PREDICTING STUDENT SUCCESS USING DIGITAL TEXTBOOK ANALYTICS IN ONLINE COURSES

by

Dustin Lee Williams

Liberty University

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

Liberty University

2019

PREDICTING STUDENT SUCCESS USING DIGITAL TEXTBOOK ANALYTICS

IN ONLINE COURSES

by Dustin Lee Williams

A Dissertation Presented in Partial Fulfillment

Of the Requirements for the Degree

Doctor of Education

Liberty University, Lynchburg, VA

2019

APPROVED BY:

Michael Shenkle, Ed.D., Committee Chair

Scott Watson, Ph.D., Committee Member

ABSTRACT

In the digital era, students are generating and institutions are collecting more data than ever before. With the constant change in technology, new data points are being created. Digital textbooks are becoming more popular, and textbook publishers are shifting more of their efforts to creating digital content. This shift creates new data points that have the potential to show how students are engaging with course material. The purpose of this correlational study is to determine if digital textbook usage data, pages read, number of days, reading sessions, highlights, bookmarks, notes, searches, downloads and prints can predict student success. This study used a multiple regression to determine if digital textbook usage data is a predictor of course or quiz success in five online undergraduate courses at a private liberal arts university. The analysis used digital textbook data from VitalSource and consisted of 1,602 students that were enrolled in an eight-week online course at a private liberal arts university. The analysis showed that there is a significant relationship between digital textbook usage data and total points earned and average quiz grade. This study contributes to the limited knowledge on digital textbook analytics and provides valuable insight into how students engage with digital textbooks in online courses.

Table of Contents

ABSTRACT 3

List of Tables 7

List of Figures 8

List of Abbreviations 9

CHAPTER ONE: INTRODUCTION 10

Overview 10

Background 10

Historical Context 11

Theoretical Framework 12

Problem Statement 14

Purpose Statement 15

Significance of the Study 16

Research Questions 16

Definitions 17

CHAPTER TWO: LITERATURE REVIEW 18

Overview 18

Theoretical Framework 19

Student Engagement Theory 19

Self-Regulated Learning Theory 23

Related Theories 26

Related Literature 27

Learning Analytics 27

Textbooks..39

Summary 46

CHAPTER THREE: METHODS 48

Overview..48

Design 48

Research Questions..48

Null Hypotheses...49

Participants and Setting..49

Instrumentation 50

Procedures..50

Data Analysis...51

Assumption Testing 52

CHAPTER FOUR: FINDINGS..53

Overview..53

Research Questions..53

Null Hypotheses...53

Descriptive Statistics..54

Results..56

Null Hypothesis One..56

Null Hypothesis Two 60

CHAPTER FIVE: DISCUSSION...72

Overview..72

Discussion..72

Research Question One (Final Grade) 73

Research Question Two (Quiz Average) 77

Implications 80

Limitations 82

Recommendations for Future Research 83

REFERENCES 85

APPENDIX: IRB Exemption Letter 98

List of Tables

Table 1. Descriptive Statistics of Criterion and Predictor Variables 54

Table 2. Coefficients 56

Table 3. Model Summary 57

Table 4. ANOVA of Digitial Textbook Event Data and Overall Final Grade 57

Table 5. Coefficients of All Predictor Variables and Overall Points Earned..............................58

Table 6. Model Summary 60

Table 7. ANOVA of Digital Textbook Event Data and Average Quiz Score 60

Table 8. Coefficients of All Predictor Variables and Average Quiz Score.................................61

List of Figures

Figure 1. Final Grade Histogram 59

Figure 2. Scatterplot of Final Grades 59

Figure 3. Quiz Average Histogram 62

Figure 4. Scatterplot of Quiz Scores 62

Figure 5. Scotterplot of Reading Session & Final Grade 63

Figure 6. Scatterplot of Days Read & Final Grade 63

Figure 7. Scatterplot of Pages Read & Final Grade 64

Figure 8. Scatterplot of Highlights & Final Grade 64

Figure 9. Scatterplot of Notes & Final Grade 65

Figure 10. Scatterplot of Searches & Final Grade 65

Figure 11. Scatterplot of Print & Final Grade 66

Figure 12. Scatterplot of Bookmarks & Final Grade 66

Figure 13. Scatterplot of Downloads & Final Grade 67

Figure 14. Scatterplot of Reading Sessions & Quiz Average 67

Figure 15. Scatterplot of Days Read & Quiz Average 68

Figure 16. Scatterplot of Pages Read & Quiz Average. 68

Figure 17. Scatterplot of Highlights & Quiz Average 69

Figure 18. Scatterplot of Notes & Quiz Average 69

Figure 19. Scatterplot of Searches & Quiz Average 70

Figure 20. Scatterplot of Print & Quiz Average 70

Figure 21. Scatterplot of Bookmarks & Quiz Average 71

Figure 22. Scatterplot of Downloads & Quiz Average 71

List of Abbreviations

Analytics and Decision Support (ADS)

Customer Relationship Management System (CRM)

Educational Data Mining (EDM)

General Data Protection Regulation (GDPR)

Learning Management System (LMS)

Personalized Learning Paths (PLP)

Student Information System (SIS)

Variance Inflation Factors (VIF)

CHAPTER ONE: INTRODUCTION

Overview

Learning analytics is a growing trend in higher education; with the increase of data availability on students throughout their educational journey, there are constantly new data points becoming available for institutions to explore. Major publishers are shifting their strategies and offering universities access to a large amount of content, which is helping push the use of digital textbooks. The purpose of this correlational study seeks to identify if there are any relationships between student success and digital textbook usage. The following will review the background of learning analytics and digital textbook usage as well as discuss the problem, purpose, and significance of this study.

Background

The digital revolution has enabled institutions to collect information throughout the student lifecycle, from pre-admissions all the way through graduation. The popularity of online education has also allowed for a large amount of data to be collected on how students learn, and a large majority of this data can be pulled from the learning management system (Chaurasia & Rosin, 2017; Siemens, 2013). Data from the various systems are being used to help identify students who are at risk for not completing a course, better equip professors to understand how their students are learning, and increase retention. Textbook publishers are also seeing a shift in the demand for digital material: what was once a print-heavy industry has seen an increase in demand for digital material from students, professors, and institutions (deNoyelles & Raible, 2017; Duncan Selby, Carter, & Gage, 2014). The rise in popularity of digital textbooks gives faculty and institutions access to a new data point around how students are engaging with course material.

Historical Context

Learning analytics is a newer trend in higher education; it has become more popular with the increasing amount of learner data that is being collected on students. Since this concept is newer, the definition of learning analytics is evolving. However, most researchers reference the definition that was adopted at the First International Conference on Learning Analytics and Knowledge (LAK). LAK defines learning analytics as "the measurement, collection, analysis, and reporting of data about learners and their contexts, for the purpose of understanding and optimizing learning and the environments in which it occurs" (Clow, 2013, p. 685).

Learning analytics shares its roots in educational data mining, and both of these frameworks overlap in many areas. Educational data mining (EDM) can be defined as "developing, researching and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist" (Romero & Ventura, 2013, p. 12). Both of these research methodologies share a common interest in collecting, processing, and analyzing student data (Papamitsiou & Economides, 2014). These research methods also share a common interest in providing actionable insight to institutions to use in decision making that impact students, faculty, staff, and university administration (Papamitsiou & Economides, 2014).

There are several core difference between learning analytics and EDM: learning analytics tends to focus more on human judgment while EDM focuses on automation; learning analytics focuses on holistic systems where EDM focuses on individual components; learning analytics has origins in intelligent curriculum where EDM has roots in educational software; learning analytics focuses on empowering students and learners, compared to EDM which focuses on automation (Liñán & Juan Pérez, 2015; Romero & Ventura, 2013). In summary, learning analytics provides more actionable data around a learner and seeks to improve how learners learn. In comparison, EDM has a strong emphasis on refining and developing the tools and technologies around making data mining easier.

Digital textbooks are becoming more popular in higher education. A longitudinal study showed digital textbook usage rise from 42% in 2012 to 66% in 2016 (deNoyelles & Raible, 2017). One of the explanations for this could be the increase in professors or institutions requiring digital textbooks. In the same study, students reported an increase in digital material being required in their courses from 45% in 2012 to 55% in 2016. This study also found that students' preference for print textbooks decreased over this time period; in 2012, 38% of students stated that they preferred print textbooks; this number increased in 2014 to 42% and then took a sharp decrease to 17% in 2016 (deNoyelles & Raible, 2017).

Theoretical Framework

Engagement has long been associated with student success: research dates back to the 1970s when Tinto conducted a study on higher education dropouts (Tinto, 1975). Tinto developed a foundational theory that suggested that the more a student was engaged at the institution, the less likely that student would drop out (Tinto, 1975). His work was the cornerstone of the student engagement theory that has evolved over time. Tinto's model is based on interactions between the student, institution, academic and social systems. Students have both goal commitments, which consists of preference for a particular degree or career, and institutional commitments, which consist of financial, time, or family preference. As students progress through college, the integration into the social and academic systems impact the students' commitment level and therefore impact if they remain at the institution. A student can be integrated well into the academic systems of the institution and doing well in their courses but not integrated into the social systems of the college. This can impact their institutional commitments and therefore make a student at risk of attrition.

In 1984, Astin added to Tinto's research by proposing a student involvement theory that focuses on the deficiencies of three other pedagogical theories: subject-matter theory, resource theory, and individualized theory. Astin's theory shifted the focus to more on how the student is engaging and less on what the institution is doing, with a strong focus on the processes that help facilitate student development (Astin, 1999).

In the early 2000s, Arend identified that student engagement patterns were changing and that higher education institutions were not adjusting their strategies to meet the needs of their students (Arend, 2004). Technology started to play a more active role in the life of students, and higher education institutions could not keep up with the rapid change. Institutions were utilizing technology but not in the right way and could not adapt fast enough to accommodate the changes. Arend's (2004) study showed that students desired to engage with technology, but institutions were not meeting their needs.

This study is also viewed through the lenses of self-regulated learning theory. One of the core principles of this theory is the idea that the student is an active participant in the learning process (Zimmerman, 1986). The core principles of self-regulated learning can be found in Bandura's work on social learning theory. Bandura believed that learning can happen outside of direct experience and that one has the ability to self-regulate (Bandura, 1971). Piaget believed that an individual's progress through development stages and to progress through the stages requires awareness, interaction, and the ability to attempt to control objects in his or her environment (Fox & Riconscente, 2008). Vygotsky also believed in self-regulation; he believed that self-regulation happens through the control of attention, thoughts, and action (Wertsch &

Tulviste, 1992). Zimmerman has conducted numerous studies on self-regulated learning and discovered three characteristics of a self-regulated learner. Self-regulated learners tend to be metacognitively, motivationally, and behaviorally active participates in the learning process (Zimmerman, 1986). Metacognitive self-regulated learners are organized and evaluate their progress. Self-regulated learners are also motivated; they have goals and believe they are capable of learning. In regard to behavior, self-regulated learners have the ability to select the appropriate learning environment.

These two theories overlap in how they view engagement. Student engagement theory believes that the more the student is engaged, the higher the likelihood that the student will be successful. Self-regulated learning believes that students need to be active in the learning process. In alignment with these theories, this study seeks to determine if there is a correlation between how the student engages with digital textbooks and student success.

Problem Statement

Previous research about digital textbooks has primarily focused on two areas: students' and professors' preference of digital textbooks and the impact digital textbooks have on student performance in a course (deNoyelles & Seilhamer, 2013; Millar & Schrier, 2015; Rockinson-Szapkiw, Courduff, Carter, & Bennett, 2013; Weisberg, 2011). Research is clearly showing an increase in the popularity of digital textbooks for students along with the increase of professors adopting digital textbooks (Duncan Selby et al., 2014). In regards to student performance, research has mixed reviews on the impact digital textbooks have versus print textbooks (Fike, Fike, & St. Clair, 2016; Terpend, Gattiker, & Lowe, 2014). Some studies show students perform worse when using a digital textbook while other students show the opposite. Parallel to digital textbook growth, there is also a growing trend in higher education in the field of learning analytics. Institutions are investing more resources in collecting and analyzing data about how students learn (Clow, 2013). As the demand for digital textbooks rises and the interest in mining learner data increases, these two areas should intersect. However, there has only been one study that examines the engagement metadata of digital textbooks and the predictive value of the data. Previous research on digital textbook usage data analyzed seven data points: pages read, number of days read, reading sessions, time reading, highlights, bookmarks, and notes (Junco & Clem, 2015). The study concluded that digital textbook usage data are a significant predictor of course success. This study had several limitations: the population of the study was small, there was a lack of overall usage, and this study was conducted only in a traditional residential classroom setting. The problem is, there is a growing trend around collecting and analyzing learner data but there is a lack of research on how students are interacting with digital textbooks and the potential predictive value that this dataset has on student success in courses.

Purpose Statement

The purpose of this correlational study is to determine if digital textbook usage data, pages read, number of days, reading sessions, time reading, highlights, bookmarks, and notes can predict student success. This study measured student success by examining the total number of points a student earns in the course and the average exam score. There is a lack of research on how digital textbook usage data can be used in online courses as an early warning predictor. The information from this study may provide insight into how students are interacting with digital textbooks and determine if these metrics should be further explored in higher education early warning systems. Participants in this study are undergraduate students that are enrolled in online courses that offer a free digital textbook as part of tuition.

Significance of the Study

This aim of this correlational study is to add to the limited research on digital textbook analytics. Previous studies were limited in size and scope (Junco & Clem, 2015). This study added to the current research by having a larger sample size and by focusing on courses that are taught in predominately online environments.

Another limitation of the study conducted by Junco and Clem (2015) was the fact that students did not engage with the digital textbook. In the previous study, the digital textbook was optional for the student to use. This study used courses where students were given a free digital textbook that was accessed through the learning management system (LMS). This study seeks to increase insight into how online students are interacting with digital textbooks.

This study seeks to explore the relationship between digital textbook usage and test success. With the lack of textbook usage, digital and print, professors look for strategies to increase reading; this is typically in the form of quizzes (Harrison, 2016). This study seeks to see if digital textbook usage is correlated to student success on exams.

Overall, this study seeks to add to the limited knowledge of the predictive nature of the metadata that is being generated from digital textbooks usage. Universities are building at-risk models in the hope of helping students succeed in courses and persist to graduation (McGuire, 2018). If a strong correlation is present, this data may be beneficial to add to university early warning models.

Research Questions

RQ1: Can digital textbook usage data predict course success in an undergraduate online course at a private liberal arts university?

RQ2: Can digital textbook usage data predict test success in an undergraduate online course at a private liberal arts university?

Definitions

1. *Learning Analytics* – An analytical framework that focuses on data generated by learners, with the aim of understanding and enhancing learning (Clow, 2013).
2. *Educational Data Mining*- A framework that focuses on applying computerized methods to analyze large amounts of educational data that would be difficult to calculate manually (Romero & Ventura, 2013).
3. *Digital Textbook* – Texts that are offered digitally and can be accessed on digital devices (Rockinson-Szapkiw et al., 2013).
4. *Early Warning System* – An automated system that uses institutional data to calculate the risk of a student completing a course or remaining enrolled at the university (Jokhan, Sharma, & Singh, 2018).
5. *Learning Management System* - An online system where learners can go to access course content, interact in discussions, and take assessments (Chaw & Tang, 2018).

CHAPTER TWO: LITERATURE REVIEW

Overview

One of the by-products of a digital society is the massive amount of data that is being produced from using technology. Ninety percent of the data that is in existence was created in the past two years (DOMO, 2017). In a given day, it is estimated that 2.5 quintillion bytes of data are created. It is estimated that by the year 2020, that 1.7MB of data will be generated per second for each person on earth (DOMO, 2018). Students are generating more and more data as higher education institutions are capitalizing on the benefits of data analytics. Higher education institutions are collecting data from their infrastructure, networks, servers, applications, learning management systems, and other ancillary systems (Chaurasia & Rosin, 2017; Siemens, 2013). Most institutions are collecting data throughout the educational lifecycle of a student, from initial application through graduation. Institutions are leveraging the large amounts of data for four primary reasons: reporting and compliance; analysis and visualization; security and risk mitigation; and predictive analytics (Chaurasia & Rosin, 2017). Learning analytics is a growing field in higher education and is used heavily in higher education predictive analytics (Ben, 2015). Institutions are using this data to help identify at-risk students. Purdue signals was a popular retention initiative that utilized data about students' past and current performance to predict success in courses as well as retention at the school (Pistilli & Arnold, 2010).

Recently, there has been a developing theme from three major educational publishers—McGraw Hill, Cengage, and Pearson—called Inclusive Access (McKenzie, 2017). This new textbook model enables institutions to provide digital textbook and content to students inside of their course and make it available to them on the first day of the course (McKenzie, 2017). Inclusive Access also provides easy access to course materials through the learning management system, reduced cost of materials, and the flexibility to access content on mobile devices (McGraw-Hill Education, n.d.). In addition to the above benefits of Inclusive Access, this opens up another potential data point for institutions to explore: digital textbook analytics.

This literature review has two major sections: theoretical framework and related literature. The first section of this literature review will focus on the two theoretical frameworks of the study: student engagement theory and self-regulated learning theory. The second section will examine related literature on learning analytics in higher education, with a specific focus on learning analytics and student success, as well as literature on the evolution of textbooks in higher education.

Theoretical Framework

There are two theoretical frameworks that were used as a basis for the study: student engagement theory and self-regulated learning theory. The following section will review the development of these theories.

Student Engagement Theory

The study is viewed through the lens of the student engagement theory. Early research on this theory focused on student engagement in relation to retention. The core of the theory hypothesizes that the more the student is engaged with the course, the higher the likelihood that the student will be retained. As the theory has advanced, other researchers have found additional correlations between student engagement and student success metrics. These metrics include increased critical thinking, skill transferability, increased self-esteem, moral and ethical development, student satisfaction, improved grades, and persistence (Badura, Millard, Peluso, & Ortman, 2000; Gellin, 2003; Kuh, 1995; Trowler, 2010). Digital textbooks give the student the ability to engage with the text. Therefore, in alignment with this theory, the more a student engages with the textbook, the more likely his or her success in the course.

Tinto. Tinto's theory of integration is one of the founding theories on engagement. This theory was influenced by previous work completed by Durkheim and Spady. Durkheim's work was conducted on suicide, and Spady applied the theory to student drop out. Durkheim's theory concluded that suicide occurs more when individuals are not connected to society (Tinto, 1975). Tinto leveraged the work of Spady and Durkheim along with other studies around student dropout to create a model that shows the interactions that influence student dropout.

Tinto's model of dropout is based on interactions between the individual, academic, and social systems of the institution (Tinto, 1975). Students enter an institution with varying backgrounds, attributes, and experiences. These attributes include gender, social economic status, family background, and educational experiences. The student's experiences and background feed their commitments. Goal commitment is related to the student's educational goals; an example of this is a student's preference for a two-year degree or four-year degree. Ultimately goal commitment is the student's commitment to complete college. Institutional commitment is related to the student's preferences that would influence the decision to attend a specific institution. This could include financial commitments, time commitments, or family preference. The student's experience, the interaction between these commitments, determines if a student drops out from the institution or persists until graduation. As a student progresses through the college, it is the level of integration into the social and academic systems that impacts the level of commitment. Low goal commitment or low institutional commitment can lead to a student dropping out of the institution.

Astin. In 1984, Astin presented an involvement theory that is based on five assumptions. The first assumption defines involvement as the investment of physical and psychological energy in objects. The objects can be broad or specific (Astin, 1999). Secondly, involvement occurs along a continuum (Astin, 1999). Third, the involvement has both quantitative and qualitative properties. Fourth, student learning and development is proportionally related to the quality and quantity of involvement (Astin, 1999). Lastly, educational policy effectiveness is related to how much that policy can increase student involvement (Astin, 1999).

Astin's theory is founded on the deficiencies of three pedagogical theories: subject-matter theory, resource theory, and individualized (eclectic) theory. The subject matter theory is also referred to as the content theory. The foundation of this theory is based on the knowledge of the professor and the content of the course. The weakness of this theory is founded in the passive role of the student. The emphasis is on the content and the professor and not on the student (Astin, 1999).

The resource theory is focused on building or acquiring high-quality resources in the hope that these resources will enhance student learning. Resources include physical buildings, equipment, and personnel. One of the weaknesses of this is the limitation of these resources. The other problem with this theory is it focuses on the execution of the resources. Institutions are focused on acquiring resources but do not spend any effort on measuring effectiveness (Astin, 1999).

The individualized (eclectic) theory is the opposite of the content theory. The core belief of this theory revolves around student customization. Students should have their choice of electives as well as the pace of the instruction. Beyond instruction, students require individualized support from offices around campus. The biggest limitation of this study is the cost associated with producing the individualized experience for every student (Astin, 1999).

