2019

Student Learning Management System Interactions and Performance via a Learning Analytics Perspective

Gina Maria Ricker
Walden University

Walden University

College of Education

This is to certify that the doctoral dissertation by

Gina Ricker

has been found to be complete and satisfactory in all respects,
and that any and all revisions required by
the review committee have been made.

Review Committee
Dr. Gladys Arome, Committee Chairperson, Education Faculty
Dr. Tianyi Zhang Ulyshen, Committee Member, Education Faculty
Dr. Jennifer Lapin, University Reviewer, Education Faculty

Chief Academic Officer
Eric Riedel, Ph.D.

Walden University
2019

Abstract

Student Learning Management System Interactions and Performance via a Learning Analytics Perspective

by

Gina Ricker

MS, Kaplan University, 2013

BA, University of Maryland, Baltimore County, 2011

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education, Technology and Design

Walden University

May 2019

Abstract

Enrollment in full-time, virtual, K-12 schools is increasing while mathematics performance in these institutions is lacking compared to national averages. Scholarly literature lacks research studies using learning analytics to better predict student outcomes via student learning management system (LMS) interactions, specifically in the low performing area of middle school mathematics. The theoretical framework for this study was a combination of Hrastinski's theory of online learning as online participation and Moore's 3 types of interactions model of online student behavior. The purpose of this study was to address the current research gap in the full-time, K-12 eLearning field and determine whether 2 types of student LMS interactions could predict mathematics course performance. The research questions were developed to determine whether student clicks navigating course content page(s) or the number of times a student accessed resources predicted student performance in a full-time, virtual, mathematics course after student demographic variables were controlled for. This quantitative study used archived data from 238 seventh grade Math 7B students enrolled from January 8th–10th to May 22nd–25th in two Midwestern, virtual, K-12 schools. Hierarchical regressions were used to test the 2 research questions. Student clicks navigating the course content pages were found to predict student performance after the effects of student demographic covariates were controlled for. Similarly, the number of times a student accessed resources also predicted student performance. The findings from this study can be used to advise actionable changes in student support, build informative student activity dashboards, and predict student outcomes for a more insightful, data-driven, learning experience in the future.

Student Learning Management System Interactions and Performance via a Learning Analytics Perspective

by

Gina Ricker

MS, Kaplan University, 2013

BA, University of Maryland, Baltimore County, 2011

Dissertation Submitted in Partial Fulfillment

of the Requirements for the Degree of

Doctor of Philosophy

Education Technology and Design

Walden University

May 2019

Dedication

It is with gratitude and passion that I dedicate this dissertation to my mother, who was taken by cancer before she was able to see my first college commencement. While she may not witness this culmination of my knowledge and hard work, I know she would be proud of my unrelenting spirit and indefatigable drive that has fueled my success. I will continue to approach future endeavors with such enthusiasm.

Acknowledgments

My family and friends deserve special recognition for supporting me along this journey and my never-ending quest for knowledge. Without them, I would not be the confident, strong, and determined human I am today. I would also like to thank my committee for their unquestionable patience and invaluable knowledge which carried me through the dissertation process. Lastly, I would like to acknowledge Dr. Alyssa Walters who provided me with consistent, genuine personal and professional advice, support, and objectivity when I needed it most.

To all of you, I have a deep appreciation for your support and investment in education research.

From my heart, I thank you all.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction to the Study

Introduction

eLearning has become a viable option, for not only higher education, but for full-time K-12 students over the last decade (Barbour, 2013; Choi, Walters, & Hoge, 2017; Curtis & Werth, 2015; Dixson, 2016; Liu & Cavanaugh, 2012; Lowes, Lin, & Kinghorn, 2015). Recently, virtual K-12 schools have been shown to perform worse than brick-and-mortar schools in the area of mathematics proficiency (Choi et al., 2017; Woodworth et al., 2015). With the movement toward virtual learning in K-12, learning analytics has been utilized as a research approach that can identify types of participation in the virtual classroom and their relationship with student outcomes (Goggins & Xing, 2016; Xing, Guo, Petakovic, & Goggins, 2015). Learning analytics is a method of deriving meaning from student data captured by the learning management system (LMS) to advise actionable changes in teacher instruction, interventions, and support as well as predict outcomes for a more comprehensive data-driven learning experience (Khalil & Ebner, 2016; Lu, Huang, Huang, & Yang, 2017; Siemens, 2013; Templaar, Rienties, & Nguyen, 2017; Xang et al., 2015).

One area of student data that researchers have focused on to successfully predict student performance is student LMS interactions as measured by trace data (Agudo-Peregrina, Iglesias-Pradas, Conde-González, & Hernández-García, 2014; Curtis & Werth, 2015; Goggins & Xing, 2016; Pazzaglia, Clements, Lavigne, & Stafford, 2016; Xing et al., 2015; You, 2015). Trace data are the pieces of information left behind in an LMS when a student navigates and interacts with an online course and its associated contents

(Martin, Nacu, & Pinkard, 2016). Because eLearning has an inherent gap in observing student learning and behavior face-to-face, measuring student LMS interaction by trace data can close that gap by presenting a picture of student activity and ultimately how it relates to and predicts student performance (Curtis & Werth, 2015; Liu & Cavanaugh, 2012; Xing et al., 2015).

A variety of learning analytics research measuring student LMS interactions utilizing trace data has been conducted in higher education settings (i.e., blended, online, and massively open online courses) to better predict student outcomes (Agudo-Peregrina et al., 2014; Cavanaugh, Hargis, & Mayberry, 2016; Liu & Cavanaugh, 2012). However, there is a distinct lack of studies regarding the examination of student LMS interaction via trace data in the K-12 virtual classroom predicting student performance and more specifically, mathematics courses at elementary and middle school grade levels (Curtis & Werth, 2015; Liu & Cavanaugh, 2012). Researchers in the field concur that more research, especially examining full-time K-12 learners, is required to confirm that student LMS interactions as measured by trace data have the capacity to predict student outcomes (Agundo- Peregina et al., 2014; Cavanaugh et al., 2016; Liu & Cavanaugh, 2012; Goggins & Xing, 2016).

Background

Mathematics performance is subpar nationwide; however, eLearning programs face added issues in which the distance between students and their teachers and peers produce a barrier to improving student performance (Curtis & Werth, 2015; Liu & Cavanaugh, 2012; National Center for Education Statistics, n.d.). Full-time K-12 eLearning has seen increased enrollment over the last decade (Curtis & Werth, 2015; Lowes, Lin, & Kinghorn, 2015). The vast majority of learning analytics research has previously examined colleges and universities, leaving an apparent gap in research utilizing student LMS interactions to better predict K-12 student outcomes. More to the point, in the low performing area of elementary and middle school mathematics, a learning analytics research approach could illuminate the relationship between student LMS interactions and student performance (Liu & Cavanaugh 2012; Lowes, 2014; Lowes et al., 2015).

The use of student LMS interactions to determine whether a predictive relationship exists with performance is an important research topic to properly support teachers and students learning in the virtual environment. An educational theoretical framework is necessary to draw meaning from student data for results to become actionable outcomes for stakeholders (Pardo, Han, & Ellis, 2017). In this study, I used Hrastinski's (2009) online learning as online participation theory and Moore's (1989) three types of interactions model as the theoretical foundation to evaluate student LMS interactions. This theoretical framework is widely used in the literature and reflects the variables and participants involved in the current research question and purpose (Joksimovic, Gasevic, Loughin, Kovanovic, & Hatala, 2015; Kim, Park, Yoon, & Jo, 2016; Xing et al., 2016).

Student LMS interactions typically are represented by proxy variables sourced from the trace data in an LMS that approximate student learning behavior (or participation) in the eLearning classroom (Kim et al., 2016; Wickens, 1972). In this study, I used the clicks a student made navigating lesson content pages in the LMS and the total number of times resources were accessed in the virtual classroom; these pieces of trace data were based on proxy variables being used in current peer-reviewed literature to assess the relationship student LMS interactions have with student performance (Agundo-Peregrina et al., 2014; Czerkawski, 2016; Khalil & Ebner, 2016: Miller, Soh, Samal, Kupzyk, & Nugent 2015). In this research study, the trace data extracted from the LMS acted as a proxy for unobservable student behavior in the virtual learning environment. Bainbridge et al (2015) set forth that trace data captured by the LMS mirrors student performance, participation, and system use. Student LMS interactions are meaningful variables representative of larger student learning constructs vital to future research that wishes to provide a more comprehensive picture of student eLearning (Pardo et al., 2017).

Current research both supports and contradicts using student LMS interactions within learning analytics to predict student performance and provide an accurate picture of how students are learning and interacting with material online. However, the majority sees value in using a learning analytics approach that incorporates student LMS interactions of some kind along with other student assessment and qualitative data to obtain a nuanced understanding of student performance in the virtual environment. The results of this study contribute to the discussed body of literature by addressing the gap of investigating the relationship between student LMS interactions and K-12 eLearning mathematics course performance.

Problem Statement

The National Assessment for Educational Progress is delivered to fourth and eighth grade students nationally in mathematics and reading every 2 years. In 2015, the mathematics proficiency rate for eighth graders dropped to 33% of students scoring at or above the proficient level (National Center for Education Statistics, n.d.). With the growing K-12 eLearning population, there is a gap in research that addresses the utilization of learning analytics to better predict student outcomes, specifically in the low performing area of elementary and middle school mathematics via student LMS interactions. Joksimovic et al. (2015) asserted that the literature indisputably agrees that research on student LMS interactions is vital to support positive learning outcomes. While I could not locate any studies about full-time K-12 middle school students' mathematics performance via student LMS interactions, multiple recent studies have successfully endeavored to link student performance to the use of student LMS interactions in order to better support instruction and inform intervention strategies to increase student learning at the high school, undergraduate, and graduate levels (Agundo-Peregrina et al., 2014; Akcapinar, Altun, & Askar, 2015; Goggins & Xing, 2016; Miller et al., 2015; Pazzaglia et al., 2016; Xing et al., 2015).

Purpose of the Study

The purpose of this study was to address the current research gap in the full-time K-12 eLearning field from a learning analytics approach and determine whether two types of student LMS interactions, as measured by student clicks navigating the LMS course content page(s) and the total number of times resources were accessed within the course modules, can predict mathematics course performance. I tracked student LMS interactions via Google page tracking, which operates concurrently to the proprietary LMS. A nonexperimental, quantitative methodology was used as the best approach for this predictive study. Next, I used two, separate, hierarchical linear regression analyses to determine whether student LMS interactions, as measured by student clicks within the LMS course page(s) and the total number of resources accessed within the course modules, predict student final course performance.

Research Questions and Hypotheses

Research Question 1: Can student clicks navigating the LMS course content page(s) predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

$H_0$1: Student clicks navigating the LMS course content page(s) cannot significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

$H_1$1: Student clicks navigating the LMS course content page(s) can significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

Research Question 2: Can the number of times resources were accessed within the course modules predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

$H_0 2$: The number of times resources were accessed within the course modules cannot significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

$H_1 2$: The number of times resources were accessed within the course modules can significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

Theoretical Framework

The theoretical framework I used for this study was a combination of Hrastinski's (2009) theory of online learning as online participation and Moore's (1989) interaction model: three types of interactions. Hrastinski's theory filled a well-known gap in the new field of eLearning that previous seminal learning theories did not sufficiently address. The author asserted online participation is not a solitary metric of quantitative measure but an amalgamation of a myriad of online interaction points, citing the use of primary level quantitative interaction measures of frequency and duration inside of online learning systems as one means of measuring student interaction at its most basic level. Various recent learning analytics research studies have applied the theory, confirming its appropriateness as part of the framework for this research study (Iglesias-Pradas, Ruiz-De-Azcarate, & Agudo-Peregrina, 2015; Kim et al., 2016; Xing et al., 2015).

Moore (1989), whose model was a building block for Hrastinski's theory, stated that interaction in online education is a useless term without the definition of three distinct types: learner-learner interactions, learner-instructor interactions, and learner-content interactions. In this study, I focused on Moore’s learner-content interaction. Moore distinguished this interaction as “a defining characteristic of education” (p. 1). I depict how Moore’s interaction and Hrastinski’s theory fit into the structure of this research study in Figure 2 in Chapter 2. This framework directly supported my examination of student LMS interactions and their ability to predict student performance in this study; because the LMS is where students access the course content, quantitative measures derived from it can be used to measure student interaction.

Nature of the Study

The nature of this study was quantitative. While qualitative methods have been applied in previous research on this topic, the majority of student LMS interaction research is quantitative and prediction focused (Purarjomandlangrundi, Chen, & Nguyen, 2016; Strang, 2017). This approach aligned with utilizing trace data to elucidate the impact of student LMS interaction on student performance in the virtual classroom (Khalil & Ebner, 2015). Multiple researchers have used temporal data, such as time-on-task or duration and frequency of student LMS logins, as their trace data variables in assessing the relationship to and prediction of student outcomes (Akcapinar et al., 2015; Cavanaugh et al., 2016; Firat, 2016; Goggins & Xing, 2016; Liu & Cavanaugh, 2012; Pazzaglia et al., 2016; Strang, 2016b; You, 2015). The proprietary LMS used for this research study utilized Google page tracking to capture the number of student clicks navigating through course lesson content pages in the modules. This piece of trace data acted as an equivalent, or proxy, to student activity within the LMS; this approach aligned with the ones taken in other studies seeking to predict student performance via student mouse clicks (Agundo-Peregrina et al., 2014; Akcapinar et al., 2015; Martin et al., 2016). The second chosen trace data variable, total times resources were accessed by a student within the course modules, has been identified by multiple researchers as a valuable measure of trace data related to student performance in the virtual classroom (see Coker, 2015; Khalil & Ebner, 2016; Miller et al., 2015; Strang, 2016a). Resources in the mathematics course for Grade 7 can range from the resource packet, hyperlinks to external sites, course material pages, and the glossary (see Appendix B); although, the specific resource accessed cannot be differentiated between in the trace data. The quantitative approach assisted me in identifying which predictors, either student clicks navigating lesson content pages or student access to resources, are significantly predictive of student final course scores when demographic variables are controlled for. As expressed by the literature, quantitative methodologies are most appropriate for measuring the predictive power of trace data concerning student final course scores (Khalil & Ebner, 2016).

In this study, I sourced data from a full-time, Grade 7 Mathematics course available in two, separate, K-12 virtual charter schools in the United States. While the two virtual schools are locally owned and operated within the state by charter authorities or boards, they both use the same proprietary LMS and curriculum. The course ID is the same in the LMS for both school locations. A single course and curriculum that appears for seventh grade students delivered via the proprietary LMS provided the data used for this study. A homogeneous group of participants included students who are on grade-level (i.e., Grade 7 students in Grade 7 Mathematics).

Despite the entire target population being used, I conducted an a priori power analysis using G*Power software by selecting an *F* test, multiple linear regression: fixed model, R^2 increase, at $a = 0.05$ level with a moderate (.15) effect size. The result indicated a minimum of 55 student participants would need to be included in the study (see Appendix A); the target population exceeded this minimum. The power analysis output reflects each research question having a single test predictor and two covariates for a total of three predictors in each hierarchical regression. I statistically analyzed each research question independently of one another in their own hierarchical regression. Figure 1 displays how Research Question 1 in this research study translates to the hierarchical regression with one test predictor and two covariates for a total of three predictors.

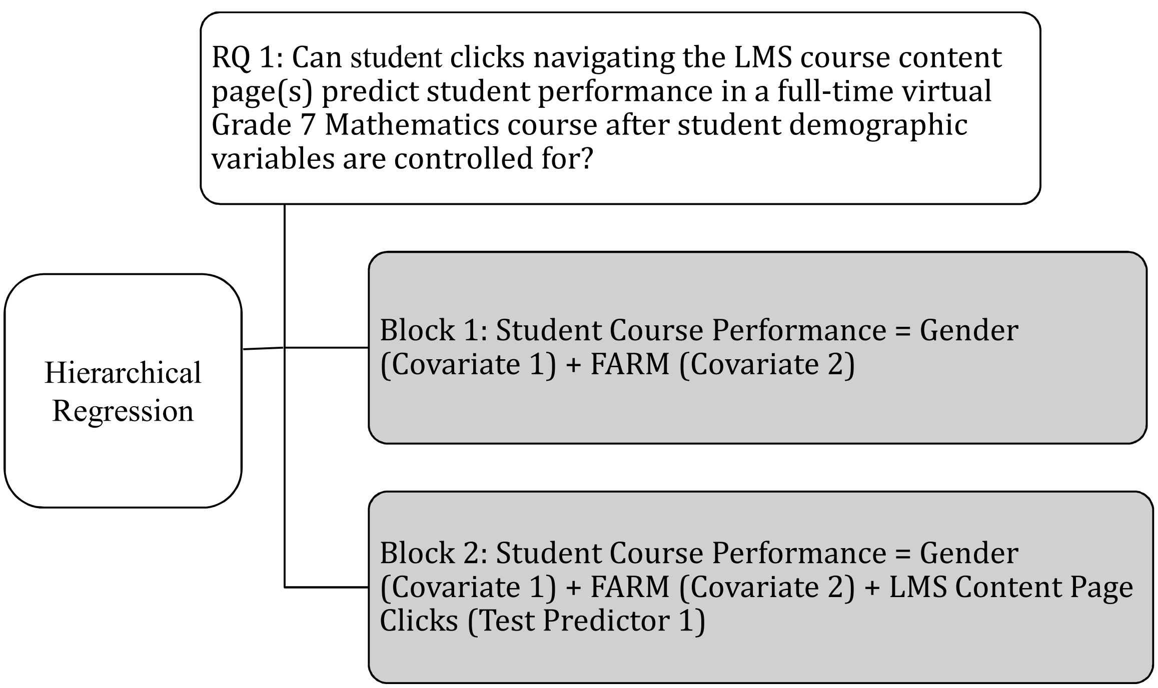


Figure 1. Hierarchical regression predictors.

I collected trace data from the proprietary LMS through Google page tracking used by the two virtual charter schools participating in the study. Trace data included the number of student mouse clicks made within the virtual classroom to navigate through the lesson page(s) and the total number of resources accessed within the course itself through the "Virtual Backpack" link (see Appendix B). The Virtual Backpack includes a glossary and course materials for students to refer to throughout the course. Student performance was the students' numerical final score in the Grade 7 Mathematics course (on a scale from 0 to 100). Student demographic information is frequently included as covariates in research regarding online education (Strang, 2017). Student demographics, such as gender and free and reduced meal (FARM) status, have demonstrated sizeable impact on mathematics performance (Tempelaar, Rienties, & Giesbers, 2015; Yarbrough, Cannon, Bergman, Kidder-Ashley, & McCane-Bowling, 2016). Similar to Liu and Cavanaugh (2012), I included demographic data (i.e., student FARM status and gender) in the hierarchical linear regression as covariates to control for their known impact on student academic performance.

Definitions

Full-time: Students who are enrolled at the institution as their primary source of public education, taking all essential courses and electives (Curtis & Werth, 2015).

Virtual K-12 school: A fully online public or charter school that serves kindergarten through 12th grade (Curtis & Werth, 2015). The terms *virtual* and *eLearning* are used interchangeably in this study.

Learning management system (LMS): The system where students and caretakers access courses and course content via the Internet and interact with teachers and peers (Firat, 2016). In this study, this term is used to refer to the proprietary LMS used by the study site schools.

Student performance: In this study, the term was defined by numerical final course score scaled from 0–100. Student performance is interchangeably used with the term *student outcome*.

Student LMS interaction: In this study, this term was defined as student mouse clicks navigating lesson content pages in course within the proprietary LMS. Additionally, student LMS interaction was also represented by student access to course resources within the virtual classroom via the "Virtual Backpack" link.

Assumptions

There were various assumptions made in this study. One assumption I made was that student LMS interactions are an accurate representation of student click behavior. I assumed that the data obtained from the Google page tracking was made by the student and not another member of their household or person with access to the student's computer/login information. Another assumption was that the student's final course grades were entered accurately and reasonably assigned. I also assumed that student LMS interactions were not representative of overall student learning. Additionally, it was not possible to know when students printed material and worked offline, which would contribute to their overall learning and final course scores. I assumed that the data obtained were representative of the population examined in this study. Assumptions related to the specific statistical analyses will be addressed further in Chapters 3 and 4.

Scope and Delimitations

The scope of this study was delimited to Grade 7 Mathematics students enrolled at two, Midwestern, full-time, K-12 virtual schools, who were present on Day 1 of the course. The Mathematics 7 course were delivered entirely online. The scope of the study included data from the B section of the course completed from January 8–10, 2018 to May 22–25, 2018. In an effort to obtain a homogeneous group of participants, I further delimited the study data to students who were on grade-level and completed the course to receive a grade. Students who received a final course score over 100 due to extra credit completed offline were removed from the data.

Limitations

The proprietary LMS does not have an activity log built into it like larger systems, such as BlackBoard and Moodle. The supplementary attachment of Google page tracking analytics allowed for some student LMS interaction data to be captured for analysis and meeting state standards for tracking eLearning student attendance and participation. The study was limited by what data could be extracted from Google page tracking to accurately reflect student LMS interactions congruent with what has previously been utilized in eLearning research to predict student performance. However, this data did not directly translate to what other LMSs capture and impacted the generalization of the findings. Additionally, because this course was taken in two, separate, Midwestern state schools, the inherent difference in teachers and school climate and culture were considered in the generalizability of the results.

Researcher bias is a consistent potential limitation to research studies. This limitation was less present in the study herein because the data were objectively collected in Google page tracking from my workplace. While the data were collected objectively via the LMS and I did not have reservations or expectations regarding the outcomes of this study, I was familiar with the nature of the data. I worked with the doctoral committee to ensure appropriate objectivity when the data results for this study were analyzed in order to remain as objective and unbiased as possible.

Significance

Learning analytics has demonstrated the potential to contribute data-driven decision making to early warning systems, student dashboards, and student interventions (Agundo-Peregrina et al., 2014; Akcapinar et al., 2015). The results of this research study contributed to a lack of information illuminating which student LMS interactions to highlight as predictors of student performance in order to better inform LMS early warning systems and dashboards (see Agundo-Peregrina et al., 2014; Akcapinar et al., 2015; Joksimovic et al., 2015). Joksimovic et al. (2015) found that the frequency and duration of student LMS interaction had a significant positive impact on student final course scores. In a growing area of education where observing students is nearly nonexistent, student LMS interactions show continuing promise in revealing information about student activity in the online classroom which impacts outcomes (Coker, 2015).

Summary

Full-time, K-12 eLearning, as expressed in the literature, is a growing education option for students (Barbour, 2013; Curtis & Werth, 2015; Liu & Cavanaugh, 2011, 2012; Lowes et al., 2015). As such, further research concerning factors that impact student success in the area of mathematics is required (Liu & Cavanaugh, 2012). More specifically still, examination of student LMS interactions is necessary to provide consistent information about activity in the online classroom that impacts student performance (Agundo-Peregrina et al., 2014; Cavanaugh et al., 2016). Towards this effort, learning analytics has been accepted as the approach by which this information can be elucidated to provide data-driven student support, interventions, and feedback (Martin et al., 2016; Martin & Ndoye, 2016). With this study, I sought to contribute information regarding the predictive capability of middle school mathematics student LMS interactions on performance to the field of eLearning education. I will discuss current, peer-reviewed contributions to the field of learning analytics and eLearning education research related to this research study in more depth in Chapter 2.

Chapter 2: Literature Review

Introduction

Full-time, K-12 eLearning has been expanding with rapid enrollment over the last decade (Curtis & Werth, 2015; Lowes et al., 2015). While mathematics performance is low nationwide, eLearning faces additional challenges in which the inherent distance between students and teachers adds an extra obstacle to increasing student performance (Curtis & Werth, 2015; Liu & Cavanaugh, 2012; National Center for Education Statistics, n.d.). Considering the amount of learning analytics research done on higher education, there is a clear gap in research utilizing student LMS interactions to better predict student outcomes, specifically in the low performing area of elementary and middle school mathematics, via a learning analytics approach (Liu & Cavanaugh 2012; Lowes, 2014; Lowes et al., 2015).

In this chapter, I will justify the use of clicks navigating lesson content pages in the LMS and the total number of times resources were accessed in the virtual classroom based on current, peer-reviewed literature despite there not being current agreement in the field on what student LMS interactions are most pertinent to predicting student performance (Agundo-Peregrina et al., 2014; Czerkawski, 2016; Khalil & Ebner, 2016: Miller et al., 2015). Several research studies have deemed the specific student LMS interactions chosen for this study as low-level "breadcrumbs" a student can leave behind in an LMS, which may not be interpreted meaningfully on their own (Lara et al., 2014; Rientes & Toetenel, 2016; Strang, 2016b). However, these perceptions appear to be in the minority in the body of literature. Student LMS interactions are not necessarily representative of genuine learning but can be used to provide a clear picture of student behavior impacting student performance (Kim et al., 2016; Percell, 2016). Winnie (2017) stated that these “fine-grained” pieces of student LMS interactions merit further focused exploration in the research seeking to predict student outcomes. Additional findings suggested that student LMS interaction data sets can indeed provide meaningful insight into the relationship between online behavior and student performance (Lowes et al., 2015). Pardo et al. (2017) envisioned that the student LMS interaction variables would continue to be used in meaningful combinations in future research to provide a better, more nuanced, picture of student learning.

In this chapter, I will delve further into the current research that both supports and contradicts student LMS interactions and their ability to predict student performance to provide an accurate picture of how this study fits into the landscape of literature. In this chapter, I will also delineate how extant research was located through detailed literature search strategies to reach saturation on the topic of relevant variables to the study and the theoretical and methodological approach used. Specific research that has direct or congruent relevance to the underresearched population of full-time, K-12 eLearners was used to provide a current picture of learning analytics and student performance prediction to better frame how the results of the current inquiry contribute to the present body of research.

Literature Search Strategy

To authenticate the importance of predicting student performance in mathematics via trace data, in this literature review I explored multiple scholarly sources to provide credibility to this study. The review began with a comprehensive search through various databases and search engines, including Academic Search Complete, Computers and Applied Sciences Complete, Dissertations & Theses at Walden University, Education Source, ERIC, ProQuest Dissertations & Theses Global, PsycARTICLES, PsycINFO, SAGE, ScienceDirect, and Google Scholar in combination with Ulrich's publications directory to determine peer-review delineation. The key search terms that I used to locate relevant, peer-reviewed literature to the study topic included:

- Environment terms: *eLearning, virtual, online, distance,* and *classroom*;
- Subject terms: *Learning, education, LMS,* and *mathematics*;
- Subject terms: *K-12, high school, student, learner, full-time,* and *CSCL*;
- Predictor terms: *trace data, log data, participation, engagement, interaction,* and *behavior*;
- Dependent variable terms: *Performance, achievement,* and *outcomes*; and
- Method terms: *Analytics, prediction, regression,* and *model*.

The use of learning analytics to predict student performance is still a novel approach with the majority of research appearing within the last 5 to 8 years. I used a customized publication date range of 2014–2018 to locate the most recent editorials and peer-reviewed materials. Backward citation chaining was used, where possible and applicable, to gain further insight into earlier works. Additionally, K-12 full-time eLearning is a small, but growing, portion of education offerings; the vast majority of learning analytics research is centralized in undergraduate education (Lowes et al., 2015;

Purarjomandlangrudi et al., 2016). I used these empirical resources as credible representations for this research study.

Theoretical Foundation and Framework

In order to provide a foundation with which to better understand student LMS interactions, I used the online learning as online participation theory (Hrastinski, 2009) and the accompanying three types of interactions model (Moore, 1989) as the theoretical framework for this study. Pardo et al. (2017) stated that an educational theory or framework is required to draw meaning from the student digital footprint in order for actionable outcomes to result from learning analysis. In the traditional, educational, brick-and-mortar environment, student interaction with education materials and activities as they relate to student outcomes are directly observable (Coker, 2015). However, with the increased enrollment in various types of eLearning programs, there was a need in the field to have theories that applied specifically to the learning environment that exhibits a distinct temporal and spatial gap between student and teacher preventing direct observation of student behavior (Martin & Ndoye, 2016; Purarjomandlangrudi et al., 2015). Figure 2 displays the connections between the focus of this study, student LMS interactions, and the theory and model used to frame the participants and variables chosen.

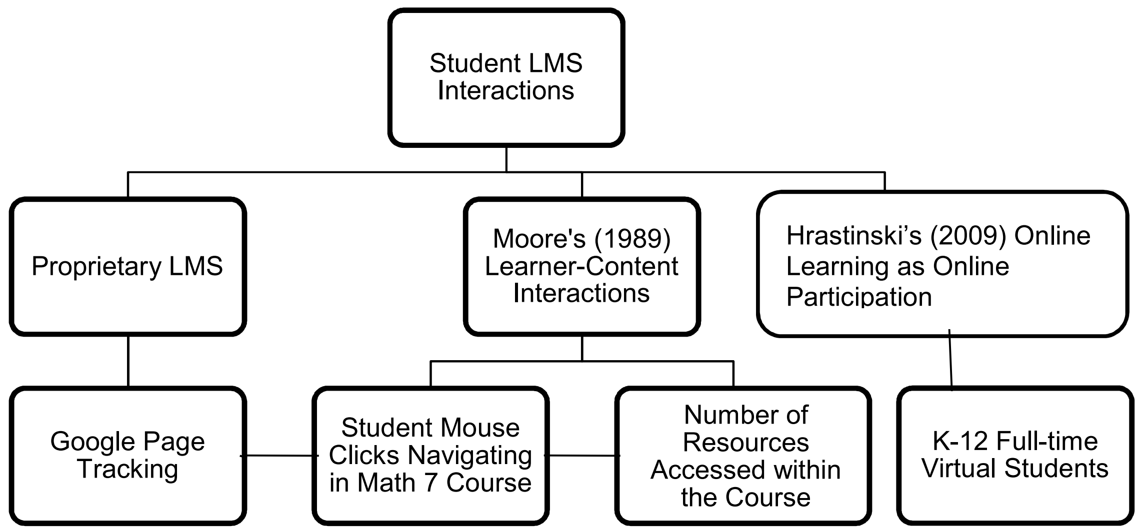


Figure 2. Theoretical framework.

Hrastinski (2009) developed the online learning as online participation theory in which the complexity of online interactions were expressed to represent student learning. Hrastinski stated "online learner participation (1) is a complex process of taking part and maintaining relations with others, (2) is supported by physical and psychological tools, (3) is not synonymous with talking or writing, and (4) is supported by all kinds of engaging activities" (p. 81). Hrastinski's second, third, and fourth point indicated that student LMS interactions are representations of student participation in the learning process, connecting them to student outcomes. Similar to my approach in this study, Xing et al. (2015) applied Hrastinski's theory to build prediction models that utilize student trace data to quantify student participation and predict student performance. Additionally, recent research has examined the capability of learning analytics to predict student participation and engagement through the lens of online learning as online participation (Iglesias-Pradas et al., 2015). Kim et al. (2016) elaborated upon Hrastinski's theory, stating that active participation was a measure of student engagement known to positively impact student learning by creating a model from trace data captured by the LMS and representative of active participation in the virtual classroom to predict low and high achievers. Congruent to the scholarly work mentioned, in this study, I applied this theory in such a way that demonstrates student LMS interactions are representative of participation known to impact student performance.

In the theory of online learning as online participation theory, Hrastinski (2009) referenced Moore's (1989) model of three interactions for distance learning to build the theory around areas of student interactions. In this study, I drew upon Moore's model to frame specific student-LMS interactions impacting student performance. The three interactions outlined in the model are learner-learner, learner-instructor, and learner-content (Moore, 1989). Moore asserted that in order to better understand how and what interactions impact distance education outcomes, the distinction between them was crucial. For the purposes of this study, my focus was on learner-content interaction, with content being presented through a virtual classroom housed in an LMS. Student LMS interactions, such as navigating through course lesson pages and accessing module materials in the virtual classroom, best represented the area of Moore's learner-content interaction distinction.

Moore's model has been used and expanded upon by researchers to frame LMS trace data and quantify the three interactions to demonstrate a significant relationship between them and student final course score in eLearning (Joksimovic et al., 2015; Purarjomandlangrudi et al., 2016; Rientes & Toetenel, 2016). Joksimovic et al. (2015) expanded upon Moore's model by adding learner-system interactions, which are aligned more closely with the research questions of this study. Joksimovic et al. found that learner-system activity positively impacted student outcomes, while learner-content interactions demonstrated a significant negative relationship with student outcomes. Both interactions were found to have significant relationships with undergraduate final course scores and deemed necessary in future learning analytics research (Joksimovic et al., 2015). Strang (2016a) also echoed the four most commonly used student activities in learning analytics research as being student-student, student-content, student-teacher, and student-system. While learner, or student-system interactions, are most reflective of my aim with this study, it was not part of Moore's original model and not directly used in the theoretical framework herein.

As evidenced by the literature, the theories previously discussed demonstrate a justification to use trace data to quantify student LMS interaction and determine its ability to predict student performance. While the theory and model share a common assumption that the majority of student interactions are captured by the LMS, there are opportunities where students will work offline. However, despite this limitation, measurable trace data has been widely accepted to represent student interactions in the virtual classroom (Joksimovic et al., 2015; