructional Delivery Method Dependent	.370
		Final Grade in college Math course Dependent	.348

Treating Final Grade as dependent variable, Eta is .348, as shown in Table 3.12. When double-clicking on this value, the precise value of Eta is shown as .347679. To find Eta Squared, square the value of Eta, meaning $.347679^2 = .121$, similar to what was found for Partial Eta Squared, since the bias decreases as the sample size increases.

Triangulating further, the measure of effect size ω^2 or Omega Squared was computed to (1) determine the practical significance in the differences among all groups, (2) correct some of the bias that occurs due to overestimations of the actual population effect based on samples averages by eta squared and partial eta squared, and (3) to contextualize the magnitude of the overall differences in means as the robust ANOVA Welch was used due to not meeting the homogeneity of variance assumption (Thompson, 2007). Omega Squared measures the degree of association for a population (Olejnik & Algina 2003).

Cohen (1988) created the following categories to interpret omega squared (ω^2) as strength of association between the independent and dependent variable as small = .01, medium = .06, and large = .14.

The calculated partial eta-squared ($\eta_p{}^2$), eta-squared (η^2), and omega squared (ω^2) showed similar results, 12.1%, 12,1%, and 11.6% respectively. The sample size of this study is large and so the bias associated with overestimating the effect size grows smaller as the sample size grows larger

(Borenstein et al., 2011; Fritz, Morris & Richler 2012; Hayes, 1963). Therefore, it can be noted that the calculated effect sizes of the overall group differences are similar. The effect size of the overall group differences estimates the magnitude of the differences among the four delivery methods. It does not, however, identify differences between the specific IDMs, as this was calculated by the effect sizes of each pairwise comparison.

3.5 Reliability and Validity

3.5.1 Introduction

The research is a comparative analysis of students' grades in different instructional methods. Positivist research includes cause and effect relationships, deductive reasoning, hypotheses testing, mathematical calculations, and then forming conclusions. (Kivunja & Kuyini 2017). In their book on Research Methods in Education, Cohen, Manion & Morrison (2011) state that positivist paradigm research is based on four assumptions for its reliability and validity, namely: determinism, empiricism, parsimony, and generalizability. The assumption of determinism implies that the observed events originated from different factors. This assumption is met, as the study observed students' grades influenced by four different factors, namely four IDMs. Furthermore, the assumption of empiricism is satisfied as the research collects verifiable empirical data, students' grades, which allows hypotheses testing. Additionally, the assumption of parsimony indicated that the research is performed as economically as possible. This research has been performed free of charge; hence the assumption of frugal research is met. Lastly the assumption of generalizability is satisfied because the results and findings of the research were inferenced and generalized using inferential statistics. Satisfying the assumptions, quantitative research methods are appropriate to collect, analyze, interpret, infer, and explain cause and effect within the parameters of the research (Cohen, Manion & Morrison, 2011; Kivunja & Kuyini, 2017).

The validation of a positivist paradigm consists of the following four criteria: internal validity, external validity, reliability, and objectivity (Burns, 2000). These criteria are described in the following sections.

3.5.2 Internal Validity

According to Cohen, Manion & Morrison (2011) internal validity concerns the following question: “do the experimental treatments, in fact, make a difference in the specific experiments under scrutiny”? This study analyzes whether the different IDMs made a difference in the students’ grade by using inferential statistic. This type of question is addressed by statistical hypothesis testing in which the Null Hypothesis of equal means can be retained or rejected (Bluman, 2007; Sim & Reid, 1999). A Type I error, or false positive, occurs when the Null Hypothesis is rejected, when it is in fact true. As explained in the 3.5 Inferential Statistics section on page 9, the alpha error or statistical significance level is set at 5%, indicating that mean differences were found at a rigorous level of statistical significance to ensure the mean differences were unlikely to have occurred by chance (Cohen, Manion & Morrison, 2011). In a mathematical equation P (type I error) = α, or the maximum probability of committing a Type I error is 5%. Inversely related, a Type II error, or false negative, takes place when the Null Hypothesis is wrongly accepted (Cohen, Manion & Morrison, 2011; Muijs, 2004). Besides reporting on significance, and using a large sample size, this research also reports on confidence intervals and effect sizes that are making the difference (Cohen, Manion & Morrison, 2011). Significance, confidence intervals, and effect sizes, respectively, are explained in great detail in section 3.4.2 ANOVA F-Test for Research Question 1 on page 37, and section 3.5.3 Post Hoc

Analyses for Research Question 2 on page 37, and 3.4.3 Post Hoc Analyses for Research Question 2 on page 43.

3.5.3 External Validity

External validity corresponds to what Cohen, Manion & Morrison (2011) explain as generalizability described in section 3.6.1 Introduction on page 29. The study used inferential statistics as scientific method (Bluman, 2007; Sim & Reid, 1999) to generalize the findings of a sample to the population. Research question one used only descriptive statistics to make conclusions about the used sample. Additionally, research questions two and three used inferential statistics to hypothesize with a significance level of 5% that the mean difference in the sample has a low probability of having occurred if there is no relationship in the population (Muijs, 2004).

Furthermore, external validity as mentioned by Cohen, Manion & Morrison (2011) can be negatively influenced by a certain number of factors. Following those factors, this made sure to

1) explicitly describe independent variables,
2) generalize the findings based on good sampling with a target population,
3) sufficiently use dependent variables as effect of the independent variable.

These factors are explained in in section 3.2.3 Variables on page 28.

Additionally, the study used reliable and valid instruments to collect and analyze the data by using statistical programs like SPSS and Excel for hypothesis testing.

3.5.4 Reliability

To guarantee reliability of quantitative research, the conditions underlying hypotheses were tested. Depending on whether the conditions were satisfied, violated, or unclear, appropriate tests were performed. When conditions were violated or unclear, nonparametric tests were carried out and when conditions were met parametric tests (Cohen, Manion & Morrison, 2011; Vargha & Delaney, 1998), which are more powerful (Cohen, Manion & Morrison, 2011) were produced. Furthermore, the research applied internal validation techniques such as running a variety of tests for the same hypothesis, in both SPSS and Microsoft Excel. As such, theoretical and methodological triangulations were used by bringing different and competing theories, mathematical formulas, and tests into play on the same object of study (Cohen, Manion & Morrison, 2011) to check for convergence between the results. The used techniques are meticulously described in both the Methodology section and the Results and Findings section.

3.5.5 Objectivity

If another researcher would reach the same conclusion when investigate the same research questions, with the same participants and data collection (Best, 2012). The study sets aside any "preferences, personality, beliefs and values" (Cohen, Manion & Morrison, 2011) so that the results are neutral and dependent on the data. The three research questions are objective and value free. Additionally, the research conducted data collection and data analysis in an objective manner.

3.6 Ethical Considerations

As a positivist research paradigm, the beneficence axiology intends to reduce risk, harm, or wrong (Kivunja & Kuyini, 2017). No harm was inflicted on the participants. Informed consent is a fundamental concept in research (Best, 2012; Cohen, Manion & Morrison, 2011). Permission to use students' data was sought from the HEI management (see Appendix B). Sensitive or identifying student information was deleted after collection, and described explicitly in section 3.2.4 Data Collection Procedures on page 29.

Chapter 4: Data Analysis & Findings

The results of the data analysis are described in this chapter. The goal of the study was to examine and compare students' final grades in four different IDMs when teaching mathematics to ELL in a HEI in the UAE. The four groups were face-to-face, blended, hybrid, online IDM.

This study analyzes the following research questions:

1. Are there differences in students' final grades among the face-to-face, blended, hybrid, and online IDMs?
2. If there is/are statistically significant difference(s) between the four methods, exactly which one(s) is/are?
3. What is/are the effect size(s) of the statistically significant different methods?

4.1 Descriptive Statistics

The descriptive statistics associated with the final grade of 574 students across the four different IDMs are reported in Table 4.1. It can be seen that the face-to-face group was associated with the numerically smallest mean final grades (M = 1.42) and the online group was associated with the numerically highest mean of final grades (M = 3.65). The numerical mean of the blended method (M = 2.14) lower than the hybrid method (M = 2.59). All four groups' distributions have a skew less than and kurtosis less than , hence the assumption of normality has been satisfied.

Table 4.1 Descriptive Statistics for ELL Students' Final Grades in Mathematics courses across four IDMs in a HEI in the UAE

Method	*n*	*Mean*	*SD*	Skew	Kurtosis
Face-to-face	175	1.42	1.59	.44	-1.51
Blended	256	2.14	2.49	-.35	-1.38
Hybrid	119	2.59	1.32	-.83	-.52
Online	24	3.65	.37	-1.54	2.33

The distribution of the grades in each group and each group's success rate is show in Table 4.2.

Table 4.2: Distribution of grades and success rate in each IDM

Letter Grade	**GPA**	**F2F**	**Blended**	**Hybrid**	**Online**
A	4	7.43%	7.42%	12.61%	29.17%
A-	3.7	5.71%	5.47%	3.36%	20.83%
B	3.3	2.29%	10.55%	13.45%	20.83%
B-	3	4.57%	8.59%	8.40%	8.33%
B+	2.7	3.43%	3.91%	10.08%	12.50%
C	2.3	6.29%	10.94%	10.92%	8.33%
C-	2	5.14%	3.91%	4.20%	0.00%
C+	1.7	2.86%	6.64%	9.24%	0.00%
D	1.3	8.00%	10.16%	4.20%	0.00%
D+	1	2.86%	6.25%	10.08%	0.00%
F	0	51.43%	26.17%	13.45%	0.00%
Success Rate		48.57%	73.83%	86.55%	100.00%

Figure 4.1 shows the side-by-side bar graph which displays percentage of the students who received a certain grade.

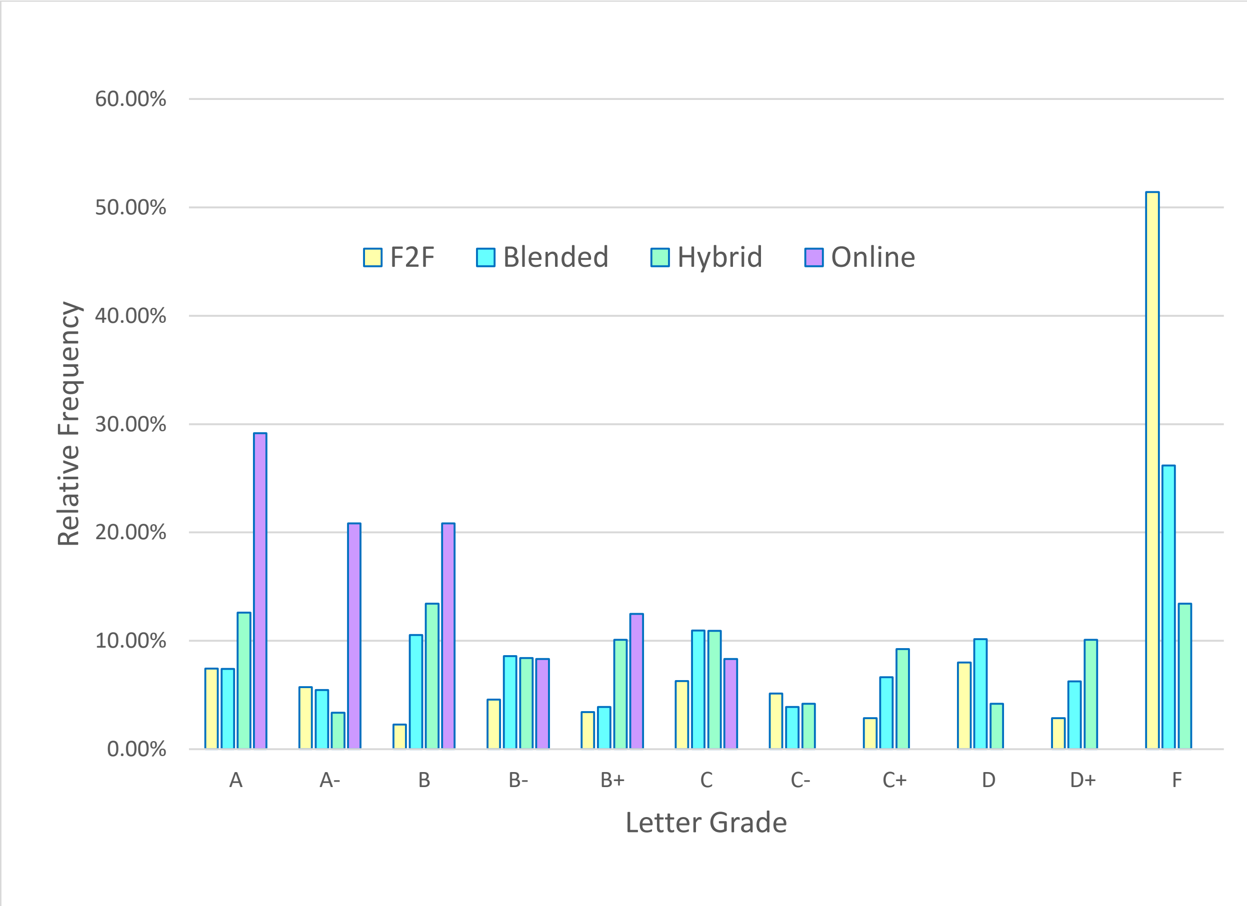


Figure 4.1: Side-By-Side Bar Graph of the Relative Frequencies of each grade in each delivery method.

Figure 4.2 visualizes the measures of the location of the students' final grades and the skewness in a boxplot through showing the quartiles.

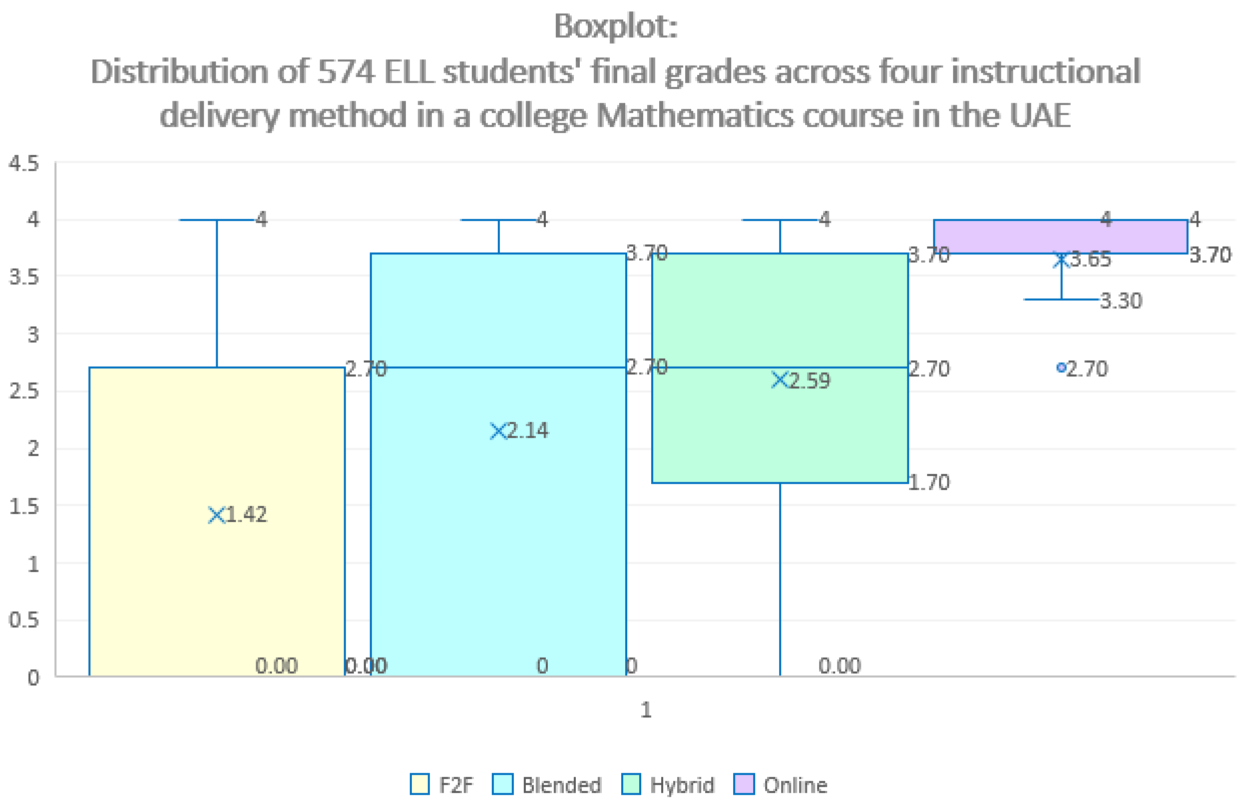


Figure 4.2: Distribution of students' final grades in the four instructional delivery methods

Similarly, Table 4.3 represents the data from the boxplot organized in a table.

Table 4.3: Measures of location of grades across the four IDM

Groups	Min	Q1	Median	Mean	Q3	Max	Range
F2F	0	0	0.00	1.42	2.70	3.70	3.70
Blended	0	0	2.70	2.14	3.70	4.00	4.00
Hybrid	1.70	1.70	2.70	2.59	3.70	4.00	2.30
Online	2.70	3.70	3.70	3.65	4.00	4.00	1.30

4.2 Inferential Statistics

To test the hypothesis that the IDM (face-to-face, blended, hybrid, and online) had an effect on the students' final grades in their college mathematics courses, *F*-tests were performed. After this, to test exactly which means of the four IDMs differed, post hoc comparisons were found. Lastly, the effect size for each of the statistically significant differences, as well as the overall effect size of the instructional methods on students' grades were found. Before conduction the ANOVA, post hoc test, and effect size, the assumptions underlying these tests were examined.

4.2.1 Results of *F*-tests for Research Question 1

To test if there are differences in students' final grades among the face-to-face, blended, hybrid, and online IDMs, four omnibus *F*-tests were performed: the ANOVA, Welch *F*-test, the Brown-Forsythe *F*-test, and the nonparametric Kruskal-Walis test.

Prior to conduction these omnibus tests, the assumptions for carrying out these tests were evaluated. The samples are independently chosen samples. The normality assumption was evaluated and determined to be satisfied as the four groups' distributions were associated with skew and kurtosis less than and, respectively (Schmider et. al, 2010), see Table 4.1 on page 60. Furthermore, the assumption of homogeneity of variances was tested and not satisfied based on Levene's *F* test, $F(3, 570) = 27.59$, $p < 0.001$).

Using the means of the students' final grades, expressed as GPA, the Analysis of Variance test (ANOVA) was run with a 0.05 level of significance. The ANOVA was based on the following Null Hypothesis: , that the means across the four teaching methods are equal.

Table 4.4: Summary (Mean ± Standard Error) – Final Course Grade

Method of delivery	Final Course Grade
Face-to-Face	1.42 ± 0.12
Blended	2.14 ± 0.09
Hybrid	2.59 ± 0.12
Online	3.65 ± 0.07

Two Robust ANOVA tests, Welsch test and Brown-Forsythe *F* test, are preferred instead of the regular ANOVA when testing mean differences as the equal variance assumption is violated and samples sizes are unequal ($N_f = 175$, $N_b = 256$, $N_h = 119$, $N_o = 24$).

The nonparametric Kruskal-Walis Test was used, to triangulate the results of Robust ANOVA tests and is suggested by scholars when the assumptions of ANOVA are violated (Vargha and Delaney, 1998).

Table 4.5: Output of ANOVA, Robust Tests of Equality of Means, and Kruskal-Wallis test in SPSS

ANOVA

Final Grade in college Math course

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	167.435	3	55.812	26.125	.000
Within Groups	1217.693	570	2.136		
Total	1385.128	573			

Robust Tests of Equality of Means

Final Grade in college Math course

	Statistic[a]	df1	df2	Sig.
Welch	103.503	3	185.424	.000
Brown-Forsythe	37.179	3	498.856	.000

a. Asymptotically F distributed.

Hypothesis Test Summary

	Null Hypothesis	Test	Sig.	Decision
1	The distribution of Final Grade in college Math course is the same across categories of Instructional Delivery Method.	Independent-Samples Kruskal-Wallis Test	.000	Reject the null hypothesis.

Asymptotic significances are displayed. The significance level is .05.

All *F*-tests showed the same result: that students' average score on the measure of IDMs indicated a statistically significant difference in the means across the four methods, e.g. Welch's $F(3, 185.424) = 103.503$, $p < .001$.

Since $p < .001$, for all *F*-tests the null-hypothesis that the means are equal, was rejected.

4.2.2 Results of Post Hoc testing for Research Question 2

The Omnibus *F*-test was statistically significant, therefore a post-hoc test and a contrast test were used to show precisely which means are not equal. Specifically, to evaluate the nature of the differences between the four means further, the statistically significant Analysis of Variances was followed-up with Games-Howell's post-hoc tests (Field, 2009) and triangulated with Contrast Tests.

In Section 3.4.3, Table 3.7: Output of Game-Howell Post Hoc teston page 45 shows the Games-Howell's results, and Table 3.9 on page 47 shows the Contrast Test results. In Table 3.7 statistically significant mean differences are indicated with an asterisk. Furthermore all "Sig." are less than 0.05, indicating a statistically significant differences between all means. The 95% confidence intervals (between lower and upper bound) of the difference do not include zero, hence it is unlikely that there is zero difference between the means in the population.

The conclusions from the Games-Howell (**Error! Reference source not found.**) and the Contrast Tests (Table 3.9) are:

1) The face-to-face method has a mean students' grade of 0.72 lower than the blended delivery method.
2) The face-to-face method has a mean students' grade of 1.17 lower than the hybrid delivery method.

3) The face-to-face method has a mean students' grade of 2.23 lower than the online delivery method.

4) The blended method has a mean students' grade of 0.45 lower than the hybrid delivery method.

5) The blended method has a mean students' grade of 1.51 lower than the online delivery method.

6) The hybrid method has a mean students' grade of 1.06 lower than the online delivery method.

It can be observed that students' grades are reported using GPA scores falling between 0 and 4, a minimum and maximum score respectively. Furthermore, students' scores are highest in the online, then the hybrid, followed by the blended and in the last position is the face-to-face delivery method.

A visual depiction of the means and 95% confidence intervals are presented in Figure 4.3.

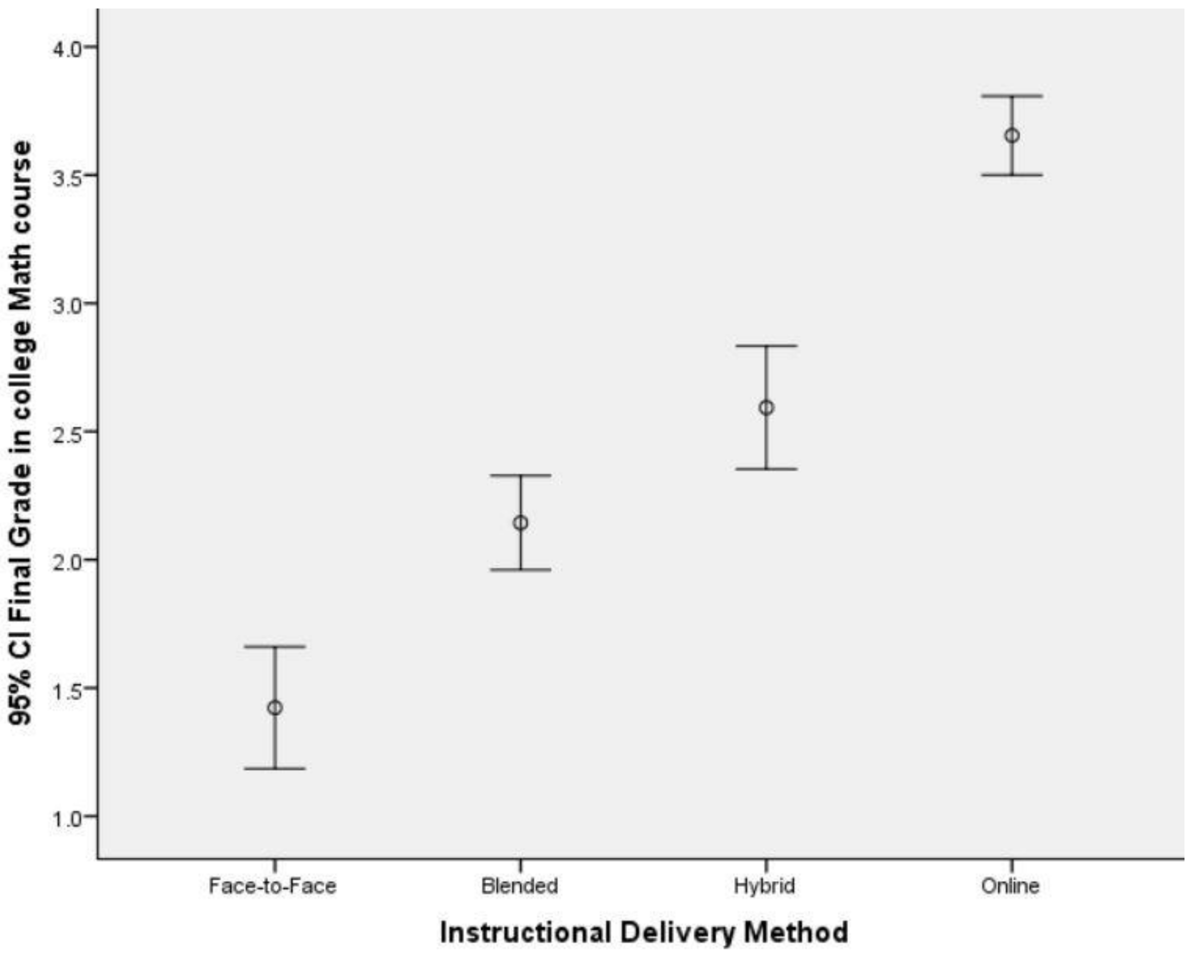


Figure 4.3: Graph of students' grade confidence means and 95% confidence intervals across the four methods in SPSS.

4.2.3 Results of Effect Size for Research Question 3

4.2.3.1 Effect sizes for each pairwise comparison

Table 4.6 shows the effect size for each of the six comparisons by calculating the Cohen's *d* and Hedges' *g* for each of the six comparisons to measure the effect size between the groups. The standardized mean differences express how much variance in the students' grades are accounted for by the IDM.

Table 4.6: Cohen's d and Hedges' g for each of the six comparisons

	F2F vs Blended	**F2F vs Hybrid**	**F2F vs Online**	**Blended vs Hybrid**	**Blended vs Online**	**Hybrid vs Online**
Cohen's *d*	0.470	0.787	1.486	0.312	1.054	0.873
Hedges's *g*	0.469	0.784	1.480	0.312	1.051	0.868

The differences between Cohen's *d* and Hedges's *g* are minimal.

1) The face-to-face method versus the blended delivery method approaches a <u>medium</u> effect size of almost .5
2) The face-to-face versus the hybrid delivery method <u>approaches a large</u> effect size of almost .8
3) The face-to-face method versus the online delivery method has a <u>very large</u> effect size of almost 1.5
4) The blended method versus the hybrid delivery method approaches a <u>small</u> effect size of .3

5) The blended method versus the online delivery method

   has a large of approximately 1

6) The hybrid method versus the online delivery method

   has a large effect size of approximately .9

4.2.3.2 Effect size for overall group differences

The calculated partial eta-squared (η_p^2), eta-squared (η^2), and omega squared (ω^2) showed similar results of around .12 or 12%, which is a moderate effect size (Cohen, 1988). This means that about 12% of the variance in the students' grades are accounted for by the IDM.

4.3 Conclusion

A total of 574 participants were enrolled in one of the four IDMs: face-to-face (n = 175), blended (n = 256), hybrid (n = 119), or online (n = 24). The *F*-analyses were conducted to assess if face-to-face, blended, hybrid, or online instructional methods differ with respect to students' final grades. The independent variable represented the different instructional methods with four groups being represented: 1) face-to-face; 2) blended, 3) hybrid; and 4) online. The dependent variable was the final grade that the students received in their college mathematics course with a range of 0 to 4. An significance level (alpha level) of .05, and therefore a confidence level of 95%, was set for all following analyses.

The test for normality were evaluated and were found to satisfy all four groups' distributions with skew less than and kurtosis less than . However, the Levene's *F* test showed that the assumption of homogeneity of variances was not met ($p < .001$). Therefore, the Welch' F test was used. The one-way ANOVA, analyzing if a difference existed for the mean grades within the four methods, suggested there were statistically significant differences in mean grades by instructional method, Welch' $F(3, 185.424) = 103.503$, $p < .001$, indicating that not all IDMs had the same mean score. The adjusted omega squared ($\omega^2 = 0.12$) indicated that about 35% of the total variation in mean score on students' measure of final grades is attributed to differences between the four IDMs. In other words, 12% of the variance in students' final grades is explained by the IDM.

The pair-wise comparisons (Games-Howell Post Hoc test) were performed to find out which pairs of the four IDMs had significantly different means. Cohen's *h* effect sizes for each significant comparison is given.

The results of the first comparison indicated that students scored statistically significantly lower grades in the face-to-face group (M =1.42, SD = 1.59) than in the blended group (M =2.14, SD = 2.49, $p < .001$) with a mean difference of 0.72 and an effect size of .47.

The second pairwise comparison indicated that students in the face-to-face group (M =1.42, SD = 1.59) scored statistically significantly lower than the students in the hybrid group (M =2.59, SD = 2.32, $p < .001$), with a mean difference of 1.17 and an effect size of .79.

The third comparison, between the face-to-face group (M =1.42, SD = 1.59) and the online group (M =3.65, SD = .37, $p < .001$) also displayed statistically significant mean difference. The face-to-face group has a mean grade of 2.23 lower than the online group, with a very large effect size of 1.49, which was also the greatest effect size of the study.

The fourth pairwise comparison showed that students scored statistically significantly lower grades in the in the blended group (M =2.14, SD = 2.49) than hybrid group (M =2.59, SD = 1.32, $p = .004$), with a mean different of .45 and an effect size of 0.31.

The students in the fifth comparison revealed that students in the blended group ($M = 2.14$, $SD = 2.49$) scored lower grades than in the online delivery ($M = 3.65$, $SD = .37$, $p < .001$), with a mean difference of 1.51 and an effect size of 1.05. Notably, this result revealed the second most significant difference between groups.

The sixth, and last comparison showed that the hybrid group ($M = 2.59$, $SD = 1.32$) scored statistically significantly lower than the online group ($M = 3.65$, $SD = 0.36$, $p < .001$), with a mean difference of 1.06 and an effect size of 0.87.

Chapter 5: Discussion & Recommendations

5.1 Introduction

The start of online education goes hand in hand with Internet technology, pedagogical theories, and computer technology (Allen et al., 2016; Moore, 2013; Peters, 1971). In only a few decades, improvements in technology have broadened instructional delivery possibilities and made more meaningful pedagogical experiences for students feasible, which in turn caused online courses to gain in popularity (Wavle & Ozogul, 2019). This study conveyed findings of prior research that effective instructional methods, including conventional technology, could increase students' grades, and compared the findings of the present study to prior studies to give meaning to the effect sizes.

The research questions analyze if there was a difference in the students' final grades between the four IDMs. The research questions were:

Research Question (1):

Are there differences in students' final grades among the face-to-face, blended, hybrid, and online IDMs?

Research Questions (2):

If there is/are statistically significant difference(s) between the four methods, exactly which one(s) is/are?

Research Questions (3):

What is/are the effect size(s) of the statistically significant different methods?

The following sections will discuss an overview of the study, key findings and conclusions, limitations, implications, and recommendations.

5.2 Overview Study

A comparative analysis probed different IDMs at a single HEI in the UAE to decide if the delivery method made a difference in students' mathematics grades. The quantitative research, vested in a positivist paradigm, used all the students' grades in their mathematics courses taught by one lecturer during three consecutive academic years: from Fall 2017 until Summer 2020. As such, 574 individual student grades were collected from an archived database. Data analysis was performed using Excel and SPSS to calculate descriptive and inferential statistics. Descriptive statistics such as measures of location, measures of central tendencies, and measures of dispersion were found. Graphs and tables were used to visualize and summarize the data. The descriptive statistics was used to start the statistical inference (Verma & Abdel-Salam, 2019). F-tests (ANOVA, Welch F-test, the Brown-Forsythe F-test, and the nonparametric Kruskal-Walis test) were performed to test the hypothesis that the IDM (face-to-face, blended, hybrid, and online) had an effect on the students' grades in their college mathematics courses. Then post hoc comparisons (Games-Howell) were completed to test exactly which means of the four IDMs differed. These results were verified using the contrast tests. Finally, the effect size for each of the statistically significant differences, and the overall effect size of the instructional methods on students' grades were found. Both Cohen's d and Hedges'g were calculated and the differences between Cohen's d and Hedges's g are minimal. Assumptions underlying the tests were examined before performing each test.

5.3 Key Findings and Conclusions

This section provides a summary of the findings and relevant conclusions for the three research questions.

A total of 574 participants were enrolled in one of the four IDMs: face-to-face (n = 175), blended (n = 256), hybrid (n = 119), or online (n = 24). A significance level (alpha level) of .05, and therefore a confidence level of 95%, was set for all following analyses. Note that students' grades are expressed as GPA, ranging between zero and four.

5.3.1 Face-to-Face versus Blended and Hybrid IDM

The results indicated that students scored statistically significantly higher grades in the blended group ($M = 2.14$, $SD = 2.49$, $p < .001$) with a mean difference of 0.72 and a nearly medium effect size ($d = 0.47$) than the face-to-face group ($M = 1.42$, $SD = 1.59$). Furthermore, when comparing the hybrid group ($M = 2.59$, $SD = 2.32$, $p < .001$) with the face-to-face group, an even higher mean difference of 1.17 and a remarkable higher effect size ($d = 0.79$) were found. The effect is categorized as approaching a large size (Cohen, 1988).

The nearly medium and large effect size favoring the blended method and hybrid method respectively over the traditional method is supported by previous studies. However, when comparing prior studies, it should be noted that the present study classified the blended method and hybrid method as two different methods, the latter one started out as a blended method but was moved to completely online instruction during the pandemic. Most studies do not differentiate between blended and hybrid IDMs (Allen et al., 2016; NCES, 2020). Vo, Zhu, and Diep (2017) found a similar effect size ($g^{+} = .496$) favoring the blended method of STEM-disciplines over the face-to-face method, while Means et al. (2010) found a small mean size effect of +0.35 favoring the blended method of all disciplines over the face-to-face method. Furthermore, Mahmud (2018) also concluded that blended learning can improve learner's performance for language acquisition with a medium effect size of d = 0.5. Additionally, Darling-Aduana & Heinrich (2018) also derived that blended instructional methods are more effective in yielding better outcomes for ELLs in Mathematics classes; size effect (d = 0.29). Overall, it can be concluded that integrating some form of online method is beneficial for ELLs' grades when taking college mathematics courses.

5.3.2 Face-to-Face versus Online IDM

The third comparison, the online group (M =3.65, SD = .37, p < .001) had a mean grade of 2.23 higher than the face-to-face group (M =1.42, SD = 1.59), with a very large effect size of 1.49, which was the greatest effect size of the study. The very large effect size is higher than most findings of prior studies (Xu et al., 2018). For instance, in the meta-analysis, Means et al. (2010) found a small size effect averaging +0.20 for online students compared to face-to-face students. Contrary, Bernard et al. (2004) found only a close to zero effect size in a similar meta-analysis. Prior research found that ELLs' grades significantly increased when effectively integrating technology (Darling-Aduana & Heinrich, 2018; Mahmud, 2018; Xu et al., 2018). Xu et al., 2018 found a large mean effect size of 1.3 in their meta-analysis, when comparing the effect of technology on language acquisition in adult ELLs. The very large effect size of the present study is much bigger than prior studies.

5.3.3 Blended versus Hybrid IDM

When comparing the blended group (M = 2.14, SD = 2.49) and the hybrid group (M = 2.59, SD = 1.32, p = .004), the hybrid group outperforms the blended group with a mean different of 0.45

and a small effect size of 0.31. This result indicated a trend that students performed better when classes were taught more online and hence less face-to-face. Students' grades were positively impacted during the semester in which teaching had to be moved to completely online instruction during the pandemic. No studies were found comparing blended with hybrid methods, as most classifications use blended as hybrid as synonyms (Allen et al., 2016; NCES, 2020).

5.3.4 Blended and Hybrid versus Online IDM

The online delivery ($M = 3.65$, $SD = .37$, $p < .001$) demonstrated a higher mean difference of 1.51 and a large effect size of 1.05. than the blended method ($M = 2.14$, $SD = 2.49$). Notably, this result revealed the second most significant difference between groups. Furthermore, the online group scored statistically significantly higher than the hybrid group ($M = 3.65$, $SD = 0.36$, $p < .001$), with a mean difference of 1.06 and a similar large effect size of 0.87. Prior studies found that technology has a positive impact in higher education learning (Bernard et al., 2014), especially when language acquisition is a moderator (Darling-Aduana & Heinrich, 2018; Mahmud, 2018; Xu et al., 2018).

5.3.5 Effect size for Overall Group Differences

The overall group differences showed an effect size of 0.12 ($\eta_p^2 = \eta^2 = \omega^2$) which approaches a large effect size. 12% of the variance in the students' grades are accounted for by the IDM. Students' grades are influenced by other factors, which are mentioned in the conceptual framework, such as amount of hours students spent studying (Ebbinghaus, 1913), the social cognitive ability of students (Bandura, 1999), and the factors that play a role in the theory of transactional distance (Moore, 1997). Furthermore, method of assessment, ability of instructor, and major of students can play an impact on students' grades.

5.3.6 Summary

Taking the mathematic discipline and ELLs in higher education into account, the results revealed that students in the online method scored best, followed by the hybrid method, and in a third position is the blended method. The face-to-face method was the lowest. The more online teaching took place, the better students' grades were. Technological and pedagogical developments have contributed to facilitate students' learning process of "linear thinking" by designing instruction in a structured and organized sequenced content on the learning management system.

The conceptual frameworks of the present study aid the view that by including online IDMs overlearning can take place, as described in the "Ebbinghaus' studies on memory" framework (Ebbinghaus, 1913). Overlearning is possible because students are provided content material at any time, and anywhere, as explained in the "Study at Own Pace: Anytime Anywhere Concept" framework. During the Covid-19 pandemic remote education has been a necessity. Learning takes place in societal experiences which was supported by the "Social Cognitive Theory" framework. The four IDMs include these social experiences. However, as online learning became the new way of teaching, students showed that they were able to handle a more self-regulated learning effectively regardless of the geographical separation of instructor and students, which pertains to the "Theory of Transactional Distance" framework. Distance seemed to affect the structural communication and students' autonomy positively.

Chapter 6: Conclusion

IDMs rapidly change as they are influenced by technological and pedagogical improvements. The UAE is a technological advanced country and students grew up using technology from a young age. While the blended instructional method was in place before the outbreak of Covid-19, students were forced in hybrid, and then completely online methods.

This aim of the study was to answer the following three research questions:

1. Are there differences in students' final grades among the face-to-face, blended, hybrid, and online IDMs?
2. If there is/are statistically significant difference(s) between the four methods, exactly which one(s) is/are?
3. What is/are the effect size(s) of the statistically significant different methods?

Similar studies have been performed before outside the UAE. The present research is new to the literature as it compared four IDMs for ELLs taking a higher education mathematics course in the UAE.

The findings of the study suggest that the online IDM outperforms the other three methods. Second best is the hybrid method, then the blended method and the face-to-face method is last. The more content is taught online, the better the students' grades get. The results encourage the implementation of the online method in mathematics higher education courses for ELLs in the UAE.

6.1 Limitations and Implications for Future Research

Attempts to generalize the results should compel the recognition of the context of the study. The sample consists of ELLs in the UAE whose first language is Arabic and are enrolled in a college mathematics course taught in English. Furthermore, the dependent variable, students' grades, is used to measure effectiveness of each IDM. Other unobserved variables such as students' gender, motivation, and retention rates, other disciplines could influence the effectiveness of the methods and causes a limited generalizability to other contexts. Limitations of the study imply that the generalization of the research's findings should be made cautiously. Therefore, future research is recommended including more variables to measure student performance in each method.

The study included the end-of-semester grades of 574 homogenous students enrolled in one faculty member's mathematics course. However, the instructor and students were not randomly selected to each IDM. Rather the HEI decided each semester which method had to be used. The faculties' technological and non-technological abilities to teaching in the different instructional methods might moderate the effect studied. Therefore, research with randomly assigned students and instructors is recommended for further studies. Controlling confounding variables, such as different instructors, or qualifications of instructors, or years of experiences, or major of students, age of students, undergraduate versus graduate level, size of classes, learning objectives, similar assessments, and gender, should be kept constant to ensure validity of the research (Cohen, Manion & Morrison, 2011).

Due to geographic and context-specific studies, further research could include other courses besides mathematics to examine the effect of IDMs on students' grades in other courses.

Additionally, the researcher recommends the incorporations of more variables, such as motivation, retention rates, different disciplines, and more instructors in further studies. Furthermore, a larger sample, across different HEI, and including more disciplines could investigate the same research questions to generalize the findings more.

6.2 Recommendations

The fact that the effect sizes increased from medium to large, and very large as the amount of online teaching increased, instructors and HEI should adapt their teaching to include more online technology and pedagogy. Certainly, during this time of the Covid-19 pandemic, these novel findings for ELLs are promising for the educational sector in the UAE.

The question that needs to be asked is if students score significantly higher in their online courses, what might be the cause of this. Do students feel that learning mathematics online develops linear thinking and problem-solving skills? Do instructors and students perceive students' skills are indeed increasing? Or is there something else happening? Are the exams too easy now that they are taken online from home? Are students able to cheat during exams now? Although various systems are in place to monitor students during exams, are students able to find alternative ways? Advancements in technology also means that students often find new ways to cheat the system. If that is the case, should exam questions be rephrased to more problem or project-based exams? Qualitative research is needed to analyze students and instructors' perspectives on why students perform better online.

On a last note, will students tire of taking all their classes online during the pandemic? Future research analyzing the effect of students' performance as the pandemic continues is recommended.

References

Abbott, M. (2011). *Understanding educational statistics using Microsoft Excel[a] and SPSS[a]*. New York: John Wiley & Sons.

Aker, L.B. (2016). *A Meta-Analysis of Middle School Science Engagement* (Doctoral dissertation, Seattle Pacific University).

Aldarayseh, Abdulla. (2020). The Impact of COVID-19 Pandemic on Modes of Teaching Science in UAE Schools. *Journal of Education and Practice*, vol 11, pp. 110. doi:10.7176/JEP/11-20-13.

Allen, E., Seaman, J., Poulin, R. & Straut, T. (2016). *Online Report Card: Tracking Online Education in the United States*. 13th annual report. Babson Survey Research Group. Babson Par, MA. [Accessed 8 November 2020]. Available at: https://eric.ed.gov/?id=ED572777.

Almansoori, S., & Akre, V. L. (2016). Roadmap for enhancing efficiency and effectiveness of Blended E-learning in Higher Education: A UAE Case Study. *International Journal of Education and Information Technologies*, vol. 10, pp. 176-185.

Annand, D. (2011). Article review—social presence within the community of inquiry framework. *International Review of Research in Open and Distance Learning*, vol. 12(5), pp. 40–56. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/924/1873.

Bahrick, H. P., Hall, L. K., Goggin, J. P., Bahrick, L. E., & Berger, S. A. (1994). Fifty years of language maintenance and language dominance in bilingual Hispanic immigrants. *Journal of Experimental Psychology: General*, vol. 123(3), pp. 264.

Baki, A., & Çakıroğlu, Ü. (2010). Learning objects in high school mathematics classrooms: implementation and evaluation. *Computers & Education,* vol. *55*(4), pp. 1459–1469 doi:10.1016/j.compedu.2010.06.009.

Bandura, A. (1999). Social cognitive theory of personality. *Handbook of personality*, vol. 2, pp. 154-96.

Bernard, R., Abrami, P., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., Wallet, P., Fiset, M. & Huang, B. (2004). How Does Distance Education Compare With Classroom Instruction? A Meta-Analysis of the Empirical Literature. *Review of Educational Research*, vol. 74 (3), pp. 379-439.

Bernard, R., Borokhovski, E, Schmid, R., Tamim, R., & Abrami, P. (2014). A meta-analysis of blended learning and technology use in higher education: From the general to the applied. *Journal of Computing in Higher Education*, vol. 26(1), pp. 87–122.
DOI: http://dx.doi.org/10.1007/s12528-013-9077-3

Best, S. (2012). *Understanding and Doing Successful Research: Data Collection and Analysis for the Social Sciences*. Taylor & Francis Group, London.

Bluman, A. (2007). *Elementary statistics: A step by Step Approach*. 6th edn. 2007: McGraw-Hill Higher Education.

Boneau, C. (1960). The effects of violations of assumptions underlying the t test. *Psychological Bulletin*, vol. 57(1), pp. 49-64.

Bonk, C. J., & Graham, C. R. (2012). *The handbook of blended learning: Global perspectives, local designs*. John Wiley & Sons.

Borba, M. C. (2012). Humans-with-media and continuing education for mathematics teachers in online environments. *ZDM–The International Journal on Mathematics Education*, vol. 44(6), pp. 801–814. doi:10.1007/s11858-012-0436-8.

Borenstein, M., Hedges, L. V., Higgins, J. P., and Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. Hoboken, NJ: Wiley.

Borgioli, G. M. (2008). Equity for English Language Learners in the Mathematics Classroom. *Teaching children mathematics*, vol. 15(3), pp. 185-191.

Burns, B. R. (2000). *Introduction to Research Methods,* 4th Edn. Frenchs Forest, Pearson education.

Castle, S. R., & McGuire, C. J. (2010). An analysis of student self-assessment of online, blended, and face-to-face learning environments: Implications for sustainable education delivery. *International Education Studies*, vol. 3(3), pp. 36-40.

Chan, Y. & Walmsley, R. (1997). Learning and Understanding the Kruskal-Wallis One-Way Analysis-of-Variance-by-Ranks Test for Differences Among Three or More Independent Groups. *Physical Therapy*, vol. 77 (12), pp. 1755-1761.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. New York, NY: Routledge Academic.

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

“Congress.gov”. (2015) Every Student Succeeds Act of 2015. (ESSA). (2015) 20 U.S.C. §6301 et seq. (2015) [online]. [Accessed 1 December 2020]. Available at: https://www.congress.gov/114/plaws/publ95/PLAW-114publ95.pdf

Darling-Aduana, J. and Heinrich, C.J., 2018. The role of teacher capacity and instructional practice in the integration of educational technology for emergent bilingual students. *Computers & Education*, vol. 126, pp. 417-432.

Driver, M. K., & Powell, S. R. (2017). Culturally and linguistically responsive schema intervention: Improving word problem solving for English language learners with mathematics difficulty. *Learning Disability Quarterly*, vol. 40(1), pp. 41-53.

Ebbinghaus, H. (1913). Memory: A contribution to experimental psychology. 1885. *New York: Teachers College, Columbia University*.

European Distance and E-Learning Network (EDEN). (2006). Theory and theorists: Michael G. Moore. Evolution of Theory of Transactional Distance. [online]. [Accessed 1 December 2020]. Available at: http://www.eden-online.org.

Falloon, G. (2011). Making the connection: Moore’s theory of transactional distance and its relevance to the use of a virtual classroom in postgraduate online teacher education. *Journal of Research on Technology in Education*, vol 43(3), pp.187-209.

Feir-Walsh, B. J. and Toothaker, L. E. (1974) ‘An Empirical Comparison of the Anova F-Test, Normal Scores Test and Kruskal-Wallis Test Under Violation of Assumptions’, *Educational and Psychological Measurement*, vol. 34(4), pp. 789–799. doi: 10.1177/001316447403400406.

Field, A. (2009). *Discovering statistics using SPSS*. 3rd edn. London:SAGE.

Fritz, C., Morris, P. & Richler, J. (2012). "Effect size estimates: Current use, calculations, and interpretation": Correction to Fritz et al. (2011). *Journal of Experimental Psychology: General*, vol. 141(1), pp. 30-30.

Fry, R. (2007). How Far behind in Math and Reading Are English Language Learners? Report. *Pew Hispanic Center*.

Garrison, D. R., Shale, D. (1987). Mapping the boundaries of distance education: Problems in defining the field. *The American Journal of Distance Education*, vol. 1(1), pp. 7-13.

Glantz, S., Slinker, B. & Neilands, T. (2016). *Primer of Applied Regression & Analysis of Variance*. 3rd edn. McGraw-Hill Education / Medical.

Graham, C. R. (2006). Blended learning systems. *The handbook of blended learning: Global perspectives, local designs*, vol. 1, pp. 3-21.

Gürsul, F., & Keser, H. (2009). The effects of online and face to face problem based learning environments in mathematics education on student's academic achievement. *Procedia-Social and Behavioral Sciences*, vol. 1(1), pp. 2817-2824.

Hayes, W. L. (1963). *Statistics for Psychologists*. New York, NY: Holt, Rinehart and Winston.

Hedges, L. V. (1981). Distribution theory for Glass's estimator of effect size and related estimators. *Journal of Educational Statistics*, vol. 6(2), pp. 107-128. doi:10.3102/10769986006002107.

Hrastinski, S. (2008). Asynchronous & Synchronous E-Learning. *Educause Quarterly*, vol. 4, pp. 51 – 55.

Huizingh, E. (2007). *Applied statistics with SPSS*. London: SAGE Publications, Ltd doi: 10.4135/9781446249390.

Hussein, E., Daoud, S., Alrabaiah, H., & Badawi, R. (2020). Exploring Undergraduate Students' Attitudes towards Emergency Online Learning during COVID-19: A Case from the UAE. *Children and Youth Services Review*.

Johnson, S. D., Aragon, S. R., & Shaik, N. (2000). Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments. *Journal of interactive learning research*, vol. 11(1), pp. 29-49.

Kersaint, G., Thompson, D. R., & Petkova, M. (2014). *Teaching mathematics to English language learners*. Routledge.

Kish, L. (1959) Some statistical problems in research design, American Sociological Review, vol. 24, pp. 328-338.

Kivunja, C. & Kuyini, A. (2017). Understanding and Applying Research Paradigms in Educational Contexts. *International Journal of Higher Education*, vol. 6 (5), pp. 26.

Kohr, R. & Games, P. (1974). Robustness of the Analysis of Variance, the Welch Procedure and a Box Procedure to Heterogeneous Variances. *The Journal of Experimental Education*, vol. 43 (1), pp. 61-69.

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, vol. 4, pp. 863.

Lix, L., Keselman, J. & Keselman, H. (1996). Consequences of Assumption Violations Revisited: A Quantitative Review of Alternatives to the One-Way Analysis of Variance "F" Test. *Review of Educational Research*, vol. 66 (4), pp. 579.

Mahmud, M.M., 2018. Technology and language–what works and what does not: A meta-analysis of blended learning research. *Journal of Asia TEFL*, vol. 15(2), pp. 365.

Margulieux, L., McCracken, W. & Catrambone, R. (2016). A taxonomy to define courses that mix face-to-face and online learning. *Educational Research Review*, vol. 19, pp. 104-118.

Martin, J. (2004). Self-regulated learning, social cognitive theory, and agency. *Educational psychologist*, vol. 39(2), pp. 135-145.

Maxwell, S. E., and Delaney, H. D. (2004). *Designing experiments and analyzing data: A model comparison perspective*, 2nd Edn. Mahwah, NJ: Erlbaum.

Mayadas, F., Miller, G. & Sener, J. (2015). "E-Learning Definitions". *OLC* [online]. [Accessed 7 November 2020]. Available at: https://onlinelearningconsortium.org/updated-e-learning-definitions-2/

Means, B., Toyama, Y., Murphy, R., Bakia, M., & Jones, K. (2010). Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning. *Center for Technology in Learning, US Department of Education.* Retrieved from http://www.ed.gov/about/offices/list/opepd/ppss/reports.html

McGrath, R. E., & Meyer, G. J. (2006). When effect sizes disagree: the case of r and d. *Psychological Methods* 11, 386–401. doi: 10.1037/1082-989X.11.4.386

Mgqwashu, E. (2000). University Learning. *Scrutiny2*, vol. 5 (2), pp. 63-66.

Moore, M. G. (1973). Toward a theory of independent learning and teaching. *Journal of Higher Education*, vol. 44, pp. 66-69.

Moore, M. (1997). Theory of transactional distance. In Keegan, D., ed. *Theoretical Principles of Distance Education*, Routledge, pp. 22-38.

Moore, M. G. (2013.). *Handbook of distance education*. Routledge.

Muijs, D. (2004). 'Validity, reliability and generalisability', in *Doing quantitative research in education with SPSS*, SAGE Publications, Ltd, London, pp. 64-84. doi: 10.4135/9781849209014.

Muir, T. (2017). Online, Anytime, Anywhere: Enacting Flipped Learning in Three Different Secondary Mathematics Classes. *Mathematics Education Research Group of Australasia*.

NCES. (2020). "National Center for Education Statistics (NCES) Home Page, a part of the U.S. Department of Education". *Nces.ed.gov* [online] [Accessed 11 November 2020]. Available at: https://nces.ed.gov/

Neumann, R., Parry, S., & Becher, T. (2002). Teaching and learning in their disciplinary contexts: A conceptual analysis. *Studies in Higher Education*, vol. 27(4), pp. 405–417.

Olejnik, S. & Algina, J. (2003). Generalized Eta and Omega Squared Statistics: Measures of Effect Size for Some Common Research Designs. *Psychological Methods*, vol. 8 (4), pp. 434-447.

Ortiz, B. I. L. (2006). *Online collaborative problem-based learning: Design, facilitation, student work strategies and supporting technologies*. Teachers College, Columbia University, pp. 1-376

Paul, J., & Jefferson, F. (2019). A Comparative Analysis of Student Performance in an Online vs. Face-to-Face Environmental Science Course from 2009 to 2016. *Frontiers Compute. Sci.*, vol. 1, pp. 7.

Peters, O. (1971). *Theoretical aspects of correspondence instruction.* In O. Mackenzie & E. L. Christensen Edition. The Changing World of Correspondence Study University Park, PA: Pennsylvania State University.

Pickvance, C. (2001). Four varieties of comparative analysis. *Journal of Housing and the Built Environment,* vol. 16 (1), pp. 7-28. Retrieved from http://www.jstor.org/stable/41107161

Posten, H. (1992). Robustness of the two - sample t - test under violations of the homogeneity of variance assumption, part ii. *Communications in Statistics - Theory and Methods,* vol. 21 (8), pp. 2169-2184.

Russell, T. L. (1999). *The No Significant Difference Phenomenon. As reported in 355 Research Reports, Summaries and Papers.* Chapel Hill, NC: Office of Instructional Telecommunications, University of North Carolina.

Rumble, G. (1986). *The planning and management of distance education*. London: Croom Helm.

Sawilowsky, S. (2003). A Different Future For Social And Behavioral Science Research. *Journal of Modern Applied Statistical Methods*, vol. 2 (1), pp. 128–132. doi:10.22237/jmasm/1051747860

Schmider, E., Ziegler, M., Danay, E., Beyer, L. & Bühner, M. (2010). Is It Really Robust?. *Methodology*, vol. 6 (4), pp. 147-151.

Schwartz, B. L. (2020). *Memory: Foundations and applications*. SAGE Publications, Incorporated.

Shea, P., & Bidjerano, T. (2017). Online learning in the 30 community colleges of the State University of New York: Differences in outcomes between classroom and online coursework. In Johnston (Ed.), Proceedings of EdMedia: World Conference on Educational Media and Technology 2017, pp. 1192–1198. Washington, DC: Association for the Advancement of Computing in Education (AACE). Retrieved from https://www.learntechlib.org/p/178436/.

Sim, J. & Reid, N. (1999). Statistical Inference by Confidence Intervals: Issues of Interpretation and Utilization. *Physical Therapy*, vol. 79 (2), pp. 186-195.

Smith, B. & Brame, C. (2020). "Blended and Online Learning". *Vanderbilt University* [online]. [Accessed 7 November 2020]. Available at: https://cft.vanderbilt.edu/guides-sub-pages/blended-and-online-learning/ .

Snart, J. A. (2010). *Hybrid Learning: The Perils and Promise of Blending Online and Face-to-Face Instruction in Higher Education: The Perils and Promise of Blending Online and Face-to-Face Instruction in Higher Education*. ABC-CLIO.

SPSS Tutorials. (n.d.). *SPSS One-Way ANOVA with Post Hoc Tests - Simple Tutorial*. [online]. [Accessed 31 August 2020]. Available at: https://www.spss-tutorials.com/spss-one-way-anova-with-post-hoc-tests-example/.

Spurrier, J. (2003). On the null distribution of the Kruskal–Wallis statistic. *Journal of Nonparametric Statistics*, vol. 15 (6), pp. 685-691.

Sternstein Ph. D, M. (2010). *Barron's AP Statistics*. 5th edn. New York: Barron's Educational Series.

Thompson, B. (2007). Effect sizes, confidence intervals, and confidence intervals for effect sizes. *Psychol. Sch*. 44, 423–432. doi: 10.1002/pits.20234.

Thomas, L. (2020). "Confounding Variables | Definition, Examples and Controls". *Scribbr* [online]. [Accessed 25 October 2020]. Available at: https://www.scribbr.com/methodology/confounding-variables/.

UAE Government. (2020). "Distance learning in times of COVID-19 - The Official Portal of the UAE Government". [Accessed 21 November 2020]. Available at: https://u.ae/en/information-and-services/education/distance-learning-in-times-of-covid-19.

Vargha, A. and Delaney, H. D. (1998) ‘The Kruskal-Wallis Test and Stochastic Homogeneity’, Journal of Educational and Behavioral Statistics, 23(2), pp. 170–192. doi: 10.3102/10769986023002170.

Verma, J. & Abdel-Salam, A. (2019). *Testing statistical assumptions in research*. John Wiley & Sons, Inc.

Vo, H.M., Zhu, C. and Diep, N.A., 2017. The effect of blended learning on student performance at course-level in higher education: A meta-analysis. *Studies in Educational Evaluation*, vol. 53, pp. 17-28.

Wavle, S. & Ozogul, G. (2019). Investigating the Impact of Online Classes on Undergraduate Degree Completion. *Online Learning*, vol. 23 (4), pp. 281-295. DOI: http://dx.doi.org/10.24059/olj.v23i4.1558.

Wedemeyer, C. A. (1977). Independent study. In A. S. Knowles (Ed.), *The International Encyclopedia of Higher Education Boston*: Northeastern University.

Wedemeyer, C. A. (1981). Learning at the back door: Reflections on non-traditional learning in the lifespan. Madison, WI: University of Wisconsin.

Whiting, J. (2017). Caught between the push and the pull: ELL teachers' perceptions of mainstreaming and ESOL classroom teaching. *NABE Journal of Research and Practice*, vol. 8(1), pp. 9-27.

Wigfield, A., Cambria, J., & Eccles, J. S., 2012. Motivation in education. *The Oxford handbook of human motivation*, pp. 463-478.

WWC. 2020. Find What Works! [Accessed 25 October 2020]. Available at: https://whatworks.ed.gov.

Xu, Z., Banerjee, M., Ramirez, G., Zhu, G. and Wijekumar, K., 2019. The effectiveness of educational technology applications on adult English language learners' writing quality: A meta-analysis. *Computer Assisted Language Learning*, vol 32(1-2), pp.132-162. https://doi.org/10.1080/09588221.2018.1501069.

Appendix A

A digital version of the excel file can be found here.

	N	Mean	Std. Deviation
F2F	175	1.42285714	1.59270711
Blended	256	2.14375000	1.49247787
Hybrid	119	2.59327731	1.31881631
Online	24	3.65416667	.36472642
Total	574	2.08031359	1.55477508

Cohen's (1988) Conventions	
0.2	Small
0.5	Medium
>0.8	Large

	F2F vs Blended	F2F vs Hybrid	F2F vs Online	Blended vs Hybrid	Blended vs Online	Hybrid vs Online
Cohen's *d*	0.470	0.787	1.486	0.312	1.054	0.873
Hedges's *g*	0.469	0.784	1.480	0.312	1.051	0.868

1. Group Statistics F2F vs Blended

	N	Mean	SD
Blended	256	2.1437500	1.4924779
F2F *(control group)*	175	1.4228571	1.5927071

Cohen's *d*	Hedges' *g*	Glass's Δ_2
0.470	**0.469**	**0.453**
		Uses SD from Group 2

2. Group Statistics F2F vs Hybrid

	N	Mean	SD
Hybrid	119	2.5932773	1.3188163
F2F *(control group)*	175	1.4228571	1.5927071

Cohen's *d*	Hedges' *g*	Glass's Δ_2
0.787	**0.784**	**0.735**
		Uses SD from Group 2

3. Group Statistics F2F vs Online

	N	Mean	SD
Online	24	3.6541667	.3647264
F2F *(control group)*	175	1.4228571	1.5927071

Cohen's *d*	Hedges' *g*	Glass's Δ_2
1.486	**1.480**	**1.401**
		Uses SD from Group 2

4. Group Statistics Blended vs Hybrid

	N	Mean	SD
Hybrid	119	2.5932773	1.3188163
Blended *(control group)*	256	2.1437500	1.4924779

Cohen's *d*	Hedges' *g*	Glass's Δ_2
0.312	**0.312**	**0.301**
		Uses SD from Group 2

5. Group Statistics Blended vs Online

	N	Mean	SD
Online	24	3.6541667	.3647264
Blended *(control group)*	256	2.1437500	1.4924779

Cohen's *d*	Hedges' *g*	Glass's Δ_2
1.054	**1.051**	**1.012**
		Uses SD from Group 2

6. Group Statistics Hybrid vs Online

	N	Mean	SD
Online	24	3.6541667	.3647264
Hybrid *(control group)*	119	2.5932773	1.3188163

Cohen's *d*	Hedges' *g*	Glass's Δ_2
0.873	**0.868**	**0.804**
		Uses SD from G2

Appendix B

Consent Form for using Students' Data

www.buid.ac.ae

September 30, 2020

To whom it may concern

This is to certify that Miss. Lisa Brashear with student ID number 20204663 is a registered part-time student on the Master of Education in Science Education programme at The British University in Dubai since January 2020.

She is required to gather data using surveys, interviews, & databases that will help her in writing the researches in her two remaining modules and the final research.
She will make sure that:
The research is based on objective and not subjective facts.
The institution shall not be mentioned, instead the name should be "Higher Education Institute in the UAE".
Values, faculty/staff and students at your institution are protected at all times. The studies do not evaluate your institution and its operations at all times.
Principles of research ethics, such as consent and confidentiality, are followed.

This letter is issued on student`s request.
Yours sincerely,

Amer Alaya

Head of Student Administration

PO Box 345015 • Block 11 Dubai International Academic City Dubai U A E • T +971 4 279 1400 • F +971 4 279 1490
FB.com/BUiD.Team BUiD_Team youtube.com/BUiDadmin @BUiD_Team BUiD