The theory of student involvement adds a new layer to these three previous theories by shifting the role of the student. In this theory, the student plays an active role in the learning process; the focus is shifted away from what the institution is doing to what the student is doing and is more concerned with the processes that help facilitate student development (Astin, 1999). Educators need to shift the focus on what they are doing to focus on what the student is doing. If educators are only focused on the textbooks and academic resources, learning may not occur as well as if the educator focused on how to get the student involved. Educators and academic administrators also need to understand that students have a finite amount of time to spend on academic activity. Students have to split their time between their studies, work, and social life. Each policy or decision that academic institutions make can impact the amount of time students have to devote to their studies.

The theory of student involvement is based on a longitudinal study of college dropouts that sought to identify factors that impacted student persistence. The conclusion of the study found that the factors that impacted persistence the most tied back to involvement. Students who persisted had factors that showed involvement, whereas students who did not persist had factors that showed a lack of involvement (Astin, 1999). Astin's work aligns with the findings of Tinto. However, Astin provided some practical applications for faculty, administrators, and counselors. He challenged faculty to continue to focus on what the students are doing and where they are spending their time. Similar to the recommendation for faculty, Astin encourages counselors and advisors to find where students are spending their time. He proposed that advisors ask struggling students to keep a diary of their activities, so that the advisor can determine if the student's struggle is related to time management issues, study habits, or motivation (Astin, 1999).

Arend. In the early 2000s, Arend noticed that engagement patterns were changing with the increase in computers on campuses (Arend, 2004). Arend (2004) stated, "Patterns of engagement are changing due to computers, yet many institutional services are barely keeping up with high student expectations for technology, let alone capitalizing on the learning opportunities inherent in the technology" (p. 30). Institutions were using technology but only to help automate normal tasks; there was a lack of innovative use of technology among faculty. The study showed that students had a willingness to engage with technology throughout their education, and institutions were not adapting to the new level of engagement (Arend, 2004). Arend noticed a shift in how students engage with institutions in light of technological advances. As technology advances and as new avenues of engagements are created, it is important for higher education institutions to account of these new methods and research the potential impact this has on learning and student involvement.

In summary, student engagement theory is the primary viewpoint for this research. As technology has advanced, it has increased the way in which students engage in their learning process. In alignment with the student engagement theory, it is hypothesized that the more a student engages with the course material (i.e., digital textbook), the more likely the student will be successful in the course.

Self-Regulated Learning Theory

This study will also use the perspectives that are found in self-regulated learning theory. This theory considers students as active participants in the learning process (Zimmerman, 1986).

Pairing this theory with student engagement theory, the more active students are in the learning process, the more engaged they will be with their learning.

Bandura. The core principles of self-regulated learning can be traced back to the social learning theory. Many traditional theories of learning believed that learning happened through direct experience (Bandura, 1971). Bandura's theory proposed that learning can happen outside of direct experiences and can happen by observation. Bandura believed that cognitive capacity allows humans to mentally solve problems without requiring to experience all of the alternatives, which enables them to mentally process and see the potential consequences and use this information to inform their decisions. Essentially, Bandura believed that reinforcement can happen by perceiving. Bandura's theory also relied on the ability of one to self-regulate. He proposed that individuals have the capacity to manage stimulus determinants, which enables them to influence their own behavior.

Vygotsky and Piaget. Vygotsky's and Piaget's theories are foundational theories of constructivism. Even though there are some differences between their theories, they share some common viewpoints regarding self-regulation. In Piaget's work, he believed that progression through developmental stages required awareness, interaction, and the ability to attempt to control objects and others in their environment (Fox & Riconscente, 2008). In Vygotsky's work, self-regulation happens through control of attention, thoughts, and action:

> At the higher developmental stages of nature, humans master their own behavior; they subordinate their own responses to their own control. Just as they subordinate the external forces of nature, they master personal behavioral processes on the basis of the natural laws of this behavior. Since the laws of stimulus-response connections are the basis of natural behavioral laws; it is impossible to control a response before controlling

> the stimulus. Consequently, the key to the child's control of his/her behavior lies in mastering the system of stimuli. (Wertsch & Tulviste, 1992, pp. 175-176)

Both of these theorists support the idea that an individual needs to take an active role in the learning process (Phillips, 1995). The works of these authors lay the foundation for the constructivist viewpoint. Constructivists believe that learners are active in the learning process; this aligns with both student engagement theory and self-regulated learning. The more students engage and regulate their learning in order to meet their goals, the more likely they will be successful in their education and remain enrolled at the university.

Zimmerman. Zimmerman conducted numerous studies on self-regulated learning and focused on three key areas: metacognition, motivation, and behavior (Zimmerman, 1986). Student achievement was historically viewed in terms of the quality of teaching or the natural ability of the student. Self-regulated learning focuses on how students actively engage in their learning process by activating, adjusting, and maintaining their learning strategies in specific contexts (Zimmerman, 1986). Zimmerman describes self-regulated learners as being metacognitively, motivationally, behaviorally active participants in the learning process. Applying these concepts to self-regulated learning, metacognitively self-regulated learners organize, evaluate, self-teach, and monitor throughout the learning process (Zimmerman, 1986). In terms of motivation, self-regulated learners see themselves as capable, effective, and autonomous. In terms of behavior, self-regulated learners are able to select and create the appropriate learning environment (Zimmerman, 1986). All learners have been found to use regulatory processes to some extent. However, self-regulated learners are aware of the relationship between the process and the learning outcomes and intentionally use strategies to meet their academic goals (Zimmerman, 1990). Another key characteristic of a self-regulated learner is the feedback loop. During this process, learners review how well their learning methods are performing and make necessary adjustments (Zimmerman, 1990).

Related Theories

The study will also be examined in the perspectives of e-learning theory. This theory looks at how students process multimedia information and suggests a framework for how multimedia should be designed (Mayer, 1997). As digital textbooks and media continue to grow, it is important to understand basic principles of how digital curriculum should be designed.

Mayer. Mayer proposed a theory of multimedia learning that has its roots in generative theory as well as dual coding theory (Mayer, 1997). Mayer takes three elements from generative theory: "meaningful learning occurs when learners select relevant information from what is presented, organize the pieces of information into coherent mental representation and integrate the newly constructed representation with others" (Mayer, 1997, p. 4). From the dual coding theory, Mayer takes the idea that information processing has a visual and a verbal system. The theory of multimedia learning starts with analyzing the relevant text and illustrations that are presented. The key part in this step is recognizing which of the information is relevant. This part of the process is derived from the dual-coding theory. After the selection of relevant text and images, the next step is organization. In this process, the learner organizes the text information into a verbal-based model and the images into a visually-based model (Mayer, 1997). The final step in this theory is when the learner integrates the information. In this process, the leaner will build relationships with existing knowledge as well as associate the text and images with each other.

Related Literature

The following section focuses on two main topics: learning analytics in higher education, with a particular focus on student success, and history and use of textbooks in education. The literature will also examine research that was conducted for online education.

Learning Analytics

Learning analytics is a newer trend in higher education and has roots in several branches of analytical thought. Institutions are collecting information on their students from the time of application all the way through graduation, with a large portion of this data coming from the learning management system. Learning analytics is made possible through this collection of large amounts of data, commonly referred to as Big Data (Clow, 2013; Duval & Verbert, 2012). Learning analytics is still in its infancy, but the most common definition, and the one adopted by the First International Conference on Learning Analytics and Knowledge, defines learning analytics as "the measurement, collection, analysis and reporting of data about learners and their contexts, for the purposes of understanding and optimizing learning and the environments in which it occurs" (Clow, 2013, p. 685). The following will review the history of learning analytics along with how learning analytics is being used in higher education to improve student success.

"Big data." Before diving into learning analytics, there needs to be an understanding of what big data is, the benefits of big data, and how other industries are leveraging it to be successful. Big data is a relatively new concept, and the definition continues to evolve. One of the original definitions of learning analytics defined big data as datasets that are too large to be captured, managed, and processed by a general computer (Chen, Mao, & Liu, 2014). In 2011, the International Data Corporation defined big data as " big data technologies describe a new generation of technologies and architectures, designed to economically extract value from very large volumes of a wide variety of data, by enabling the high-velocity capture, discovery and/or analysis” (Klašnja-Milićević, Ivanović, & Budimac, 2017, p. 1067).

Gartner, Microsoft, and IBM typically classify big data in regard to the three Vs: volume, velocity, and variety (Chen & Zhang, 2014). Volume is related to the collection of the data from various sources (Big Data History and Current Considerations, n.d.; Bond, 2018; Chen et al., 2014). Velocity is related to the speed at which the data is generated and needs to be analyzed. Variety deals with the various types of data and the structure of the data, which can be classified as structured or unstructured data (Bond, 2018). For some, the definition of big data has expanded to include four or five Vs; these Vs vary and can stand for value, variability, veracity, or virtual (Chen et al., 2014; Chen & Zhang, 2014; Special Issue on Educational Big Data and Learning Analytics, 2018).

Big data directly impacts all major industries, including business/retail, healthcare, government, and education (Chen & Zhang, 2014). These industries are leveraging big data to increase efficiency, become more competitive, and create better customer experiences. All of these have a direct impact on the financial bottom line of a company. It is estimated that big data can save the US healthcare system 300 billion dollars, increase retail profits upwards of 60%, and make government more efficient (Chen et al., 2014). Companies are also utilizing big data to help in the recruitment of top employees, reduce turnover, mine social media, and perform employee assessment and feedback quickly (Tonidandel, King, & Cortina, 2018).

Higher education is starting to see the benefits of big data; the emergence of several disciplines like learning analytics and educational data mining are becoming more prominent. The increased demand for online learning has paved the way for large datasets being generated by the student in the learning management system (Klašnja-Milićević et al., 2017). Similarly, the increased demand for massive online open courses (MOOC) has also generated larger learner datasets (Klašnja-Milićević et al., 2017). Big data has the potential to help higher education in several key areas: improving the learner experience, improving knowledge, institutional decision making, recognizing and understanding global trends, and converting unstructured data into actionable insights (Klašnja-Milićević et al., 2017).

Big data and analytics intersect and make each other more valuable. One of the primary goals that organizations or institutions have with their data is to summarize it into actionable insights that can help further the company (Ben, 2015). Big data technologies increase the value of educational data mining, academic analytics, and learning analytics by allowing institutions to analyze large quantitates of data (Ben, 2015). Learning analytics benefits from big data by allowing institutions to mine the large amounts of data that are being generated by the learner through the LMS and other institutional systems (Ben, 2015). An example of where the two meet is the growing trend of adaptive learning in higher education. In order for adaptive learning to be successful, careful mining of the learners' paths through the learning management system as well as other integrating data from other systems is needed in order to achieve personalized adaptive learning (Liu, McKelroy, Corliss, & Carrigan, 2017).

Big data is helping shape the future of adaptive learning by being able to help develop personalized learning paths (PLP; Birjali, Beni-Hssane, & Erritali, 2018; Liu, Kang, et al., 2017). PLPs seeks to find the best teaching methodology by evaluating the learner's skills and providing recommendations of specific learning objects that are at the learner's knowledge level and also hiding learning objects that the student has already mastered or does not fit his or her learning style (Kurilovas, Zilinskiene, & Dagiene, 2015). Research has shown that personalized learning is effective in helping learning efficiency and achievement (Essa, 2016; Kurilovas et al., 2015; Simon-Campbell, Loria, & Phelan, 2016). Current PLPs lack the ability to leverage big data; Essa (2016) believes that the future PLPs and adaptive learning environments will need to be able to leverage big data in order to evaluate the large amount of data and produce the needed just-in-time notifications (Kurilovas et al., 2015). Discussed later in more detail, the future of digital textbooks lies in advancing and leveraging personalized learning and adaptive technologies (Sun, Norman, & Abdourazakou, 2018). The use of big data will be critical to the future advancement of personalized learning and digital textbooks.

Learning analytic process. The process of how learning analytics should be used and the framework of the field is constantly changing. However, Campbell and Oblinger's (2007) five-step process for learning analytics has been adopted in several studies (Junco & Clem, 2015; Weisberg, 2011). The five steps of analytics consist of the following: capture, report, predict, act, and refine (Campbell & Oblinger, 2007). The process of capturing a dataset is centered on knowing where the data is being generated from, understanding how it flows through the system(s), and knowing where it is stored (system of record). Some examples could include demographic data from the student information system (SIS), interaction data from the customer relationship management system (CRM), or academic data from the LMS. Decisions on storage, granularity, and data retention need to be determined in this part of the process (Campbell & Oblinger, 2007). After the data have been collected and stored in a centralized place, the next step is to create reporting based on that data. This type of reporting is typically descriptive in nature, looking for trends, and making simple correlational analyses. Frequently, the data are displayed in dashboards. The next stage of the process is data prediction. In this stage, more complex modeling occurs that combines data from all areas. Institutions will create models, test reliability, and determine the frequency with which the model needs to be refreshed (Campbell & Oblinger, 2007). The fourth state is the act or intervention stage, where data is used to recommend interventions to students and empower them to take action (Campbell & Oblinger, 2007). Students could be provided with data on the best course sequencing based on their degree and similarities with previous peers. Other interventions can be more direct; faculty members may receive at-risk notifications and then reach out to students through email, phone, or meeting request. It is important during this stage to determine the appropriate interventions and measure the success of the interventions (Campbell & Oblinger, 2007). The final step of the process is to monitor the impact of the analytics projects and determine how frequently the model needs to be reviewed and updated.

Similar to other aspects of learning analytics, the framework of learning analytics is still developing. Greller and Drachsler (2012) developed a learning analytics framework that has six core components: stakeholders, objective, data, instruments, external limitations, and internal limitations. In this model, each of the six dimensions has dependencies on the others and, therefore, they all need to exist for the model to function correctly. A newer framework developed by Ifenthaler and Widanapathirana (2014) uses support vector machines (learning algorithms) to fill in some of the gaps from the previous model from Greller and Drachsler. Ifenthaler and Widanapathirana (2014) proposed a framework that:

> combines data directly linked stakeholder, their interaction with the social web and the online learning environment as well as curricular requirements. Additionally, data from outside of the educational system is integrated. The processing and analysis of the combined data is carried out in a multilayer data warehouse and returned to the stakeholders, governance or institution in a meaningful way. (p. 223).

Educational Data Mining. Educational data mining (EDM) predates learning analytics by a few years. EDM started formal meetings back in 2005 and had its first conference in 2008 (Siemens, 2013). EDM draws from three major content areas: education, statistics, and computer sciences (Romero & Ventura, 2013). EDM can be defined as "developing, researching, and applying computerized methods to detect patterns in large collections of educational data that would otherwise be hard or impossible to analyze due to the enormous volume of data within which they exist" (Romero & Ventura, 2013, p. 12). EDM is a cluster of several different interdisciplinary areas that include but are not limited to "information retrieval, recommender systems, visual data analytics, domain-driven data mining, social network analysis, psychopedagogy, cognitive psychology and psychometrics" (Romero & Ventura, 2013, pp. 12-13).

The research framework of learning analytics and educational data mining overlap in several areas. However, each of these groups takes a slightly different approach to research. In terms of research discovery, both groups use both automation and human judgment to conduct research, but educational data mining puts primary focus on automation while learning analytics has a stronger focus on human judgment (Romero & Ventura, 2013; Siemens, 2013). Learning analytics research tends to focus holistically on systems, whereas educational data mining tends to analyze relationships between individual components. The origins of learning analytics have roots in intelligent curriculum, course outcome predictions, and interventions. Educational data mining has similar roots in outcome prediction but also has roots in educational software and student modeling (Romero & Ventura, 2013; Siemens, 2013). In alignment with their background, the outcomes of learning analytics research focus on enabling students and instructors with necessary information. On the other hand, educational data mining tends to focus more on automation without the need for human interaction. When comparing the techniques and methods of the two groups, learning analytics tends to place emphasis on various forms of analysis: sentiment, discourse, and concept, as well as learner success predictions. Educational data mining places importance on classification, modeling, determining relationships and visualizations (Romero & Ventura, 2013; Siemens, 2013).

Student success. One of the primary goals of using learning analytics is to provide institutions, faculty, and students data to increase learning and student success. Institutions are being scrutinized by government agencies and accrediting bodies regarding poor graduation rates (McGuire, 2018). In order to combat lower graduation rates, institutions are turning to learning analytics to help them identify students who are at risk for not succeeding in their courses or not remaining enrolled at the institution (Scholes, 2016). This is especially important for online institutions where graduation and retentions rates of students are typically lower. Learning analytics can provide several benefits to institutions, including increased student performance and retention. Increased student performance and retention will help increase graduation rates and increase student retention which in turns helps the institution financially (Scholes, 2016).

Creating early warning systems is one strategy institutions use to target at-risk students. Some early warning systems use student activity in course assignments or in course activities, others use demographic and performance data, and others have used data obtained from the LMS (Hu, Lo, & Shih, 2014). One of the most predominant and early examples of using learning analytics to create an early warning dashboard was done at Purdue University. In this study, a model was created that identified at-risk students and then presented this data to the faculty and student in simple green, yellow, or red indicators (Arnold & Pistilli, 2012). The algorithm predicted students' risk in four different categories: performance, effort, academic history, and characteristics. Instructors who had at-risk students in their course would then implement an intervention strategy that might consist of a personal outreach by email, text, or personal meeting, an outreach by an academic advisor, or a notification inside of the LMS (Arnold & Pistilli, 2012).

When using learning analytics, the LMS collects a large number of data points on students which can be used to help measure engagement (Zacharis, 2015). A study was conducted to determine which activities inside of an LMS correlate with student grades, and which of the variables could be used to predict student success. In total, 29 variables were analyzed using a stepwise multiple regression. Four variables accounted for 52% of the variance in course grade: reading and posting messages, content creation, quiz efforts, and a total number of files viewed (Zacharis, 2015). The highest of these four was reading and posting messages.

Learning analytics dashboards. One of the major outputs of learning analytics is creating dashboards to visually display the results of the analysis to the end user in a condensed and easily understandable format. A dashboard is a way to condense important information into a single view for the end user to review (Aljohani et al., 2018; Few, 2006). Learning analytics dashboards have two primary audiences: students and professors. Institutions have mixed approaches as to how they set up their dashboards. Data are made available in three primary ways: shared with faculty, shared with students, or shared with both faculty and students (Park & Jo, 2015). There is a growing trend in higher education of institutions to start displaying learning analytics data to students through dashboards (Aljohani et al., 2018; Roberts, Howell, & Seaman, 2017).

In higher education, the learning analytics dashboards that are being built by institutions display a wide range of information. Typically, these dashboards are aggregating and displaying information around login information, performance results, message analytics, at-risk predictions, content usage, and social network analysis (Park & Jo, 2015). Not only does the input data vary across institutions, but the complexity of the analysis also differs across institutions. Several institutions are simply displaying descriptive statistics while others are using more complex algorithms and crossing over to predictive analytics (Park & Jo, 2015). In a study conducted by Aljohani et al. (2018), the researchers built a learning analytics dashboard that tracked student engagement by examining the interactions inside the learning management system. This particular dashboard compared the students' engagement against each other and also provided some limited textual feedback about how good or bad they were performing. The researchers found that the group that used the dashboard throughout the course was more engaged with the course materials and earned higher marks when compared to the control group that did not use the dashboard (Aljohani et al., 2018). This aligns with the principles of student engagement theory, in that the more a student engages, in this case, logs in and interacts with the learning management system, the higher likelihood that they will be successful.

Early alert systems. Institutions are using learning analytics data to help increase the accuracy and effectiveness of early alert systems. "Early alert systems offer institutions systematic approaches to identifying and intervening with students exhibiting at-risk behaviors" (Tampke, 2013, p. 1). Early alert systems take on many forms and make recommendations on different at-risk behaviors. For example, some early alert systems focus on identifying students that are at risk to fail the course; others focus on student attendance or even mine student behavior and recommend tutoring (Cai, Lewis, & Higdon, 2015; Tampke, 2013). A study conducted by Villano, Harrison, Lynch, and Chen (2018) reviewed the relationship between early alert systems and student retention. In this study, the researchers reviewed three categories of data: demographic variables, institutional variables, and student performance variables. Unique to other studies, the researchers monitored students' use of the library system. The results of the study found that the early alert system did a significant job in determining at-risk students in their first year of school. After being enrolled for 90 weeks, the system had a harder time identifying students who were at risk of discontinuing.

Ethical Considerations. Collecting and storing all of the required data to perform the proper analysis comes at a risk to the institution. There are some ethical considerations to take into account when conducting learning analytics. There are three broad ethical considerations that institutions need to be aware of: "location and interpretation of data; informed consent, privacy, and the de-identification of data; management, classification, and storage of data" (Slade & Prinsloo, 2013, p. 1511). Johnson (2017) believes that there are four major ethical issues in learning analytics: privacy, individuality, autonomy, and discrimination.

With the wide range of data collected and the various means of reporting and visualizing the data, institutions need to be cognizant of how the data is being interpreted (Slade & Prinsloo, 2013). It is easy to oversimplify, overgeneralize, or included biases in the reporting. Institutions are using learning analytics to identify students who are at risk of not persisting or failing a course, which if not implemented correctly has potential cause for discrimination (Scholes, 2016). There is also concern that the algorithms that are behind some of the models include bias, which has the potential to discriminate (Johnson, 2017). Machine learning models have to be trained, and if it is not trained properly, they can be predisposed to bias.

Data privacy is a growing concern in the United States and in Europe. In many cases, students might not be aware of the data being collected about them. Questions are starting to arise around getting consent and giving more information to students about how their data are being used (Slade & Prinsloo, 2013; Wintrup, 2017). Some institutions and governments are considering opt-out clauses, but there has been difficulty deciding specifically what data points students can opt out of (Wintrup, 2017). The European Union passed updates to the Data Protection Regulation 1998 in 2016 with a compliant date set for May 25, 2018. This updated policy called General Data Protection Regulation (GDPR) has three main objectives:

> Provide rules for the protection of the personal data of natural persons and the processing of their personal data; to protect the fundamental rights and freedoms of natural persons, particularly with regard to their personal data; and, to ensure that personal data can move freely within the European Union (Cornock, 2018, p. A1).

GDPR requires the following be listed in a privacy policy: understanding of what data are being collected, the reasons for the collection of that data, how the data will be processed, the timeframe of how long the data will be stored, and who is the designated person to contact to have data removed or copies sent to the data owner (Renaud & Shepherd, 2018). This policy applies to all data that individuals share as well as to all of the data that is collected about the individual as they interact with the company's systems. The requirement also states that the answer to the above questions has to be concise. Higher education institutions that offer services to students that are a part of the EU are also held to this policy (McKenzie, 2018). Institutions that have a large international presence or offer education online need to be familiar with these new regulations. Failure to comply with these regulations could result in large fines. One of the issues that higher education institutions will face is the "right to be forgotten" policy that is now part of GDPR. This new requirement has the potential for students to request to be forgotten, which will require the university to eliminate email records, remove the student's information from directories, and also remove the student's admission application (McKenzie, 2018).

A study conducted by West et al. (2016) found that most institutions are aware of the ethical issues that come with using learning analytics in the areas of autonomy, privacy, confidentiality, informed consent, and data ownership (West et al., 2016). Most of the respondents believe that learning analytics should follow the guidelines of research ethics. However, there is a disconnect between how universities understand the difference between consent and informed consent as it relates to the digital data being collected. Most universities do not have a process for a student to opt out of being reported; essentially the student either accepts the university's privacy policy or has to withdrawal from the university (West et al., 2016).

West et al. (2016) proposed a four-step process for ethical decision making: explore the issue, apply an institutional lens to the issue, view the alternative actions in light of the ethical, theoretical approaches, and document the decision made. Johnson (2017) created five questions for higher education institutions to consider when using learning analytics. What is the intent of the learning analytics model: to promote the student or the institution? Does the process of creating the system get buy-in from all parties that are impacted? Is there transparency with the calculations behind the data model? Is the data being used a valid representation of the question that is being solved? Is there a connection between the identified problem and intervention without oversimplifying or being prone to high-stake errors? In most institutions, the complexity of implementing learning analytics can make the process appear like a "black box" where no one really understands everything that is happening. Johnson (2017) encourages institutions to examine any learning analytics project and use the above questions as a guide.

In summary, learning analytics is a growing trend in higher education, and with the continued advances in technology, having the ability to analyze a large amount of data that is being generated by students is crucial for higher education institutions. The field of big data has the potential to keep growing and institutions have already shown interest in analyzing student data for reporting, visualizations, security and predictive analytics. Data is being used to save institutions money, increase student learning, and increase student retention. As more data become available, institutions must be intentional on how the data are being used and ensure that any system or process that is created meets ethical standards. Digital textbook data offer a newer learner data point and can provide insight into how students are utilizing course material and how this correlates with the success of the student in the course.

Textbooks

There is still some debate on which format of book, print versus digital, is better for student learning. Research is conflicted on how textbook format impacts student learning; some studies show that students learn better with print, while other studies show that students who use digital textbooks earn a higher grade (Rockinson-Szapkiw et al., 2013). However, publishers are pushing more and more content to digital resources and making more of their resources available electronically (Millar & Schrier, 2015; Mulholland & Bates, 2014; Rockinson-Szapkiw et al., 2013). One of the reasons for this could be tied to the wide availably of e-readers, tablets, and mobile devices (Dobler, 2015; Millar & Schrier, 2015).

The popularity of e-textbooks is on the rise. In some of the earlier studies students, preference of e-books exceeded 70% (Duncan Selby et al., 2014). A more recent study conducted by the Pearson Foundation in 2011 showed 55% of the participants preferred print textbooks compared to e-textbooks. The trend seems to continue, as a recent study conducted by deNoyelles and Raible (2017) shows e-textbook usage rise from 42% in 2012 to 60% in 2014 and rise again to 66% in 2016. DeNoyelles and Raible (2017) found that professors or institutions are making digital textbooks the preferred format. In 2012, 45% said that digital textbooks were required as a course component, in 2014 the number increased to 49%, and in 2016 it increased to 55%. In the same time period, the study found that the preference for print textbooks decreased. In 2012, 38% of the participants listed that they preferred print as a reason for not using an e-textbook. Interestingly enough, that number increased to 42% in 2014 and then dramatically decreased to 17% in 2016. The sale of e-books in 2011 increased by 117% with total sales of around 970 million. In the same year, e-textbooks increased by 44% (Millar & Schrier, 2015).

Adoption of digital textbooks varies by campus, and some universities are putting more emphasis on digital textbook initiatives (deNoyelles & Seilhamer, 2013). Universities that do not put effort into marketing or training students on proper digital textbook usage see a lower adoption rate. DeNoyelles and Seilhamer (2013) found that schools that do not have an active e-textbook initiative saw digital textbook usage around 45%.

Cost is one of the biggest factors that impact student textbook buying. In a study conducted by Rajiv Sunil and Jhangiani (2017), 27% of the participants had taken fewer courses, and 26% of the participants said they had not registered for a course due to the price of textbooks. Thirty-seven percent of the participants also reported that they had earned a low grade due to textbook cost.

Evolution of digital books. Digital books date back to the early 1980s and over the years have evolved into many formats (Subba Rao, 2001). Ebooks were originally written in plain text, meaning that they lacked textual format (no color, bolding, italics, etc.). In 1993, the first ebook was written in HTML; this provided the ability to tag text with specific formattings like color, bolding, and italics (Ebook Timeline, 2002). Initially, ebooks were available only on computers and were primarily distributed through floppy disks or CD-ROMs. In the early 1990s, a prime example of this was the digitization of encyclopedias that were sold on CDs (Ebook Timeline, 2002; Subba Rao, 2001). The growth of the internet made it easier for individuals to purchase ebooks; in 1993 BiblioBytes launched the first ebook website (Ebook Timeline, 2002). Ebook readers have changed how ebooks are being consumed by allowing them to be consumed on a portable device; this has played a major role in the growth of ebooks (Gibson & Gibb, 2011).

Similar to the evolution of ebooks, the digital textbook has evolved over time, and the types of digital textbooks that are being used in courses have changed. When publishers started developing digital textbooks, they created a static digital textbook, which is a scanned copy of the existing printed textbook. These types of digital textbooks were difficult to use and did not enhance learning (Dobler, 2015; Weng, Otanga, Weng, & Cox, 2018). Publishers started to add features that enabled the reader to engage with the content of the textbooks. Early on, this included providing interactive tables, figures, and hyperlinks. As technology advanced and the market demand increased, digital textbooks added the ability to take notes, have built-in assessments, and connect to other content (Dobler, 2015). There are two emerging trends within the digital textbook industry: collaborative digital textbooks and adaptive learning textbooks. Collaborative digital textbooks provide presentation aids, learning support, an ability to integrate content with curriculum outcomes, accessibility tools, and the ability for other instructors to collaborate and add content (Grönlund, Wiklund, & Böö, 2018). Adaptive learning textbooks track students' progress, and based on the performance on the assessments, adapt the content of the textbook to meet their needs (Sun et al., 2018); this occurs by monitoring what the student is reading and by providing assessments as the student progresses through the material. The student can answer the question and also indicate their confidence level. By keeping track of the student's progress, this enables the system to offer the student real-time feedback on his or her mastery and suggest areas of improvement. Professors can also gain insight into how the students in their class are performing and adapt their approach in real time. For the purpose of this study, the digital textbooks that are used in the courses provide ways for students to interact with the text by highlighting, taking notes, making bookmarks, and searching, but are not adaptive.

Market change. Publishers have been faced with a new market and are trying to adapt their business to compete in the digital age. One of the biggest competitions to publishers is the used textbook market, which is not a new issue but one with which they are still trying to compete. Used textbooks account for 35% of textbook sales (Reynolds, 2011). Shifting to digital textbooks and trying to increase their usage is one way to compete. A newer competition is the emergence of more textbook rental companies such as Chegg. These companies are making it easier for students to rent textbooks and are thus decreasing the number of new textbook sales (Reynolds, 2011). These companies are hurting the sales of new textbooks by offering print textbooks at a reduced cost that make them more appealing to students. New publishing companies are also starting to form. These companies are adopting a digital-first strategy where they develop content digitally and offer print-on-demand functionality to students. These companies have lower overhead and can offer digital material at a lower cost, as well as offer higher royalties to their authors. Existing publishers are creating opportunities for institutions to partner with them to get access to their entire digital library, which is being referred to as inclusive access (McKenzie, 2017). This strategy benefits publishers, institutions, and students. Publishers receive guaranteed sales in their courses, which removes the competition from both the rental companies and the used textbook market. Students receive their materials at a fraction of the cost, and faculty can now be certain that the students have access to the course content on the first day of the course (McKenzie, 2017).

Reading compliance. Student compliance with required readings has been a problem for years. There is a limited amount of research on students' participation in required readings and success in the course. Previous studies have mostly focused on students' interactions with print textbooks and have relied on self-reported data to determine how students are engaging with the textbook. Research conducted from 1991 to 1997 showed that on any given day around one-third of students completed the required reading (Berry, Cook, Hill, & Stevens, 2010; Burchfield & Sappington, 2000). Similar findings were found by Clump, Bauer, and Breadley (2004) who reported that 27% of psychology students completed assigned readings. Other self-reported studies found that 77% of students rated that they read the textbook "often" (> 75% of the time) or "sometimes" (25%–75% of the time; French et al., 2015). There are four primary reasons why students do not read assigned readings: not prepared, lack of motivation, time management issues, and not fully understanding the importance of reading the required material (Kerr & Frese, 2017).

Textbook reading in relation to student success in the course has mixed findings. Students who rated that they read the textbook "often" (>75% of the time) outperformed students that rated as reading the textbook "sometimes" (25%–75% of the time). However, students who rated themselves as reading the textbook "rarely" (<25% of the time) had similar scores to students who read the book "often" (French et al., 2015).

Educators are being encouraged to select curriculum and structure classes in a way that engages students more (Lieu, Wong, Asefirad, Shaffer, & Momsen, 2017). In order for students to be successful and prepared for the interactive content, they need to know the content ahead of time. Typically, instruction is lectured based, but in order to make classrooms and content more engaging, professors are assigning reading or video content to students before the course. Educators are using different strategies in order to encourage students to read the assigned material. These strategies include announced reading quizzes, unannounced reading quizzes, short writing assignments, journal entries, mandatory reading guides, optional reading guides, or being called on in class to answer questions regarding the reading (Hatteberg & Steffy, 2013; Lieu et al., 2017). Announced assessments on the reading seem to be more effective than other methods. Studies have also found the use of reading guides to be a successful strategy for increasing student reading. A study conducted by Lieu et al. (2017) found that 80% of the students completed the reading guide; Lieu et al. also found a correlation between completing the study guide and exam scores. Digital textbook publishers are also changing the way they deliver digital textbooks to make them more interactive and engaging. Publishers are embedding content and digital material inside of the textbook that will allow students to access their knowledge as they read. Discussed earlier in this chapter, publishers are starting to explore with creating digital platforms that adapt to the students' needs (Sun et al., 2018). These adaptive learning systems present only the information that the students need to know based on their previous performance. This study seeks to add to the body of knowledge on how students are engaging with textbooks but with the focus on digital textbooks. The second research question seeks to explore how students' engagement with digital textbooks impacts their quiz score.

Print versus digital. There are several reasons that digital textbooks are growing in popularity among students and faculty. A study conducted by Weisberg (2011) found four reasons why e-textbooks are becoming more popular: convenience, lower cost, functionality, and desirability for the current generation (Jang, Yi, & Shin, 2016). Alternatively, the results showed three reasons why students did not desire e-textbooks: easier concentration, better comprehension, and personal preference. Weisberg (2011) removed cost from the equation by offering student print and digital options, and 87% of the students chose the e-textbook.

Student satisfaction is an important part of digital textbook adoption. A recent survey conducted by Hao and Jackson (2014) measured overall student satisfaction as well as satisfaction on three dimensions: usability, learning, and features. The results showed that students had an overall moderately above-neutral positive attitude toward e-textbooks. Students were most satisfied with the usability of e-textbooks and ranked learning facilitation as the least. Digital textbook usage has also seemed to help increase student motivation (Jang et al., 2016). The student's expectation on how the digital textbooks are to perform as well as its actual performance were also associated with student satisfaction (Philip & Moon, 2013). The actual performance constituted 61% of the variance compared to expectation which accounted for 11.2% and disconfirmation 9.5%. One of the selling points of digital textbooks is the extra features that these books offer. Some of these features include highlighting, note-taking, annotations, and self-exanimation questions (Van Horne, Russell, & Schuh, 2016).

Print versus digital performance. Studies are mixed in their findings between student performance using electronic textbooks versus a traditional hard-copy textbook. A recent study conducted by Fike et al. (2016) found that students who used a digital e-textbook compared to a hard copy textbook earned significantly lower scores on most of the test and quizzes in a statistics course. Overall, the final grade of the students who used digital textbooks was a letter grade lower compared to students that used a traditional print textbook. In a similar study conducted by Terpend et al. (2014), they did not find a statistically significant difference between students who used digital textbooks and those that used traditional hard-copy texts.

There is limited research on textbooks usage as it relates to student success. A study conducted by Junco and Clem (2015) examined the relationship between digital textbook usage and course success. The study first examined the CourseSmart Engagement Index which gave an overall engagement score for each student based on his or her engagement with the digital textbook. The results showed that this index was a significant predictor of student success in the course. The second part of the study disaggregated the parts of the Course Smart Engagement Index to see how they related to course performance. The seven parts of the index were pages read, number of days read, reading sessions, time reading, highlights, bookmarks, and notes (Junco & Clem, 2015). The results of the study found that number of days read was the only factor that was a statistically significant predictor of the course grade (Junco & Clem, 2015). A study conducted by Rockinson-Szapkiw et al. (2013) examined the relationship between perceived learning and type of textbook as well as final grade and type of textbook. The results of the study showed that students who used digital textbooks had higher perceived affective and psychomotor learning (Rockinson-Szapkiw et al., 2013).

Summary

The widespread growth of technology in the digital era is generating data faster than ever before (DOMO, 2018). Higher education institutions are collecting large amounts of data about their students in a variety of systems and formats (Chaurasia & Rosin, 2017; Siemens, 2013). This data is being collected throughout the life-cycle of a student, from pre-admission to graduation. These large datasets have paved the way for the field of learning analytics to continue to grow. Learning analytics is a new and growing field that has promise to assist institutions in helping to improve student learning.

One of the growing trends in the higher education sector is the increased expansion of digital textbooks. Digital textbook adoption and usage are gradually expanding throughout college campuses, partly due to the increased effort by publishers to provide digital content (McKenzie, 2017). Even though there is mixed research on the effectiveness of digital books and the impact they have on learning outcomes, institutions and publishers are still pushing adoption.

With the increase of digital content and access to students, the door has been opened for a new dataset to be explored (Junco & Clem, 2015). There is limited research on how students are using digital textbooks in their courses, especially online courses. Past research has focused primarily on student adoption and impact on learning compared to traditional print. There is a gap in the literature regarding actual textbook usage and its relationship to student success in online courses. This research seeks to add to the knowledge by reviewing the relationship between digital textbook usage metrics and course success in fully online courses.

CHAPTER THREE: METHODS

Overview

The following is an overview of the statistical methods used in this study. This section will focus on the research design, hypotheses, participants, instrumentation, procedures, and data analysis of this correlational study.

Design

This quantitative study used a correlational research design to determine if there is a significant predictive relationship between digital textbook analytics (consisting of the following predictor variables: pages read, number of days, reading sessions, highlights, bookmarks, notes, searches, downloads and prints) and the criterion variable: student success. Student success was based on performance in the course. For the first research question, the total number of points earned in the course was used to measure student success. The second research question used the quiz score average to measure success. Correlational research is appropriate for this study because the goal of this research is to determine if textbook analytics is a predictor of student success. Prediction studies allow the researcher to determine if the criterion behavior can be predicted (Gall, Gall, & Borg, 2007).

Research Questions

RQ1: Can digital textbook usage data predict course success in an undergraduate online course at a private liberal arts university?

RQ2: Can digital textbook usage data predict test success in an undergraduate online course at a private liberal arts university?

Null Hypotheses

$H_0$1: Digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, notes, searches, downloads and prints) do not significantly predict total number of points earned in an undergraduate online course at a private liberal arts university.

$H_0$2: Digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, notes, searches, downloads and prints) do not significantly predict average test scores in an undergraduate online course at a private liberal arts university.

Participants and Setting

This study used a nonprobability convenience sample. Participants were from a private nonprofit liberal arts institution located in the southern part of the United States. The institution offers both traditional residential programs and online programs. Participants for this study were undergraduate students enrolled in Psychology 255, Education 304, Apologetics 220, Computer Science Information Systems 110 and Physical Science 210 during the 2018–2019 school year from multiple programs of study. The target courses were offered fully online during an eight-week term. To be included in the study, the participants had to have used the digital textbook that was provided in the course.

The study used $N > 104 + k$, where k is the number of predictors to calculate the required sample size for medium effect at an $\alpha = .05$ (Warner, 2013). The target sample size for this study was 110 participants; this allowed for testing of multiple R as well as individual predictors (Warner, 2013). The total population of this study was 1,602 which exceeded the minimum population needed for medium effect size. The sample consisted of 444 males, 1,158 females. Ethnicity consisted of nine American Indian or Alaska native, 11Asian, 171 Black or African

American, 72 Hispanic or Latino, four Native Hawaiian or Pacific Islander, 796 White, four nonresident alien, 37 two or more races, and 498 unreported.

Instrumentation

The predictor variables (pages read, number of days, reading sessions, highlights, bookmarks, notes, searches, downloads and prints) were collected by the software and provided to the institution through a data feed. Demographic data of the participants came from the institution's student information system (SIS). Demographic data consisted of gender, age, ethnicity, current GPA, credits earned, and program. The criterion variables (points earned and average test score) were collected from the institution’s learning management system.

Using publisher and teacher made quizzes and tests in correlational research is a common practice in educational research (Gholami & Mostafa Morady, 2013; Poljicanin et al., 2009; Wagner, Ashurst, Simunich, & Cooney, 2016). This study identified quizzes through the learning management system with the associated course and used the student’s average grade on all quizzes as the criterion variable for RQ2.

Procedures

The data for this study was gathered from three sources: student information system (Banner), learning management system (Blackboard), and publisher (VitalSource). Data for these sources are streamed to the institution's data warehouse. The researcher made a formal IRB request and received approval for the research (see Appendix). A formal request was made to the Analytics and Decision Support (ADS) Office to pull the requested data points. ADS is the centralized reporting department at the university that provides data both internally and externally. The request identified the target courses and the data points that were needed.

Definitions for each of the variables were included in the request made to ADS. Pages read consisted of the total number of pages read. Number of sessions consisted of the number of times a student opened/interacted with the book. Number of days was defined as the distinct number of days a student used the textbook regardless of the amount of time. Highlights were defined as the number of highlights the student made throughout the textbook; bookmarks were the total number of bookmarks made; notes were the total number of annotations made; searches included the number of searches made; downloads referred to the number of times a student downloaded content; and print was the number of times a student used the print feature. These definitions are consistent with a previous study on digital textbook analytics (Junco & Clem, 2015).

Data from the SIS consisted mostly of demographic data. Gender was reported as 0 for female and 1 for male. Age was the age of the student as of the start date of the course. Ethnicity was pulled from admissions applications. Current GPA was the overall GPA of the student as of the start date of the course. Credits earned referred to the overall credits the student earned at of the start date of the course. Program of study was the current declared major of the student during the term the course was taken.

The criterion variables were pulled from the learning management system. The total number of points earned ranged from 0–1010. The average quiz score was the average score of all exams in the course represented as a percentage between 0%–100%. Data were anonymized and given to the researcher in Excel.

Data Analysis

This study used a multiple regression analysis to test the relationship between digital textbook usage data and student success. This statistical method was chosen because it allows the comparison of multiple predictor variables, can handle interval, ordinal, and categorical data, and provides analysis on both magnitude and significance (Gall et al., 2007). In this study nine predictor variables was analyzed. Once the request was fulfilled, the researcher received the anonymized data in Excel. Data from the Excel file was then uploaded into SPSS for analysis.

Assumption Testing

Multiple regressions require three assumption tests: the assumption of bivariate outliers, multivariate normal distribution, and the absence of multicollinearity. Scatter plots were used to determine if there were any extreme bivariate outliers. Scatters plots were also used to determine if there is a linear relationship between the variables (Warner, 2013). Variance Inflation Factors (VIF) were analyzed to determine if there was a violation of multicollinearity.

CHAPTER FOUR: FINDINGS

Overview

The purpose of this quantitative study was to analyze the predictive significance of digital textbook usage data, pages read, number of days, reading sessions, highlights, bookmarks, searches, prints, downloads and notes on final grades and quiz scores in undergraduate online courses at a private liberal arts university. A multiple regression analysis was used to determine if a predictive relationship exists between the predictor variables and the criterion variables. The results of each of the research questions is discussed in this section. Scatterplots were used to determine if the assumptions of the analysis were met. The following section analyses the results of each research question.

Research Questions

RQ1: Can digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, searches, prints, downloads, and notes) predict course success in an undergraduate online course at a private liberal arts university?

RQ2: Can digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, searches, prints, downloads, and notes) predict test success in an undergraduate online course at a private liberal arts university?

Null Hypotheses

$H_0$1: Digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, searches, print, downloads and notes) do not significantly predict total number of points earned in an undergraduate online course at a private liberal arts university.

H_02: Digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, searches, print, downloads and notes) do not significantly predict average test scores in an undergraduate online course at a private liberal arts university.

Descriptive Statistics

This study consisted of 1,627 records from 1,602 distinct students that were enrolled in five courses, Psychology 255, Education 304, Apologetics 220, Computer Science Information Systems 110, and Physical Science 210 in the Fall 2018 semester. Students that did not use the digital textbook or did not complete any of the quizzes were excluded from the analysis. An overview of the descriptive statistics of the criterion and predictor variables are listed in Table 1.

Table 1

Descriptive Statistics of Criterion and Predictor Variables

Variables	*M*	*SD*	*N*
Final Grade	808.71	157.655	1627
Quiz Average	80.24%	14.44%	1627
Pages Read	406.40	321.91	1627
Days Read	13.10	7.79	1627
Reading Sessions	19.64	15.76	1627
Highlights	40.79	159.30	1627
Bookmarks	.22	1.05	1627
Notes	.13	.86	1627
Searches	44.48	60.70	1627
Print	.52	1.05	1627
Downloads	.07	.26	1627

A multiple regression analysis was conducted to determine if there was a relationship between the predictor variables and the outcome variable. Multiple regressions have three major assumptions that need to be examined: the assumption of bivariate outliers, the assumption of multivariate normal distribution, and the test of non-multicollinearity. Multicollinearity was measured by accessing the tolerance levels and the VIF scores, which fell within normal ranges. This data is presented in Table 2. Bivariate outliers and normal distribution were examined by reviewing scatter plots. Scatter plots show a linear relationship between the dependent variable, final grade, and the independent variables: reading sessions, days read, pages read, highlights, searches, and downloads. There was not a linear relationship present between the dependent variable, final grade, and the independent variables notes, bookmarks, and prints. The scatterplots and descriptive statistics reveal that there was low usage of these features. Scatter plots show a linear relationship between the dependent variable, average quiz score, and the independent variables: reading sessions, days read, pages read, highlights, searches, prints, and downloads. There was not a linear relationship present between the dependent variable, quiz average, and the independent variables notes and bookmarks. As stated previously, these variables had low usage.

Scatterplots and boxplots were used to identify if there were any extreme outliers present. The graphs indicated the presence of outliers in each of the independent variables. Z-scores were calculated for each independent variable to identify values that had a Z-score higher than 3.29 or lower than -3.29. Once these cases were identified, they were removed from the model; this process excluded 145 outliers.

Table 2

Coefficients[a]

Model		Collinearity Statistics	
		Tolerance	VIF
1	Reading Sessions	.334	2.997
	Days Read	.309	3.241
	Pages Read	.365	2.740
	Highlights	.876	1.142
	Notes	.961	1.041
	Searches	.892	1.121
	Print	.996	1.004
	Bookmarks	.960	1.042
	Downloads	.840	1.191

a. Dependent Variable: Final Grade

Results

Null Hypothesis One

The first null hypothesis in this study states, "Digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, searches, print, downloads, and notes) do not significantly predict the total number of points earned in an undergraduate online course at a private liberal arts university." Table 4 shows that there is a significant relationship between the combination of the predictor variables and the criterion (outcome) variable, $R^2 = .154$, adjusted $R^2 = .15$, $p < .01$. Results for the predictive value of each variable are shown in Table 5. Predictors that exhibited a significant positive relationship with the criterion variable included days read ($p < .01$) , pages read ($p < .01$) and searches ($p < .01$).

Table 3

Model Summary

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
1	.393[a]	.154	.150	145.291

a. Predictors: (Constant), Downloads, Print, Notes, Bookmarks, Highlights, Searches, Days Read, Reading Sessions, Pages Read

Table 4

ANOVA[a] of Digitial Textbook Event Data and Overall Final Grade

Model		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Regression	6228130.0	9	692014.45	32.782	.000[b]
	Residual	34134135	1617	21109.545		
	Total	40362265	1626			

a. Dependent Variable: Final Grade

b. Predictors: (Constant), Downloads, Print, Notes, Bookmarks, Highlights, Searches, Days Read, Reading Sessions, Pages Read

Table 5

Coefficients[a] of All Predictor Variables and Overall Points Earned

Model		Unstandardized Coefficients		Standardized Coefficients			Correlations		
		B	Std. Error	Beta	*t*	Sig.	Zero-order	Partial	Part
1	(Constant)	707.095	7.225		97.872	.000			
	Reading Sessions	-.047	.396	-.005	-.118	.906	.294	-.003	-.003
	Days Read	5.463	.832	.270	6.562	.000	.374	.161	.150
	Pages Read	.060	.019	.123	3.253	.001	.343	.081	.074
	Highlights	.009	.024	.009	.367	.714	.122	.009	.008
	Notes	.984	4.264	.005	.231	.818	.055	.006	.005
	Searches	.171	.063	.066	2.719	.007	.159	.067	.062
	Print	-1.912	1.075	-.041	-1.778	.076	-.029	-.044	-.041
	Bookmarks	-.259	3.489	-.002	-.074	.941	.060	-.002	-.002
	Downloads	-8.094	15.217	-.013	-.532	.595	.042	-.013	-.012

a. Dependent Variable: Final Grade

A histogram was created to ensure that the data was normally distributed. Figure 1 shows the residual is closely aligned to the normal curve; it is slightly skewed to the left. Based on this data, the null hypothesis can be rejected; there is a significant predictive relationship between the predictor variables and the outcome variable. Based on the coefficients analysis in Table 5, the results indicate a significant relationship between the number of days read ($p < .01$), number of pages read ($p < .01$) and number of searches made ($p < .01$). The number of reading sessions, highlights, notes, prints, bookmarks, and downloads were not significant in this study; each of these variables had p-values greater than .05.

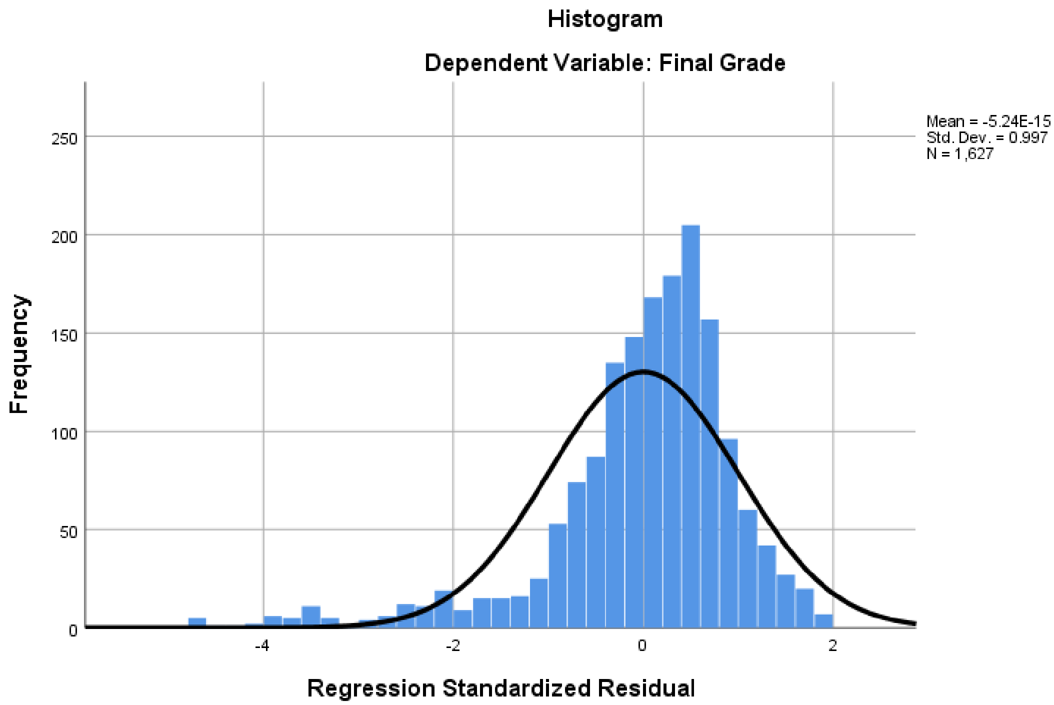

Figure 1. Final grade histogram.

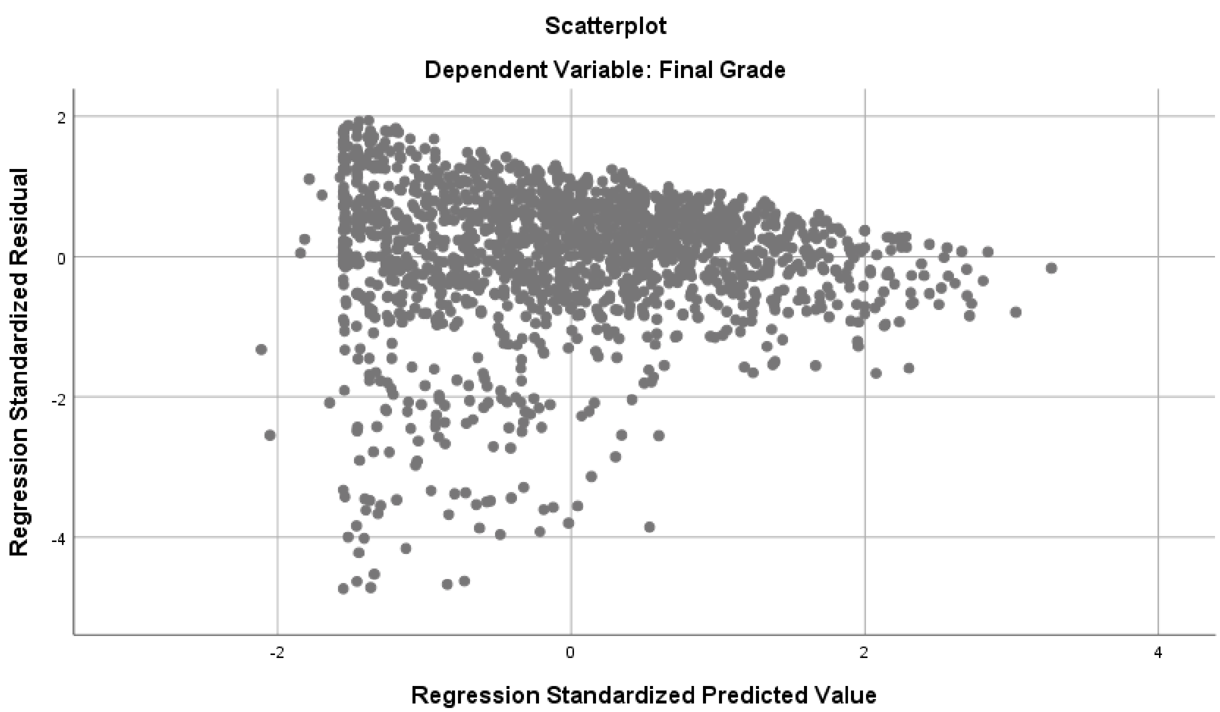

Figure 2. Scatterplot of final grades.

Null Hypothesis Two

The second null hypothesis of this study states, "Digital textbook usage data (pages read, number of days, reading sessions, highlights, bookmarks, searches, print, downloads and notes) do not significantly predict average test score in an undergraduate online course at a private liberal arts university." Table 6 shows that there is a significant relationship between the combination of the predictor variables and the criterion (outcome) variable, R^2 =.104, adjusted R^2 = .10, $p < .01$. Results for the predictive value of each variable are shown in Table 7. Predictors that exhibited a significant positive relationship with the criterion variable included days read ($p < .01$), pages read ($p < .01$) and print ($p = .04$).

Table 6

Model Summary[b]

Model	*R*	*R* Square	Adjusted *R* Square	Std. Error of the Estimate
1	.324[a]	.105	.100	.1369914

a. Predictors: (Constant), Downloads, Print, Notes, Bookmarks, Highlights, Searches, Days Read, Reading Sessions, Pages Read

b. Dependent Variable: Quiz Average

Table 7

ANOVA[a] of Digital Textbook Event Data and Average Quiz Score

Model		Sum of Squares	*df*	Mean Square	*F*	Sig.
1	Regression	3.562	9	.396	21.092	.000[b]
	Residual	30.346	1617	.019		
	Total	33.908	1626			

a. Dependent Variable: Quiz Average

b. Predictors: (Constant), Downloads, Print, Notes, Bookmarks, Highlights, Searches, Days Read, Reading Sessions, Pages Read

Table 8

Coefficients[a] of All Predictor Variables and Average Quiz Score

		Unstandardized Coefficients		Standardized Coefficients			Correlations		
Model		B	Std. Error	Beta	*t*	Sig.	Zero-order	Partial	Part
1	(Constant)	.727	.007		106.708	.000			
	Reading Sessions	-.001	.000	-.062	-1.511	.131	.224	-.038	-.036
	Days Read	.004	.001	.234	5.535	.000	.302	.136	.130
	Pages Read	.000	.000	.139	3.575	.000	.284	.089	.084
	Highlights	.000	.000	-.012	-.473	.636	.084	-.012	-.011
	Notes	.003	.004	.015	.623	.534	.054	.015	.015
	Searches	.000	.000	.045	1.789	.074	.131	.044	.042
	Print	-.002	.001	-.048	-2.051	.040	-.039	-.051	-.048
	Bookmarks	-.001	.003	-.005	-.221	.825	.045	-.005	-.005
	Downloads	.013	.014	.023	.891	.373	.055	.022	.021

a. Dependent Variable: Quiz Average

A histogram was created to ensure that the data was normally distributed. Figure 3 shows that the residual is closely aligned to the normal curve; however it is slightly skewed to the left. Based on this data, the null hypothesis can be rejected; there is a significant predictive relationship between the predictor variables and the outcome variable. Examining the coefficients in Table 8, the results of the study show three significant variables: number of days read ($p < .01$), pages read ($p < .01$), and number of print actions ($p < .01$). Reading sessions, highlights, notes, bookmarks, and downloads were not significant factors because they had *p*-values greater than .05.

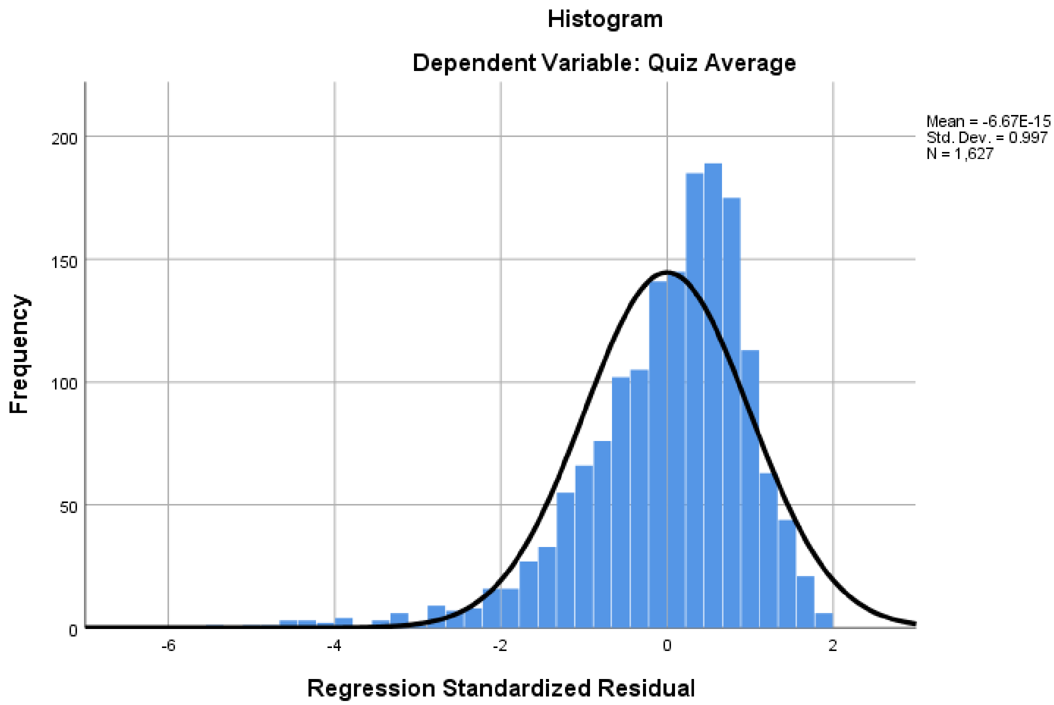

Figure 3. Quiz average histogram.

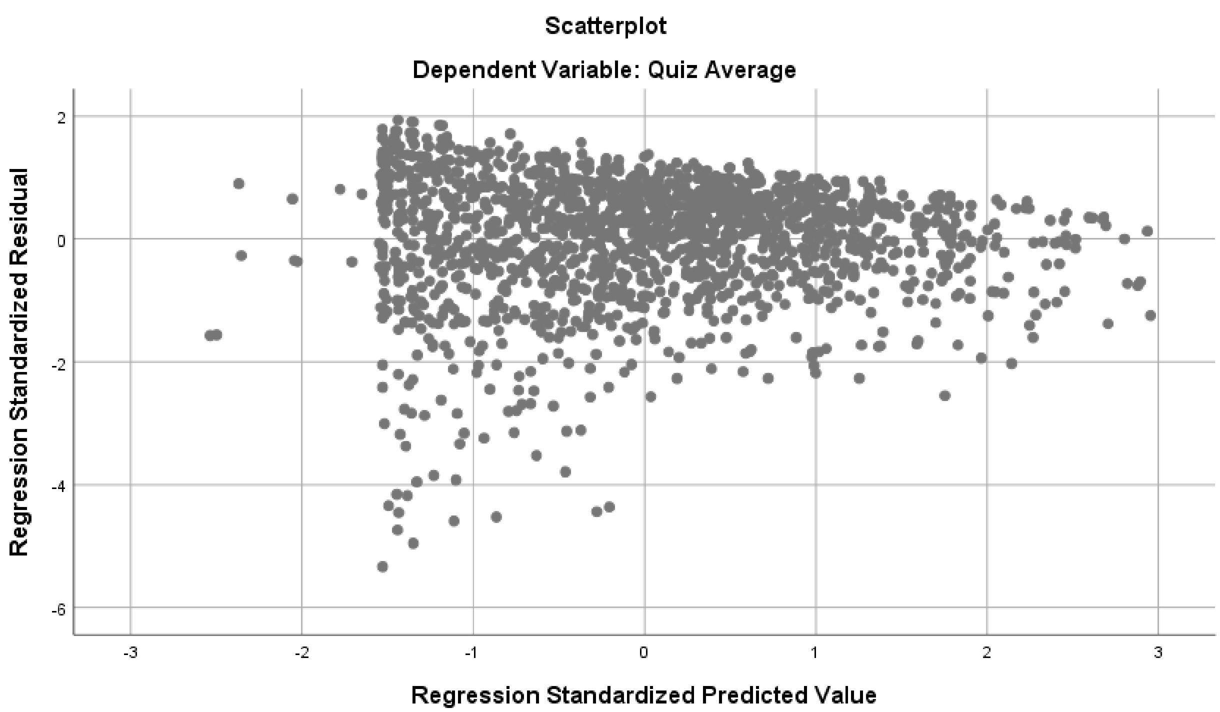

Figure 4. Scatterplot of quiz scores.

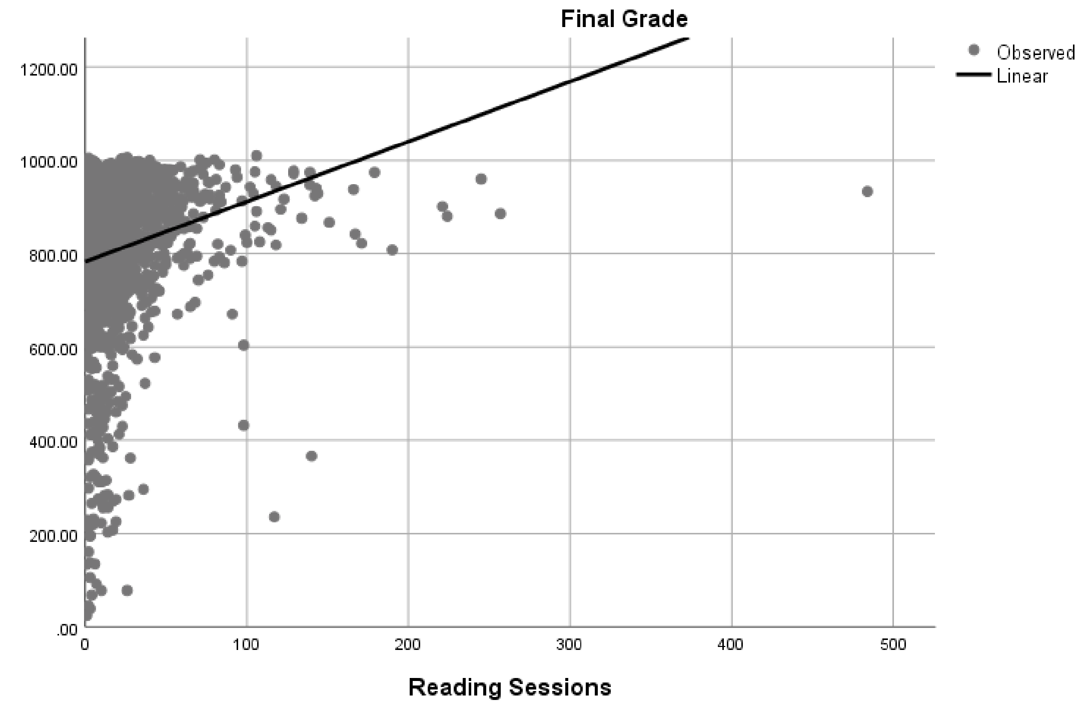

Figure 5. Scatterplot of reading sessions and final grade.

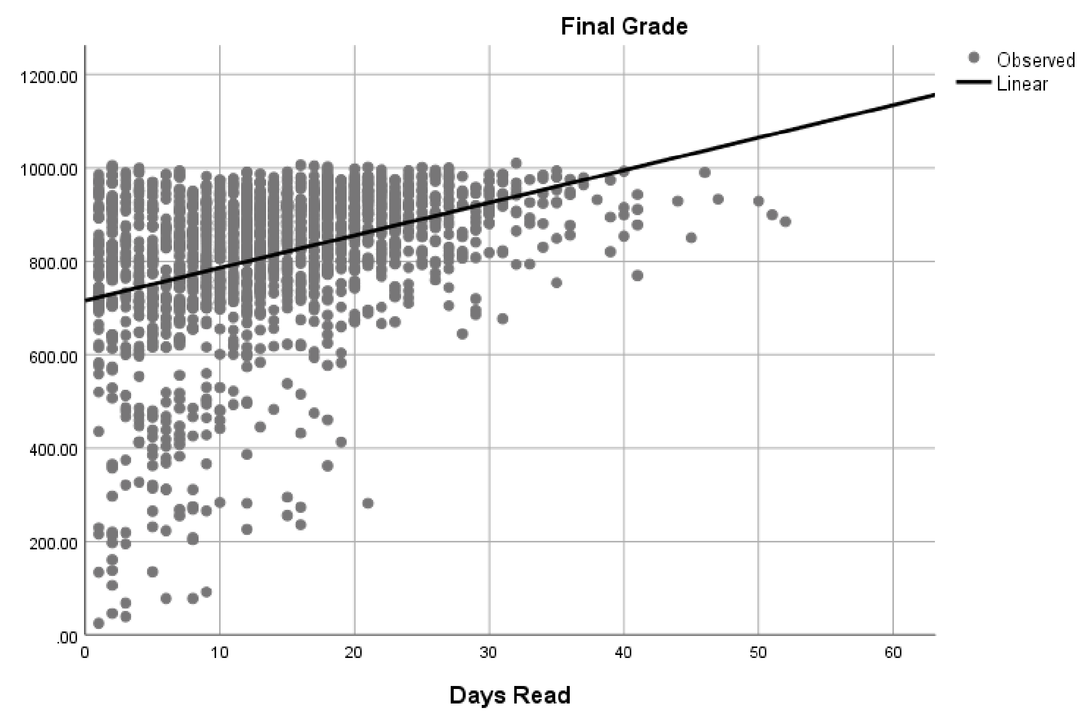

Figure 6. Scatterplot of days read and final grade.

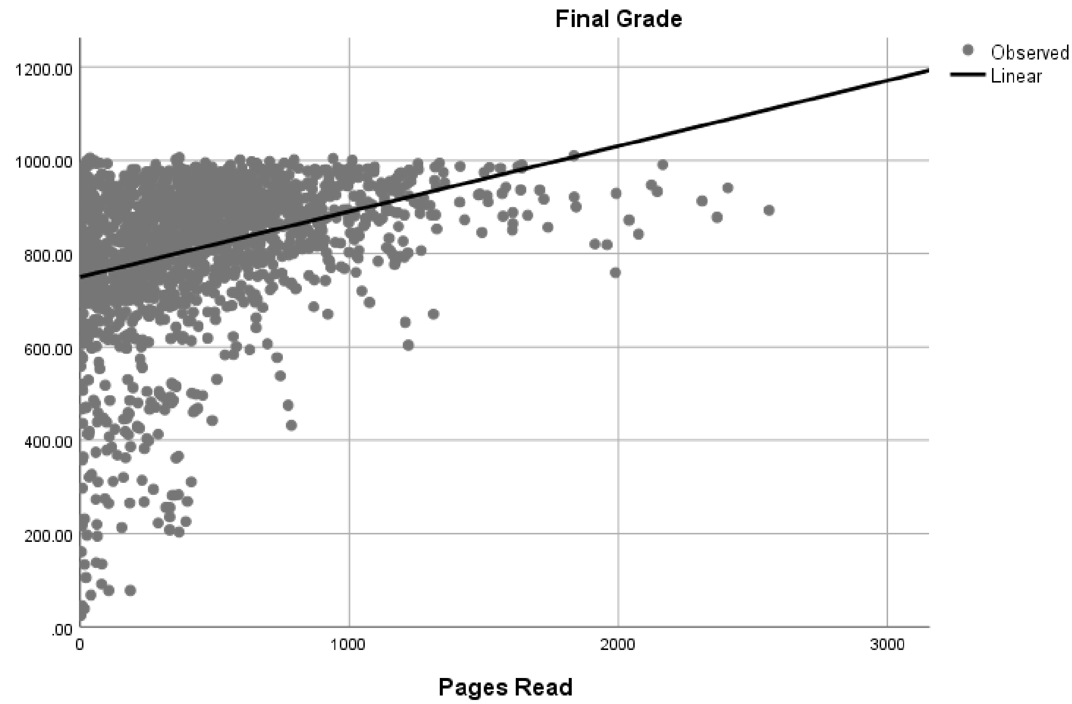

Figure 7. Scatterplot of pages read and final grade.

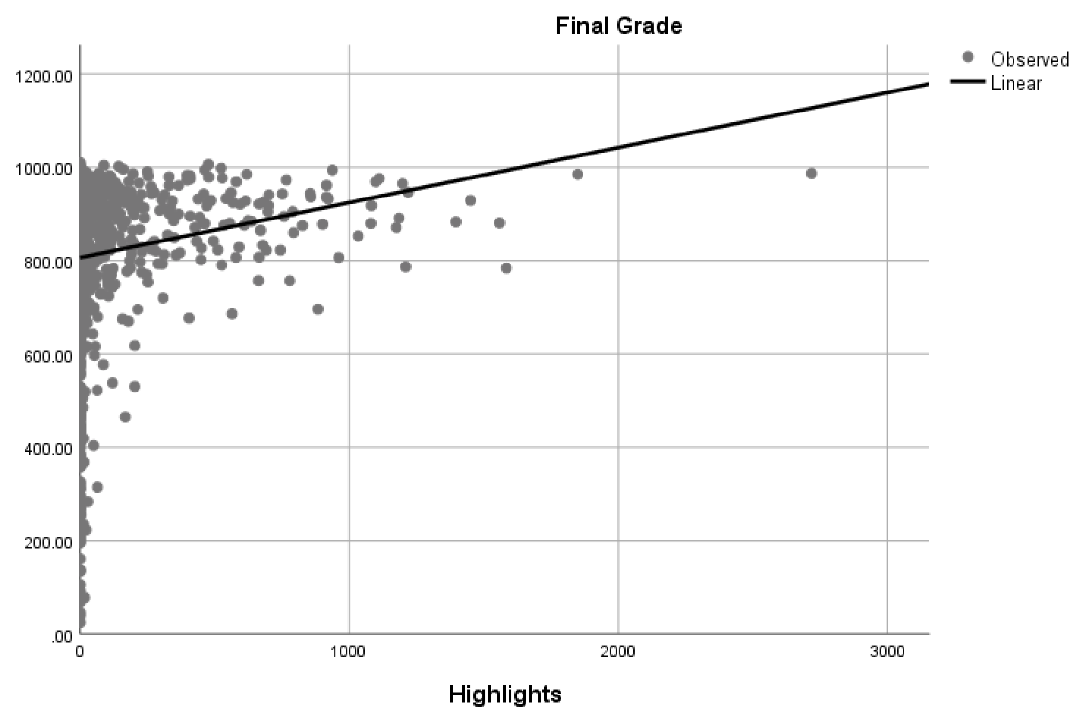

Figure 8. Scatterplot of highlights and final grade.

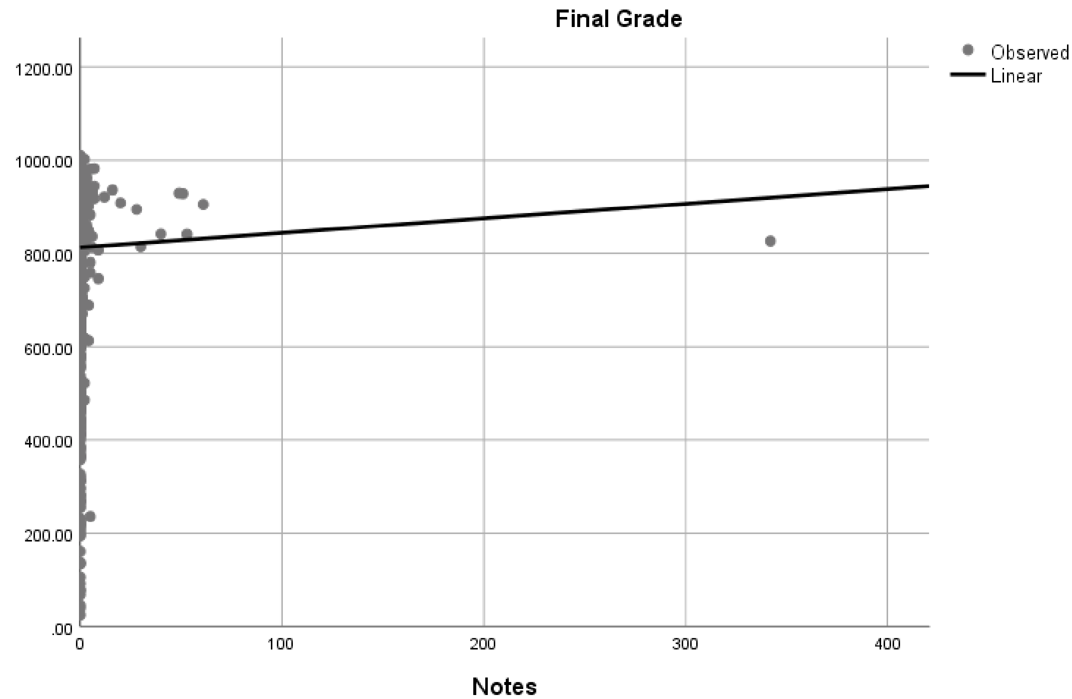

Figure 9. Scatterplot of notes and final grade.

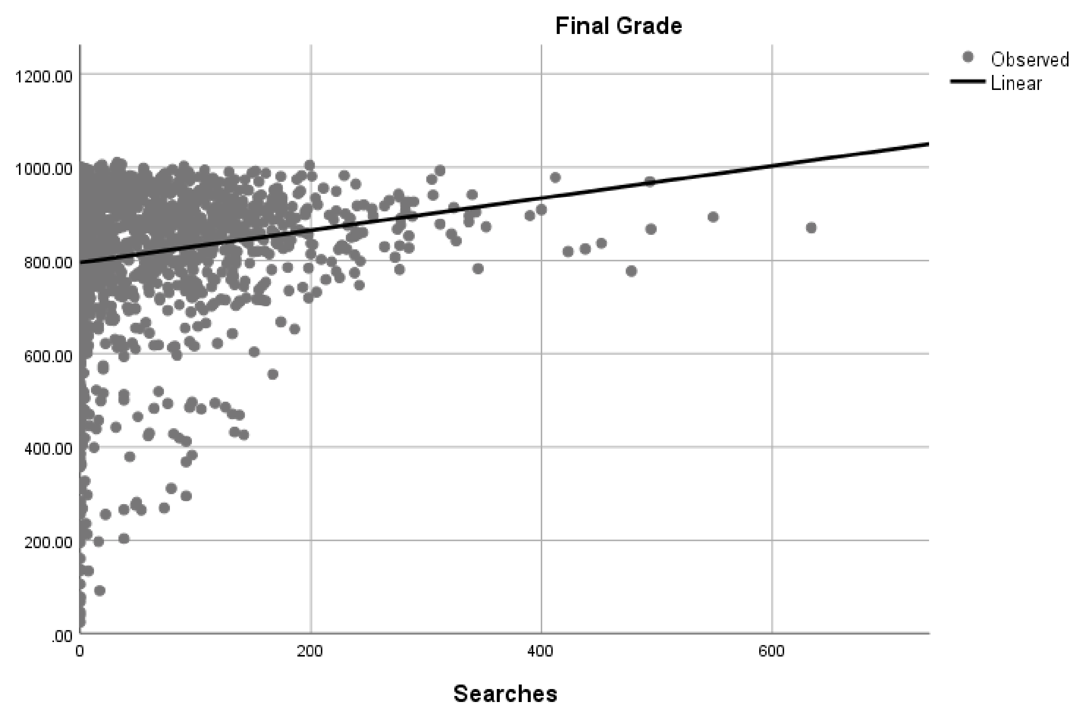

Figure 10. Scatterplot of searches and final grade.

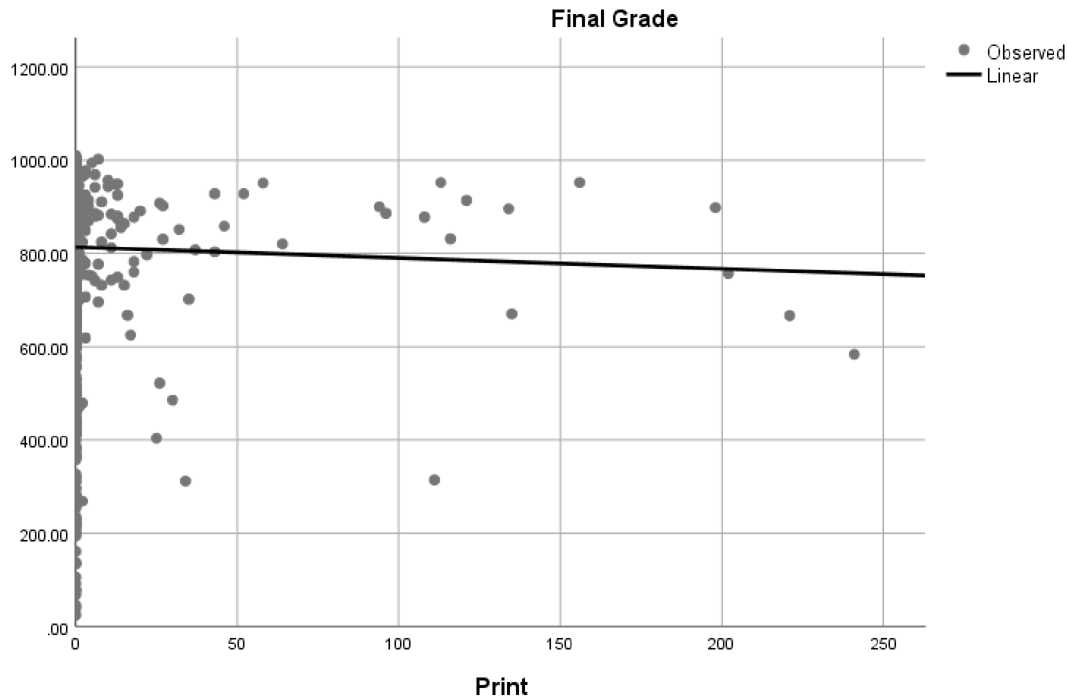

Figure 11. Scatterplot of print and final grade.

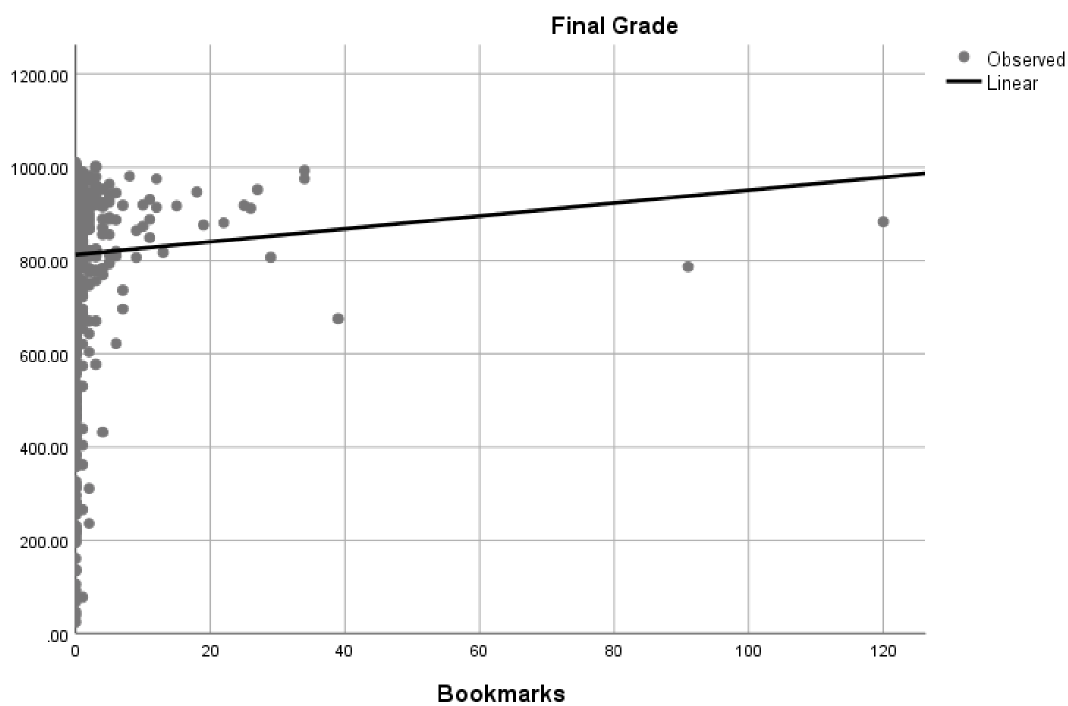

Figure 12. Scatterplot of bookmarks and final grade.

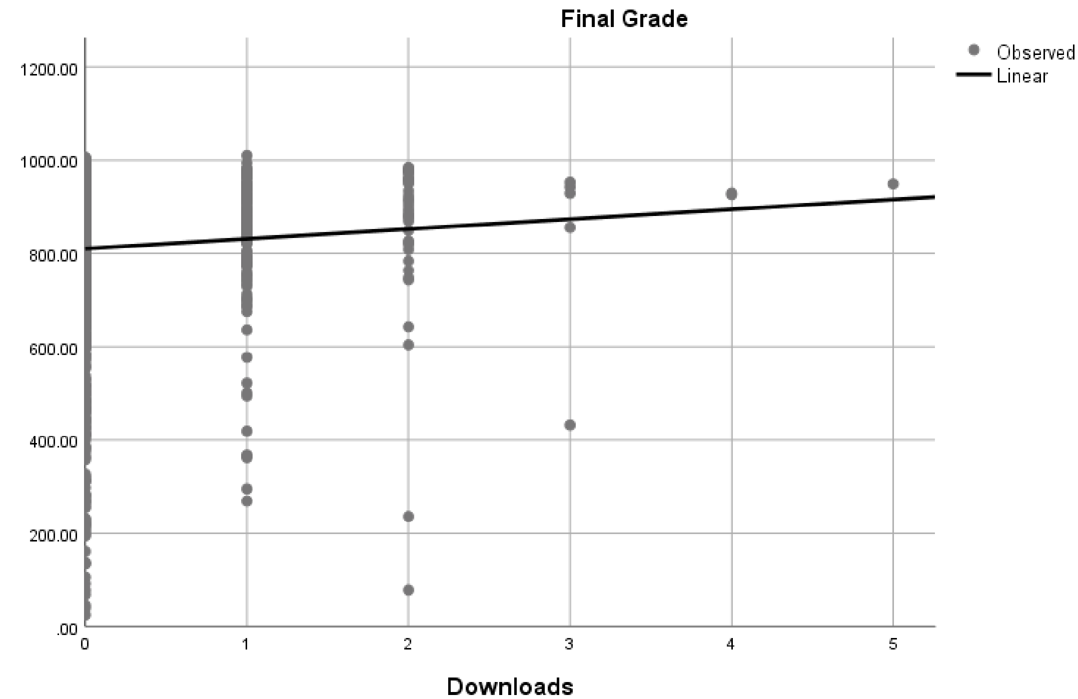

Figure 13. Scatterplot of downloads and final grade.

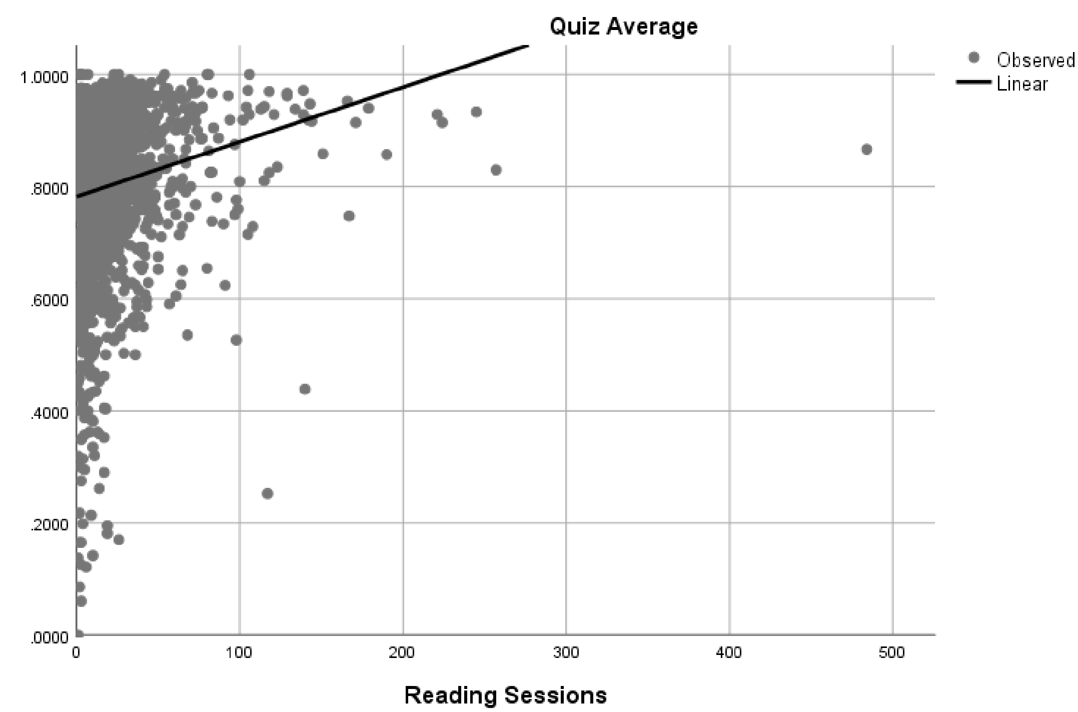

Figure 14. Scatterplot of reading sessions and quiz average.

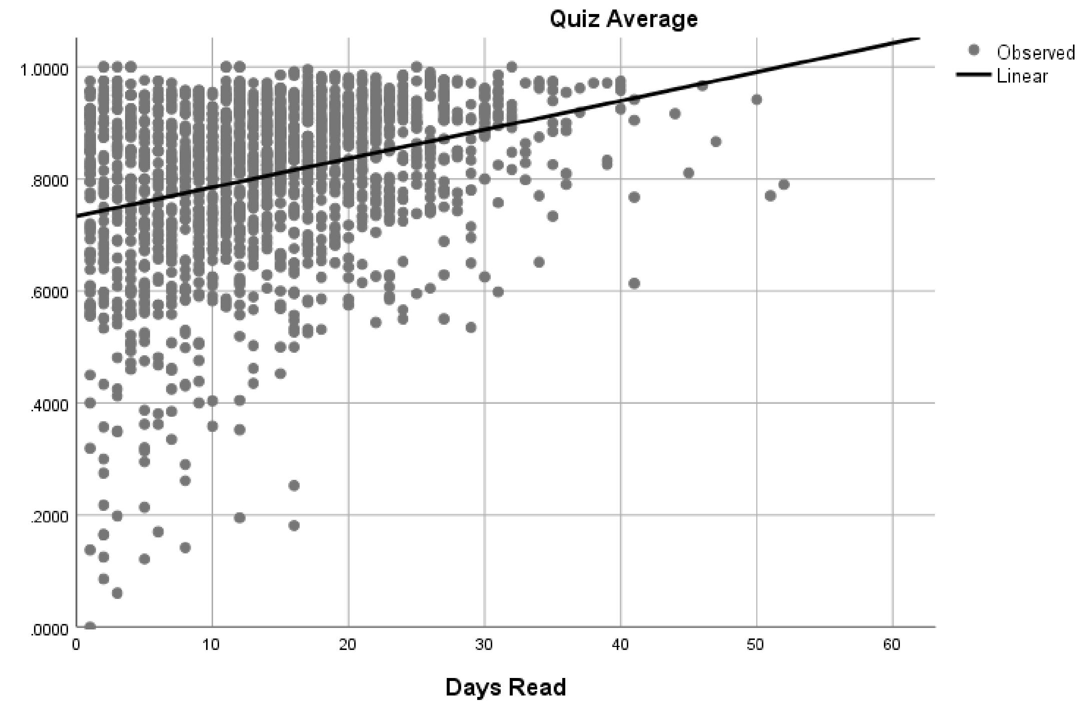

Figure 15. Scatterplot of days read and quiz average.

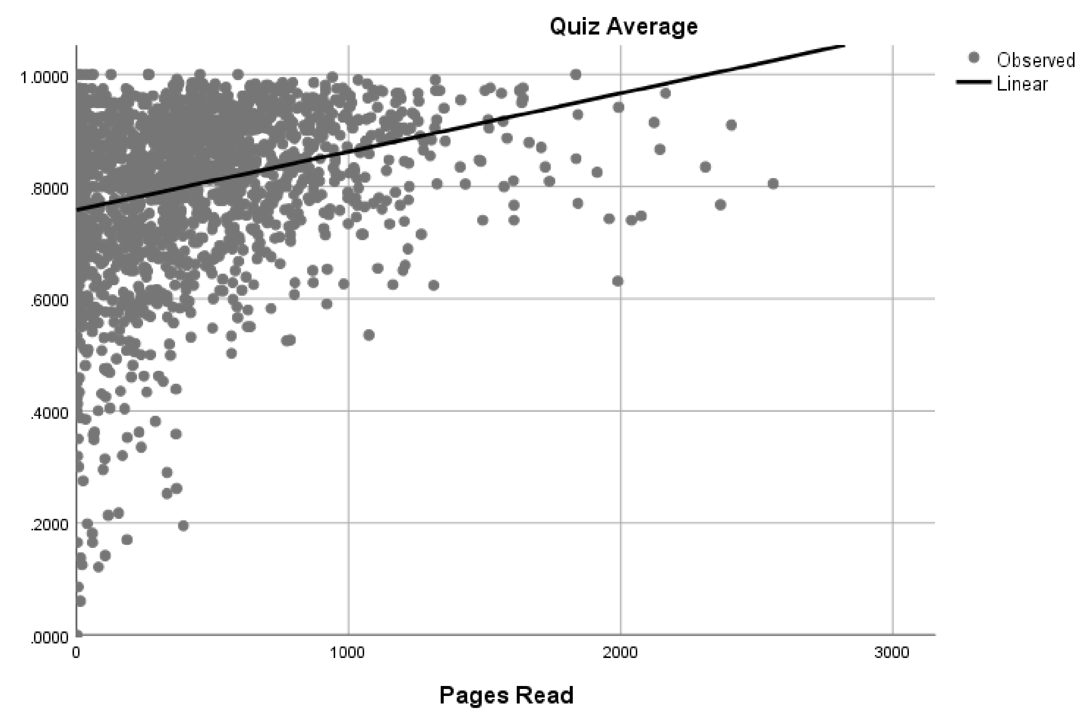

Figure 16. Scatterplot of pages read and quiz average.

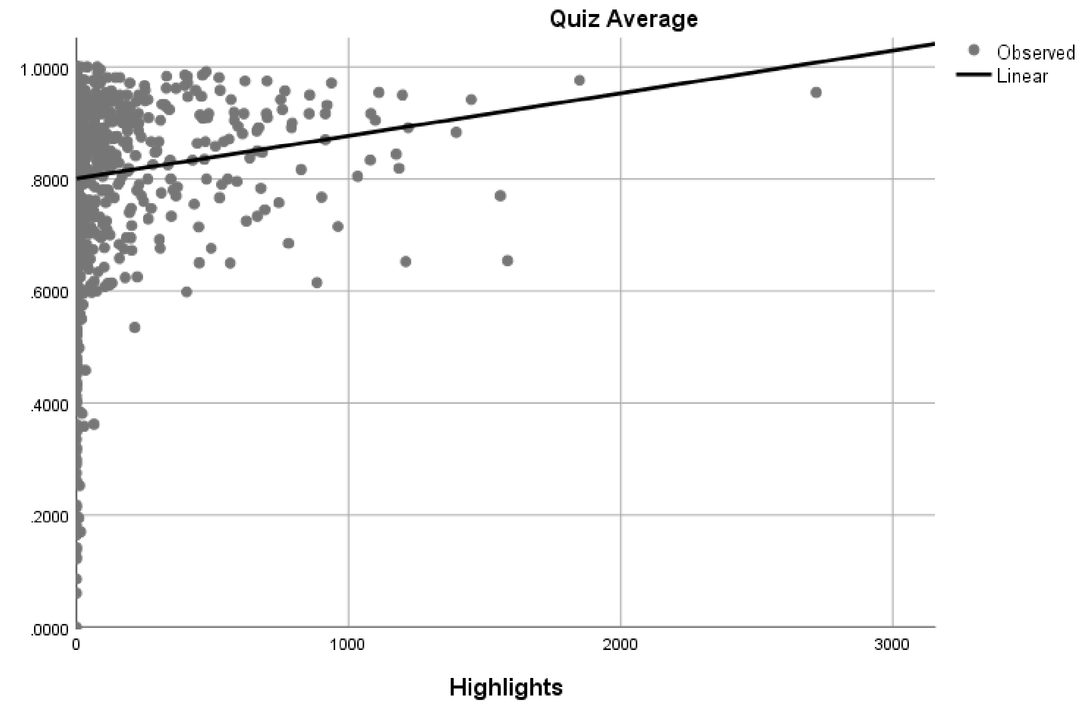

Figure 17. Scatterplot of highlights and quiz average.

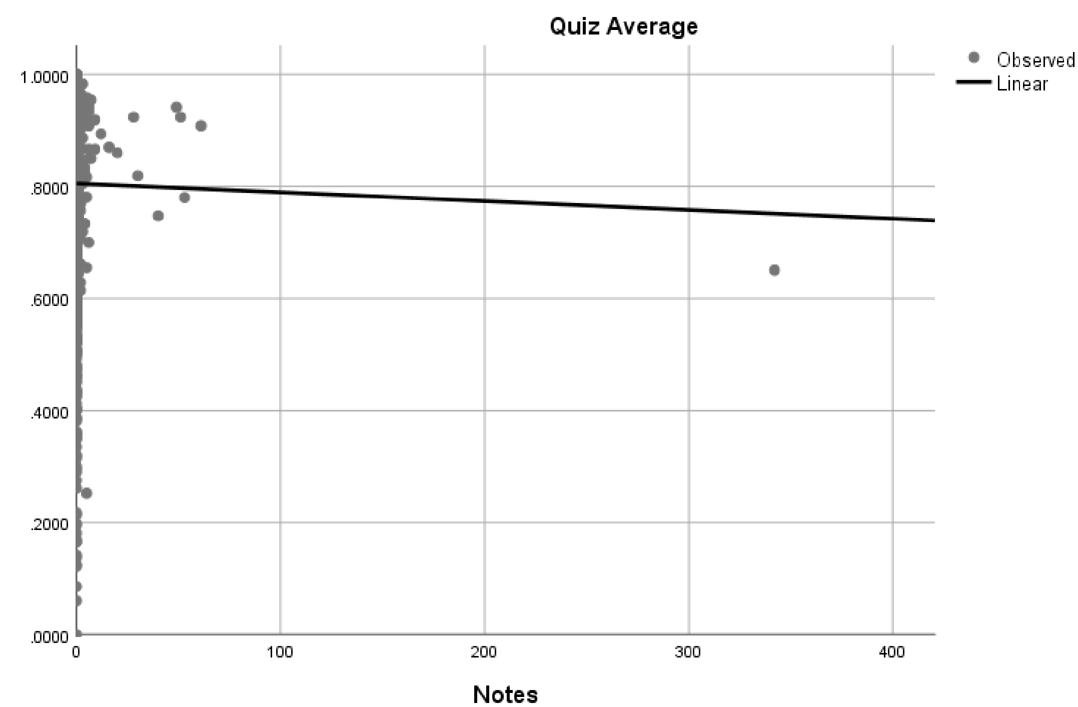

Figure 18. Scatterplot of notes and quiz average.

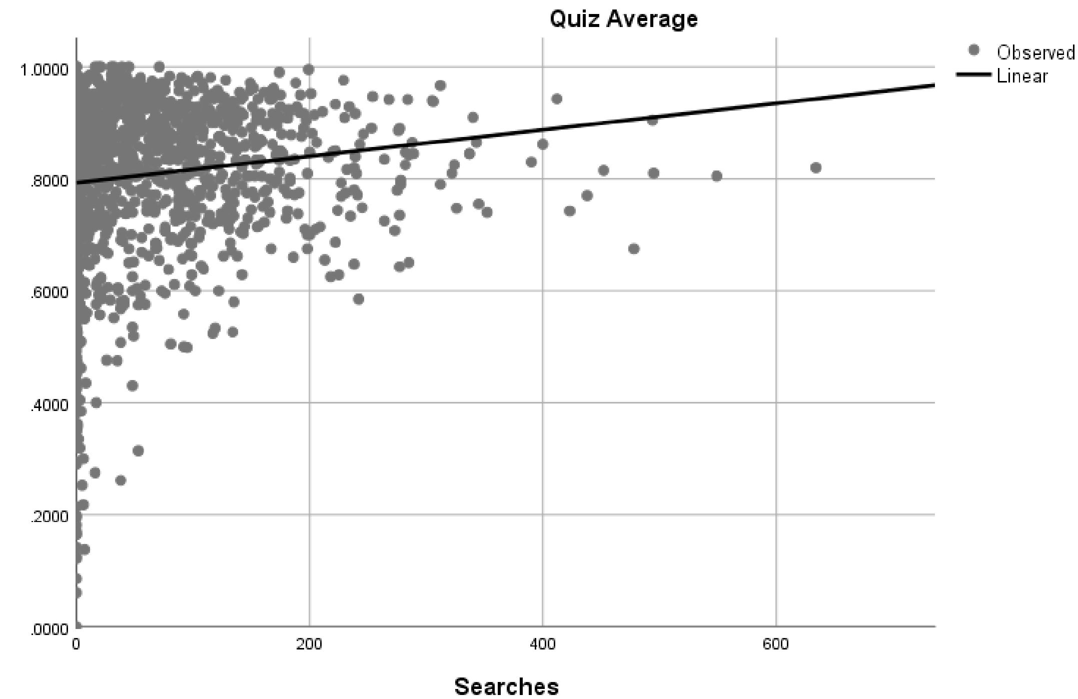

Figure 19. Scatterplot of searches and quiz average.

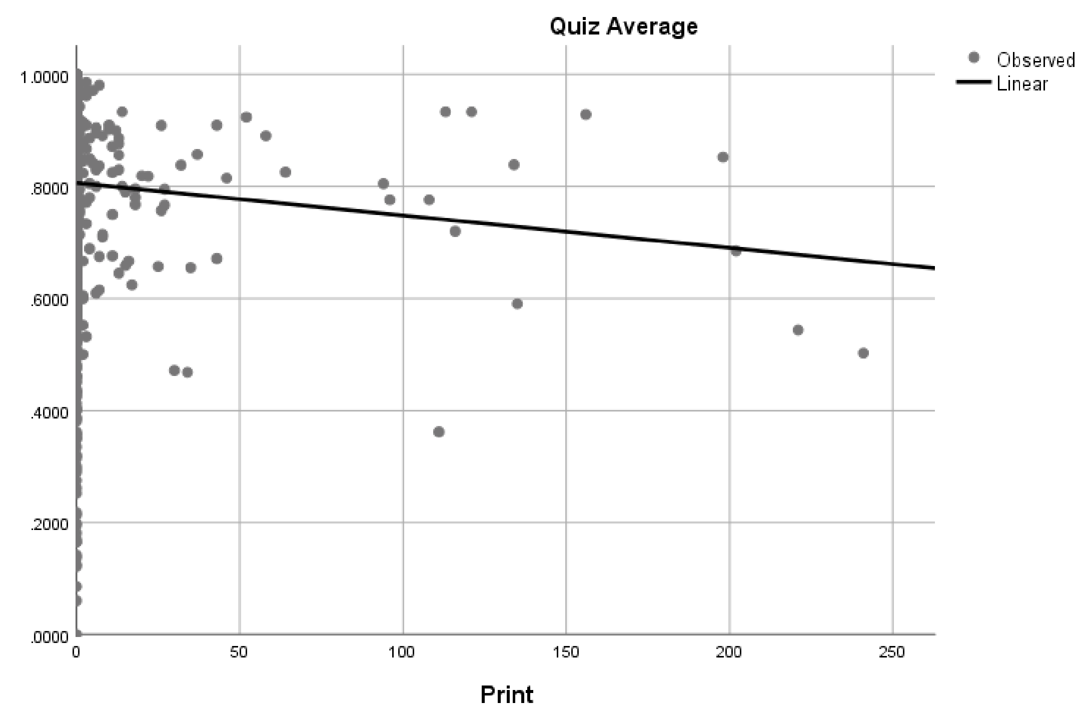

Figure 20. Scatterplot of print and quiz average.

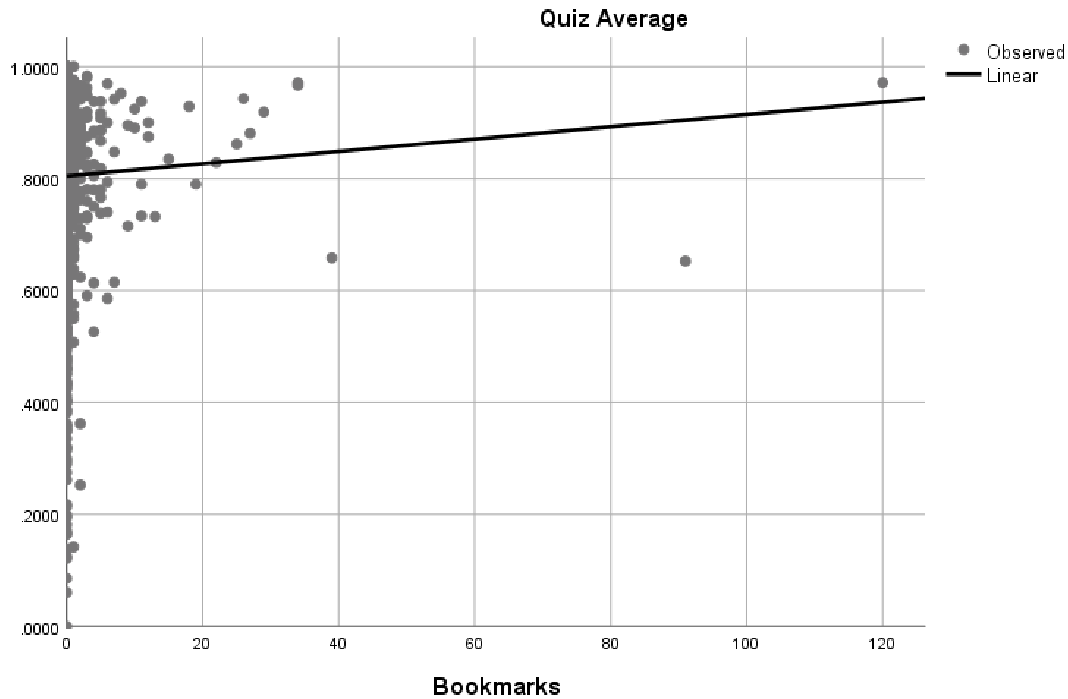

Figure 21. Scatterplot of bookmarks and quiz average.

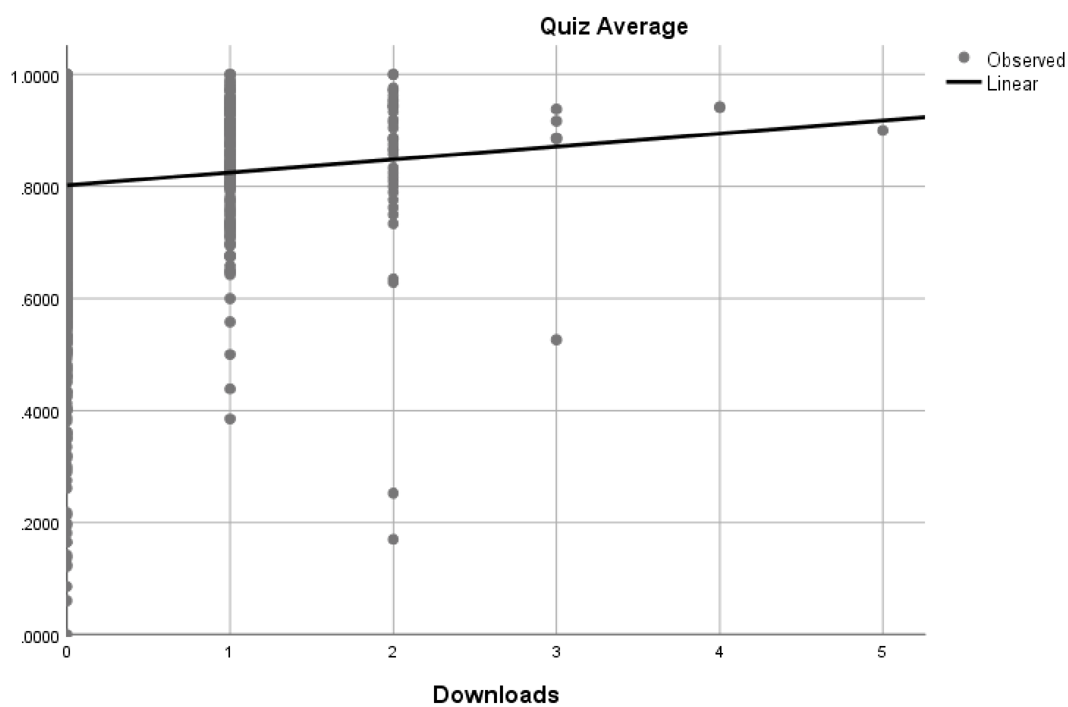

Figure 22. Scatterplot of downloads and quiz average.

CHAPTER FIVE: DISCUSSION

Overview

The purpose of the study was to explore the use of digital textbooks in online courses and determine if a predictive relationship exists between digital textbook usage and course performance. A multiple regression analysis was conducted between the predictor variables (reading sessions, days read, pages read, highlights, notes, searches, print, bookmarks, downloads) and the criterion variable of total points earned. A multiple regression analysis was also conducted between the predictor variables and the criterion variable of average quiz grade. The following section discusses the results of each of the null hypothesis in relation to the outcomes of this study, previous studies, and the overall theoretical framework.

Discussion

The purpose of the study was to examine if there was a predictive relationship between students' digital textbook usage and their overall grade or quiz average. The focus of the study was to use data that were not self-reported and were tracked using a digital textbook platform. The study examined nine digital textbook events: number of reading sessions, number of distinct days read, number of pages read, number of highlights made, number of notes taken, number of searches performed, number of prints, number of bookmarks, and number of downloads.

This research was chosen because of the increase in demand for digital textbooks in higher education courses, the change in textbook market strategy from publishers, and the increased interest in learning analytics in higher education (Clow, 2013; McKenzie, 2017; Reynolds, 2011). Research has shown an increase in digital textbook adoption by students and professors; digital textbook usage has risen from 42% in 2012 to 66% in 2016 (deNoyelles & Raible, 2017). Publishers are shifting their focus away from print textbooks and are instead focusing on increasing digital content and creating programs that provide students digital access to textbooks at the start of the course. Learning analytics is a growing field in higher education that typically leverages large datasets around how students are learning. Higher education institutions hope to understand learning at a deeper level and use data collected about student learning to increase student success (Clow, 2013; Klašnja-Milićević et al., 2017). The primary problem this research sought to address is the lack of research around how students are engaging with the digital textbook along with the potential predictive value of this growing dataset. This study seeks to add to the research conducted by Junco and Clem (2015) by analyzing additional predictor variables (searches, prints, and downloads; examining online courses) and analyzing if there is a relationship between the predictor variables and students' average quiz score.

Research Question One (Final Grade)

The first research question analyzed if a predictive relationship exists between digital textbook usage data (reading sessions, days read, pages read, highlights, notes, searches, print, bookmarks, and downloads) and the final grade of a student. Previous research has primarily focused on analyzing the impact that digital textbooks have on overall performance and the preference of digital textbooks versus print textbooks (deNoyelles & Seilhamer, 2013; Millar & Schrier, 2015; Rockinson-Szapkiw et al., 2013; Weisberg, 2011). Research conducted by Junco and Clem (2015) analyzed only seven digital textbook metrics: pages read, number of days reading, reading sessions, time reading, highlights, bookmarks, and notes. Junco and Clem's (2015) research utilized a hierarchical regression to see if the use of digital textbooks added to the predictive power of the regression model. The hierarchical regression had five blocks: gender, race/ethnicity, course/section, transfer GPA, and engagement. Each block added additional variables and measured the change of R^2. Adding in the individual digital textbook usage components accounted for a .077 increase in R^2. Analyzing the seven digital textbook usage metrics individually, Junco and Clem's (2015) study found that the number of days a student read to be the only significant predictor. The overall regression model was significant, but the distinct number of days read was the only independent variable that was a significant predictor of the final grade.

The multiple regression model in this study was significant, $p < .01$, adjusted $R^2 = .15$, meaning that the model accounted for 15% of the variance between the dependent and independent variables. Examining the predictor variables, this study found that the number of days read, the number of pages read, and the number of searches were all significant predictors of the overall grade in the course. The significance of number of days read aligns with previous research conducted by Junco and Clem (2015). However, the significance regarding the number of pages read contradicts the research by Junco and Clem (2015) where they found that this variable was not statistically significant in predicting final grades. The number of searches a student conducted is a newer metric being tracked, and there is no relevant research for comparison.

This study as well as the study conducted by Junco and Clem (2015) found that the number of distinct days a student read was a significant predictor of the final grade in the course. Junco and Clem (2015) found that the average student spent 11 distinct days in the textbook over a 16-week course. This study found that the average student spends 13 distinct days in the textbook over an eight-week course. This research continues to support research between time-on-task and student performance, as well as support the theories associated with student engagement and student success. Research conducted by French et al. (2015) found a relationship between the amount of time a student reads and the overall grade in the course.

Students who self-identified as reading 75% or more of the assigned reading performed at a higher rate than students who read the book often or rarely. Research conducted by Gyllen, Stahovich, and Mayer (2018) found that engineering students who spent time working through homework problems inside of a digital textbook had higher overall scores in the course. From a theoretical perspective, Astin's (1999) theory of student involvement believes that time on task will promote engagement and that engagement will lead to increased student learning (Junco & Clem, 2015). The research conducted by Junco and Clem (2015) focused on a residential population; this study adds to this research by showing a similar relationship in online courses. It is notable to mention that students in a condensed eight-week course utilize the textbook on average two days more than students in a 16-week course.

This study found that the number of pages read was a significant predictor, $p < .01$, of the overall grade, which contradicts the previous study conducted by Junco and Clem (2015) where there was no significant relationship found. An explanation of the difference in the findings may be attributed to the size of the population or the modality of the course. The study conducted by Junco and Clem (2015) had a sample size of 236 students (noted as a limitation in their study), whereas this study included a sample size of 1,772. In this study, the modality of the course was focused around eight-week online courses, compared to the 16-week residential courses in the Junco and Clem study.

This study aligns with the research conducted by French et al. (2015) which showed that students who self-report that they read more than 75% of the assigned reading have a higher overall grade in the course. The study conducted by Gyllen et al. (2018) showed that student engagement with course problems in the textbook was a strong predictor of student final grades. However, in the study conducted by Gyllen et al. the researcher noticed that students were not reading the textbook but instead focusing on the homework and practice problems that were included in the textbook. Additional research is needed to see if the number of pages read is a significant predictor of the final grade. Future research should also focus on the type of content that the student is engaging with inside the digital textbook.

This study also found a significant relationship between the number of searches, $p < .01$, a student conducted and the final grade in the course. The relationship between the use of the search feature and the overall grade in the course has not been formally discussed in past research, which has primarily focused on digital textbook features that students prefer along with reasons why some students prefer digital textbooks over print textbooks. Navigation and the use of searches are some of the primary features of digital textbooks that students enjoy (Dobler, 2015). The use of digital textbooks has shown to increase student engagement for some students. Dobler (2015) found that some of the digital textbook features like electronic note taking and sharing along with the use of search increased student reading habits. Increased time-on-task along with increasing student engagement has shown to increase learning and student outcomes (Astin, 1999). Additional research is needed to determine if similar results exist in other online courses as well as residential courses and in both public and private institutions.

The findings of this research align with the principles associated with both student engagement theory and self-regulated learning theory. Student engagement theory, influenced largely by the work conducted by Tinto (1975) and Astin (1999), asserts that the more a student is engaged with the institution, the higher chance of that student succeeding and persisting to completion. Tinto’s early theory centered on institutional and goal commitment. Students have varying levels of commitment as they enter and progress through their education. In regard to institutional commitment, Tinto believed that a student’s academic commitment and academic success played a large part in retention. In later research, Tinto discussed the paradox between the role of the student and the institution (Tinto, 2012). Tinto believes that there needs to be a commitment from both the student and the institution. On the institution side, Tinto believes that higher education institutions need to be open to change and rethink how learning environments are structured as well develop creative ways to keep students engaged (Tinto, 2012).

The principles of self-regulated learning theory also align with the findings of this study. Self-regulated learners tend to be actively aware of where they are in the learning process and have the ability to adapt their learning strategies to meet their educational goals. One of the key characteristics of self-regulated learners is their consistency to be active participants throughout the learning process (Zimmerman, 1986).

In summary, the findings of this study found that the distinct number of days, the number of pages read, and the number of searches conducted were significant predictors of total points earned. These predictor variables show active engagement between the student and the course material. The data from this study suggests that as students read and interact with some of the features of the textbook, the more likely they are to receive a higher grade in the course. This aligns with the principles found in both student engagement theory as well as self-regulated learning theory.

Research Question Two (Quiz Average)

The second research question analyzed if there was a predictive relationship between digital textbook usage data (reading sessions, days read, pages read, highlights, notes, searches, print, bookmarks, and downloads) and the average quiz score. There has been a lack of research between digital textbook usage and overall quiz scores. As discussed previously, previous research has primarily focused on student preferences between digital and print textbooks as well as overall student performances between digital and print textbooks. This study sought to isolate a sub-component of the overall grade, quiz scores, to see if a predictive relationship exists between digital textbook usage data and average quiz grade. The study conducted by Junco and Clem (2015) focused on understanding how students are engaging with digital textbooks; analyzing the predictive relationship between a proprietary engagement score, that was provided to the researcher by the publisher, and the final grade of the student; and analyzing the predictive relationship between the individual components of textbook usage data and the final grade of the course.

The multiple regression model in this study was statistically significant; adjusted $R^2 = .098$, $p < .01$, meaning that the model accounts for approximately 10% of the variance. An output of the multiple regression model produces a coefficients table that allows the researcher to examine the individual variables of the model. A *t*-test is performed to determine if the variable contributes to the overall significance of the regression model. Examining the coefficient table, the number of days read, the number of pages read, and the number of printing events were statistically significant.

The predictor variable (number of days read) aligns with the results of the previous research question and the previous study conducted by Junco and Clem (2015). Past research has shown the relationship between student success and time-on-task and amount of assigned reading completed (French et al., 2015; Gyllen et al., 2018). Past studies suggest that the more a student spends on task, either reading the textbook or working through practice problems, the more likelihood they will be successful in the course. This differs from the research conducted by Azorlosa (2012), who found that the amount of reading did not have any impact on student’s exam scores. Azorlosa (2012) suggests that having quizzes throughout the course help prepare students for exams. Additional research is needed to add to this knowledge base; researchers should examine whether similar trends can be found in residential courses and in other online courses.

The significance of the predictor variable, the number of pages read, aligns with the findings of the research conducted by French et al. (2015). Research conducted by French et al. (2015) found that students who read >75% of the assigned reading tend to have a higher overall score in the course compared with students that stated that they read between 50%–74% of the assigned reading. The courses that were selected for this study contained quizzes, but the number of quizzes and the number of questions per quiz were varied. The quizzes consisted of questions that were primarily based on the textbook readings in the course.

The predictor variable, pages printed, was also significant in this study. There is a slight negative correlation between pages printed and Quiz Average. The average number of print actions taken by a student was .52. Given the limited use of this digital textbook feature, additional research is needed in order to fully understand the implications this has on a student's average quiz score.

Students indicate that navigation and the use of search are some of the main benefits of using digital textbooks. Given that the quiz questions were largely compromised of textbook material and were open book, it is notable to point out that the number of searches made was significant when trying to predict the overall grade in the course but not significant when predicting the quiz average. Additional research is needed to see when students are using the search feature inside of the digital textbooks. Additional research may examine whether students are using the search feature at the same time they are taking a quiz.

The findings of this research align with both of the theoretical frameworks of this study. Similar to the findings in research question one, the number of pages read shows both student engagement and self-regulation. Both engagement and self-regulation have been correlated with stronger student performance, and the findings from this study suggest a relationship between student engagement with digital textbooks and quiz average. Students that are reading and interacting with the textbook are performing better on the quizzes.

Implications

This study has contributed to the limited knowledge on digital textbook analytics and provided valuable insight into how students engage with digital textbooks in online courses. Both of the models were statistically significant but had a low adjusted R^2, meaning that only a small amount of the variance was accounted for in the models. Given the low adjusted R^2, the models have limited use on their own but provide implications and insight for future research.

The study added to previous research by showing how students are interacting with digital textbooks. There is limited knowledge on how online students are engaging with digital textbooks; the descriptive statistics of this study provided additional insight into how students are interacting with digital textbooks. When analyzing the predictive relationship between digital textbook usage metrics and total points earned, this study found the number of days a student reads the textbook to be predictive, which aligns with the previous study conducted by Junco and Clem (2015). Both this study as well as the study conducted by Junco and Clem (2015) found that students had an overall low usage of the bookmark and notes features within the digital textbook. The analysis added information on three new digital textbook usage metrics: number of searches, downloads, and prints. Searches were the only new metric that was heavily used:

prints and downloads showed low usage. Similar to previous research, there is a high percentage of students that have low overall usage of the textbook.

The multiple regression analysis conducted between digital textbook analytics and final grade was significant but had a low R^2 of .15. One of the new digital textbook usage metrics, searchers, proved to be significant. The number of downloads and prints was rarely used and was not a significant factor in the model. The number of pages read was statistically significant, which contradicted previous research. In this study, the courses were online compared to the previous study conducted by Junco and Clem (2015) which focused on residential courses. Online courses lack the traditional lectures and more of the learning happens through reading, which may explain why pages read was a predictor of final grades in this model.

There were several metrics in this study that were predictive: even with a low R^2, the data points may be beneficial in identifying at-risk students. Tracking student engagement in online courses has been primarily focused around assessment outcomes (Junco & Clem, 2015). In online courses, there is typically a strong emphasis on using the LMS; research has shown a relationship between engagement inside of the LMS, based on clicks, and successful completion of the course (Hung et al., 2017). Expanding the dataset to include engagement metrics associated with digital textbooks may provide faculty at-risk indicators earlier in the course. The courses used in this study followed an eight-week condensed online model; in order to intervene and influence change, instructors need to know as early as possible if a student is at risk. Digital textbook analytics may provide faculty the ability to track engagement and predict course outcomes from textbook interactions, which can start generating data at the start date of the courses instead of having to wait for students to complete assessments and instructors to grade the assessments. Institutions are leveraging early warning systems that range in complexity and pull data from the entire student lifecycle. Pairing digital textbook analytic data with other student success predictor variables may strengthen the model and provide earlier at-risk classification. Outside of incorporating this data into at-risk models, faculty can access this engagement data directly, which may assist them in identifying students that are not engaging inside of the LMS or with the digital textbook.

The multiple regression analysis conducted between digital textbook analytics and the average quiz score was also significant. The adjusted R^2 was .098 for this model, which means that less than 10% of the variance is accounted for in this model. This model also found that the number of pages read and the number days read were significant. The significance of the number of pages read and the number of days read aligns with the findings in research question one and previous research. There were two notable outcomes of this analysis. The analysis showed that the number of searches a student conducted was not predictive of their quiz average. The quiz questions came primarily from the textbooks readings and were open book. Future studies may want to examine how students are using the search feature in digital textbooks to assist them with open book quizzes. The number of print actions a student took was also predictive. Looking into the relationship between print actions and quiz average, the results showed a negative correlation. The more print actions a student took, the lower the quiz average. Previous research has shown that some students have lower quiz scores and lower final grades when using digital textbooks. Additional research is needed to better understand why students are printing off pages and if they are using these printed pages during quizzes.

Limitations

There are several limitations of this study. One of the limitations is the population, which was a convenience sample and limited to undergraduate students taking online, asynchronous courses. The results of this study may only be applicable to this population. It is also important to note that the online courses used for this study were eight weeks in length and had a digital textbook that was provided to the student as part of enrollment.

The study leveraged the e-reader platform that was developed by VitalSource. It is important to note that VitalSource defined each of the events and published the definition of these events. Future research into digital textbook analytics will need to be mindful of how the publisher is defining these events. For instance, VitalSource defines a page read if the student stays on the page for at least four seconds. There are organizations seeking to standardize activity events; there is a developing standard called Caliper that may prove useful in future research studies (IMSGlobal, 2019).

Students in all undergraduate courses at the host institution were provided free digital textbooks as a part of their enrollment; however, each student has the option to buy a loose-leaf copy of the textbook for a reduced price or to purchase the textbook from the publisher or other third-parties. The researcher has no insight into the purchase of the print textbook, so it is possible that some of the students purchased a printed copy of the textbook and used it alongside the digital copy. Students that did not use the digital textbook at all were removed from the population.

Recommendations for Future Research

This study added to the limited research on digital textbook usage analytics. Learning analytics is an emerging field with many avenues for further research. Based on the outcomes and limitations of this study, below are recommendations for further research.

1. Replication of this study should be conducted with a focus on graduate online courses.

2. A similar study should analyze if digital textbook usage data increase the strength of an already existing at-risk model.
3. Replication of this study should be conducted in different online course formats that vary in length (7-week, 16-week, etc.) and modality (synchronous, asynchronous, hybrid).
4. Replication of this study should be conducted in different university settings—public, private, community college.
5. A similar study should be conducted with courses that do not have quizzes associated with the textbook.
6. A similar study should be conducted that uses different publishers and eBook readers outside of VitalSource.
7. A study should focus on incorporating digital textbook analytics into already running university at-risk models.
8. A longitudinal study should analyze student digital textbook behaviors throughout their education.
9. A study should focus on identifying the content students are interacting with inside the digital textbook (i.e., are they reading, or working on homework problems?).
10. Additional research should examine how students are using the digital textbook search feature; this study should seek to determine if students are utilizing this feature while students are taking a quiz and the impact this has on the student's quiz score.
11. Additional research should examine why students are printing pages out of the digital textbook and how they are using these pages throughout the course.

REFERENCES

Aljohani, N. R., Daud, A., Abbasi, R. A., Alowibdi, J. S., Basheri, M., & Aslam, M. A. (2018). An integrated framework for course adapted student learning analytics dashboard. *Computers in Human Behavior*. doi:10.1016/j.chb.2018.03.035

Arend, B. D. (2004). Bottom line—new patterns of student engagement. *About Campus, 9*(3), 30–32. doi:10.1002/abc.98

Arnold, K. E., & Pistilli, M. (2012). *Course signals at Purdue: Using learning analytics to increase student success*. ACM International Conference Proceeding Series. doi:10.1145/2330601.2330666

Astin, A. (1999). Student involvement: A development theory for higher education. *Journal of Student Development, 40*(5), 518–529.

Azorlosa, J. L. (2012). The effect of announced quizzes on exam performance: quiz content. *Journal of Instructional Psychology, 39*(1), 12.

Badura, A. S., Millard, M., Peluso, E. A., & Ortman, N. (2000). Effects of peer education training on peer educators: Leadership, self-esteem, health knowledge, and health behaviors. *Journal of College Student Development, 41*(5), 471.

Bandura, A. (1971). *Social Learning Theory*. New York: General Learning Press.

Ben, D. (2015). Big data and analytics in higher education: Opportunities and challenges. *British Journal of Educational Technology, 46*(5), 904–920. doi:doi:10.1111/bjet.12230

Berry, T., Cook, L., Hill, N., & Stevens, K. (2010). An exploratory analysis of textbook usage and study habits: Misperceptions and barriers to success. *College Teaching, 59*(1), 31–39. doi:10.1080/87567555.2010.509376

Big Data History and Current Considerations. (n.d.) Retrieved from https://www.sas.com/en_us/insights/big-data/what-is-big-data.html

Birjali, M., Beni-Hssane, A., & Erritali, M. (2018). A novel adaptive e-learning model based on big data by using competence-based knowledge and social learner activities. *Applied Soft Computing, 69*, 14–32. doi:10.1016/j.asoc.2018.04.030

Bond, C. (2018). Big data. *International Journal of Pharmacy Practice, 26*(1), 1–2. doi:10.1111/ijpp.12434

Burchfield, C. M., & Sappington, J. (2000). Compliance with required reading assignments. *Teaching of Psychology, 27*(1), 58.

Cai, Q., Lewis, C. L., & Higdon, J. (2015). Developing an early-alert system to promote student visits to tutor center. *The Learning Assistance Review, 20*, 61–72.

Campbell, J. P., & Oblinger, D. G. (2007). Academic anlaytics. Retrieved from https://www.educause.edu/ir/library/pdf/PUB6101.pdf

Chaurasia, S. S., & Rosin, A. F. (2017). From big data to big impact: Analytics for teaching and learning in higher education. *Industrial and Commercial Training, 49*(7/8), 321–328. doi:10.1108/ICT-10-2016-0069

Chaw, L. Y., & Tang, C. M. (2018). What makes learning management systems effective for learning? *Journal of Educational Technology Systems, 0*(0), 0047239518795828. doi:10.1177/0047239518795828

Chen, C. L., & Zhang, C.-Y. (2014). Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. *Information Sciences, 275*, 314–347. doi:10.1016/j.ins.2014.01.015

Chen, M., Mao, S., & Liu, Y. (2014). Big data: A survey. *Mobile Networks and Applications, 19*(2), 171–209. doi:10.1007/s11036-013-0489-0

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683–695. doi:10.1080/13562517.2013.827653

Clump, M. A., Bauer, H., & Breadley, C. (2004). The extent to which psychology students read textbooks: A multiple class analysis of reading across the psychology curriculum. *Journal of Instructional Psychology, 31*(3), 227.

Cornock, M. (2018). General data protection regulation (GDPR) and implications for research. *Maturitas, 111*, A1–A2. doi:10.1016/j.maturitas.2018.01.017

deNoyelles, A., & Raible, J. (2017). Exploring the use of e-textbooks in higher education: A multiyear study. *EduCause Review*.

deNoyelles, A., & Seilhamer, R. (2013). eTextbook access, usage, and beliefs: implications for adoption in higher education. *Journal of Applied Research in Higher Education, 5*(2), 189–201. doi:doi:10.1108/JARHE-12-2012-0065

Dobler, E. (2015). e-Textbooks: A Personalized learning experience or a digital distraction? *Journal of Adolescent & Adult Literacy, 58*(6), 482–491. doi:10.1002/jaal.391

DOMO. (2017). Data never sleeps 5.0. Retrieved from https://www.domo.com/learn/data-never-sleeps-5#/

DOMO. (2018). Data never sleeps 6.0. Retrieved from https://www.domo.com/learn/data-never-sleeps-6#/

Duncan Selby, R., P. Carter, K., & H. Gage, S. (2014). Survey concerning electronic textbooks. *International Journal of Sustainability in Higher Education, 15*(2), 142–156. doi:10.1108/IJSHE-03-2012-0022

Duval, E., & Verbert, K. (2012). Learning analytics. *E-learning and Education, 1*(8).

Ebook timeline. (2002, January 3). *The Guardian.* Retrieved from https://www.theguardian.com/books/2002/jan/03/ebooks.technology

Essa, A. (2016). A possible future for next generation adaptive learning systems. *Smart Learning Environments, 3*(1), 1–24. doi:10.1186/s40561-016-0038-y

Few, S. (2006). *Information dashboard design.* Burlingame, CA: Analytics Press.

Fike, D., Fike, R., & St. Clair, N. (2016). Do e-textbooks impact learning outcomes? *International Journal on E-Learning, 15*(3), 313–325.

Fox, E., & Riconscente, M. (2008). Metacognition and self-Regulation in james, piaget, and vygotsky. *Educational Psychology Review, 20*(4), 373–389. doi:10.1007/s10648-008-9079-2

French, M., Taverna, F., Neumann, M., Paulo Kushnir, L., Harlow, J., Harrison, D., & Serbanescu, R. (2015). Textbook use in the sciences and Its relation to course performance. *College Teaching, 63*(4), 171–177. doi:10.1080/87567555.2015.1057099

Gall, M. D., Gall, J. P., & Borg, W. R. (2007). *Educational research: An introduction* (8th ed.). New York, NY: Allyn & Bacon.

Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the literature 1991–2000. *Journal of College Student Development, 44*(6), 746–762. doi:10.1353/csd.2003.0066

Gholami, V., & Mostafa Morady, M. (2013). The effect of weekly quizzes on students' final achievement score. *International Journal of Modern Education and Computer Science, 5*(1), 36–41. doi:10.5815/ijmecs.2013.01.05

Gibson, C., & Gibb, F. (2011). An evaluation of second-generation ebook readers. *The Electronic Library, 29*(3), 303–319. doi:10.1108/02640471111141061

Greller, W., & Drachsler, H. (2012). Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society, 15*(3), 42–57.

Grönlund, Å., Wiklund, M., & Böö, R. (2018). No name, no game: Challenges to use of collaborative digital textbooks. *Education and Information Technologies, 23*(3), 1359–1375. doi:10.1007/s10639-017-9669-z

Gyllen, J., Stahovich, T., & Mayer, R. (2018). How students read an e-textbook in an engineering course. *Journal of Computer Assisted Learning, 34*(6), 701–712. doi:10.1111/jcal.12277

Hao, Y., & Jackson, K. (2014). Student satisfaction toward e-textbooks in higher education. *Journal of Science and Technology Policy Management, 5*(3), 231–246.

Harrison, M. (2016). Effect of prelecture quizzes on exam scores in a management course. *Journal of the Academy of Business Education, 17*, 257–269.

Hatteberg, S. J., & Steffy, K. (2013). Increasing reading compliance of undergraduates: An evaluation of compliance methods. *Teaching Sociology, 41*(4), 346–352. doi:10.1177/0092055x13490752

Hu, Y.-H., Lo, C.-L., & Shih, S.-P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469–478. doi:10.1016/j.chb.2014.04.002

Hung, J.-L., Wang, M. C., Wang, S., Abdelrasoul, M., Li, Y., & He, W. (2017). Identifying at-risk students for early interventions-a time-series clustering approach. *IEEE Transactions on Emerging Topics in Computing, 5*(1), 45–55. doi:10.1109/TETC.2015.2504239

Ifenthaler, D., & Widanapathirana, C. (2014). Development and validation of a learning analytics framework: Two case studies using support vector machines. *Technology, Knowledge and Learning, 19*(1), 221–240. doi:10.1007/s10758-014-9226-4

IMSGlobal (2019). Real-time, cross application educational data and analytics. Retrieved from http://www.imsglobal.org/initiative/real-time-cross-application-educational-data-and-analytics

Jang, D.-h., Yi, P., & Shin, I.-s. (2016). Examining the effectiveness of digital textbook use on students' learning outcomes in South Korea: A Meta-analysis. *The Asia - Pacific Education Researcher, 25*(1), 57–68. doi:10.1007/s40299-015-0232-7

Johnson, J. A. (2017). Ethics and justice in learning analytics. *New Directions for Higher Education, 2017*(179), 77–87. doi:10.1002/he.20245

Jokhan, A., Sharma, B., & Singh, S. (2018). Early warning system as a predictor for student performance in higher education blended courses. *Studies in Higher Education*, 1–12. doi:10.1080/03075079.2018.1466872

Junco, R., & Clem, C. (2015). Predicting course outcomes with digital textbook usage data. *The Internet and Higher Education, 27*, 54–63. doi:10.1016/j.iheduc.2015.06.001

Kerr, M. M., & Frese, K. M. (2017). Reading to learn or learning to read? Engaging college students in course readings. *College Teaching, 65*(1), 28–31. doi:10.1080/87567555.2016.1222577

Klašnja-Milićević, A., Ivanović, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education, 25*(6), 1066–1078. doi:10.1002/cae.21844

Kuh, G. D. (1995). The other curriculum: out-of-class experiences associated with student learning and personal development. *The Journal of Higher Education, 66*(2), 123–155. doi:10.1080/00221546.1995.11774770

Kurilovas, E., Zilinskiene, I., & Dagiene, V. (2015). Recommending suitable learning paths according to learners' preferences: Experimental research results. *Computers in Human Behavior, 51*, 945–951. doi:10.1016/j.chb.2014.10.027

Lieu, R., Wong, A., Asefirad, A., Shaffer, J. F., & Momsen, J. (2017). Improving exam performance in introductory biology through the use of preclass reading guides. *CBE—Life Sciences Education, 16*(3), ar46. doi:10.1187/cbe.16-11-0320

Liñán, L. C., & Juan Pérez, Á. A. (2015). Educational data mining and learning analytics: Differences, similarities, and time evolution. *Minería de datos educativos y análisis de datos sobre aprendizaje: diferencias, parecidos y evolución en el tiempo, 12*(3), 98–112. doi:10.7238/rusc.v12i3.2515

Liu, M., Kang, J., Zou, W., Lee, H., Pan, Z., & Corliss, S. (2017). Using data to understand how to better design adaptive learning. *Technology, Knowledge and Learning, 22*(3), 271–298. doi:10.1007/s10758-017-9326-z

Liu, M., McKelroy, E., Corliss, S. B., & Carrigan, J. (2017). Investigating the effect of an adaptive learning intervention on students' learning. *Educational Technology Research and Development, 65*(6), 1605–1625. doi:10.1007/s11423-017-9542-1

Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions? *Educational Psychologist, 32*(1), 1–19. doi:10.1207/s15326985ep3201_1

McGraw-Hill Education. (n.d.). Inclusive access. Retrieved from https://www.mheducation.com/highered/services/enterpriseandservices/inclusive-access.html

McGuire, P. (2018, April 12). Want more college students to graduate? Fix the high schools. *The Chronicle of Higher Education.* Retrieved from https://www.chronicle.com/article/Want-More-College-Students-to/243101

McKenzie, L. (2017, November 7). "Inclusive access" takes off. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/news/2017/11/07/inclusive-access-takes-model-college-textbook-sales

McKenzie, L. (2018, March 13). Europeand rules (and big fines) for American colleges. *Inside Higher Ed.* Retrieved from https://www.insidehighered.com/news/2018/03/13/colleges-are-still-trying-grasp-meaning-europes-new-digital-privacy-law

Millar, M., & Schrier, T. (2015). Digital or printed textbooks: Which do students prefer and why? *Journal of Teaching in Travel & Tourism, 15*(2), 166–185. doi:10.1080/15313220.2015.1026474

Mulholland, E., & Bates, J. (2014). Use and perceptions of e-books by academic staff in further education. *The Journal of Academic Librarianship, 40*(5), 492–499. doi:10.1016/j.acalib.2014.05.018

Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society, 17*(4), 49–64.

Park, Y., & Jo, I.-H. (2015). Development of the learning analytics dashboard to support students' learning performance. *Journal of Universal Computer Science, 21*(1), 110–133. doi:10.3217/jucs-021-01-0110

Philip, G. C., & Moon, S.-Y. (2013). An investigation of student expectation, perceived performance and satisfaction of e-textbooks. *Journal of Information Technology Education: Innovations in Practice, 12*, 287–298. doi:10.28945/1903

Phillips, D. C. (1995). The good, the bad, and the ugly: The many faces of constructivism. *Educational Researcher, 24*(7), 5–12. doi:10.2307/1177059

Pistilli, M. D., & Arnold, K. E. (2010). In practice: Purdue signals: Mining real-time academic data to enhance student success. *About Campus, 15*(3), 22–24. doi:10.1002/abc.20025

Poljicanin, A., Carić, A., Vilović, K., Kosta, V., Marinović Guić, M., Aljinović, J., & Grković, I. (2009). Daily mini quizzes as means for improving student performance in anatomy course. *Croatian Medical Journal, 50*(1), 55–60.

Rajiv Sunil, J., & Jhangiani, S. (2017). Investigating the perceptions, use, and impact of open textbooks: A survey of post-secondary students in British Columbia. *International Review of Research in Open and Distance Learning, 18*(4).

Renaud, K., & Shepherd, L. (2018). GDPR: its time has come. *Network Security, 2018*(2), 20. doi:10.1016/S1353-4858(18)30017-5

Reynolds, R. (2011). Trends influencing the growth of digital textbooks in US higher education. *Publishing Research Quarterly, 27*(2), 178–187. doi:10.1007/s12109-011-9216-5

Roberts, L. D., Howell, J. A., & Seaman, K. (2017). Give me a customizable dashboard: Personalized learning analytics dashboards in higher education. *Technology, Knowledge and Learning, 22*(3), 317–333. doi:10.1007/s10758-017-9316-1

Rockinson-Szapkiw, A. J., Courduff, J., Carter, K., & Bennett, D. (2013). Electronic versus traditional print textbooks: A comparison study on the influence of university students' learning. *Computers & Education, 63*, 259–266. doi:10.1016/j.compedu.2012.11.022

Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 3*(1), 12–27. doi:10.1002/widm.1075

Scholes, V. (2016). The ethics of using learning analytics to categorize students on risk. *Educational Technology Research and Development, 64*(5), 939–955. doi:10.1007/s11423-016-9458-1

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380–1400. doi:10.1177/0002764213498851

Simon-Campbell, E., Loria, & Phelan, J. (2016). Effectiveness of an adaptive quizzing system as an institutional-wide strategy to improve student learning and retention. *Nurse Educator, 41*(5), 246–251. doi:10.1097/NNE.0000000000000258

Slade, S., & Prinsloo, P. (2013). Learning analytics: Ethical issues and dilemmas. *American Behavioral Scientist, 57*(10), 1510–1529. doi:10.1177/0002764213479366

Special Issue on Educational Big Data and Learning Analytics. (2018). *Interactive Learning Environments, 26*(5), 712–715. doi:10.1080/10494820.2018.1472430

Subba Rao, S. (2001). Familiarization of electronic books. *The Electronic Library, 19*(4), 247–256. doi:10.1108/02640470110402045

Sun, Q., Norman, T. J., & Abdourazakou, Y. (2018). Perceived value of interactive digital textbook and adaptive learning: Implications on student learning effectiveness. *Journal of Education for Business, 93*(7), 323–331. doi:10.1080/08832323.2018.1493422

Tampke, D. R. (2013). Developing, implementing, and assessing an early alert system. *Journal of College Student Retention: Research, Theory & Practice, 14*(4), 523–532. doi:10.2190/CS.14.4.e

Terpend, R., Gattiker, T. F., & Lowe, S. E. (2014). Electronic textbooks: Antecedents of students' adoption and learning outcomes. *Decision Sciences Journal of Innovative Education, 12*(2), 149–173. doi:10.1111/dsji.12031

Tinto, V. (1975). Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research, 45*(1), 89–125. doi:10.2307/1170024

Tinto, V. (2012). *Leaving college: rethinking the causes and cures of student attrition* (2nd ed.). Chicago, IL: University of Chicago Press.

Tonidandel, S., King, E. B., & Cortina, J. M. (2018). Big data methods: Leveraging modern data analytic techniques to build organizational science. *Organizational Research Methods, 21*(3), 525–547. doi:10.1177/1094428116677299

Trowler, V. (2010, November). *Student engagement literature review*. Heslington, York, UK: The Higher Education Academy. Retrieved from https://www.heacademy.ac.uk/system/files/studentengagementliteraturereview_1.pdf

Van Horne, S., Russell, J.-e., & Schuh, K. L. (2016). The adoption of mark-up tools in an interactive e-textbook reader. *Educational Technology Research and Development, 64*(3), 407–433. doi:10.1007/s11423-016-9425-x

Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: a survival analysis approach. *Higher Education, 76*(5), 903–920. doi:10.1007/s10734-018-0249-y

Wagner, B. J., Ashurst, J. V., Simunich, T., & Cooney, R. (2016). Correlation between scores on weekly quizzes and performance on the annual resident in-service examination. *The Journal of the American Osteopathic Association, 116*(8), 530–534. doi:10.7556/jaoa.2016.106

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques*. Thousand Oaks, CA: Sage.

Weisberg, M. (2011). Student attitudes and behaviors towards digital textbooks. *Publishing Research Quarterly, 27*(2), 188–196. doi:10.1007/s12109-011-9217-4

Weng, C., Otanga, S., Weng, A., & Cox, J. (2018). Effects of interactivity in E-textbooks on 7th graders science learning and cognitive load. *Computers & Education, 120*, 172–184. doi:10.1016/j.compedu.2018.02.008

Wertsch, J. V., & Tulviste, P. (1992). L. S. Vygotsky and contemporary developmental psychology. *Developmental Psychology, 28*(4), 548-557. doi:10.1037/0012-1649.28.4.548

West, D., Henk, H., Heath, D., Lizzio, A., Toohey, D., Miles, C., . . . Bronnimann, J. (2016). Higher education teachers' experiences with learning analytics in relation to student retention. *Australasian Journal of Educational Technology, 32*(5), 48–60. doi:10.14742/ajet.3435

Wintrup, J. (2017). Higher education's panopticon? Learning analytics, ethics and student engagement. *Higher Education Policy, 30*(1), 87–103. doi:10.1057/s41307-016-0030-8

Zacharis, N. Z. (2015). A multivariate approach to predicting student outcomes in web-enabled blended learning courses. *The Internet and Higher Education, 27*, 44–53. doi:10.1016/j.iheduc.2015.05.002

Zimmerman, B. J. (1986). Becoming a self-regulated learner: Which are the key subprocesses? *Contemporary Educational Psychology, 11*(4), 307–313. doi:10.1016/0361-476X(86)90027-5

Zimmerman, B. J. (1990). Self-regulated learning and academic achievement: An overview. *Educational Psychologist, 25*(1), 3–17. doi:10.1207/s15326985ep2501_2

APPENDIX: IRB Exemption Letter

LIBERTY UNIVERSITY.
INSTITUTIONAL REVIEW BOARD

January 4, 2019

Dustin Williams
IRB Application 3639: Predicting Student Success Using Digital Textbook Analytics in Online Courses

Dear Dustin Williams,

The Liberty University Institutional Review Board has reviewed your application in accordance with the Office for Human Research Protections (OHRP) and Food and Drug Administration (FDA) regulations and finds your study does not classify as human subjects research. This means you may begin your research with the data safeguarding methods mentioned in your IRB application.

Your study does not classify as human subjects research because it will not involve the collection of identifiable, private information.

Please note that this decision only applies to your current research application, and any changes to your protocol must be reported to the Liberty IRB for verification of continued non-human subjects research status. You may report these changes by submitting a new application to the IRB and referencing the above IRB Application number.

If you have any questions about this determination or need assistance in identifying whether possible changes to your protocol would change your application's status, please email us at irb@liberty.edu.

Sincerely,

[redacted]

G. Michele Baker, MA, CIP
Administrative Chair of Institutional Research
The Graduate School

LIBERTY
UNIVERSITY.
Liberty University | Training Champions for Christ since 1971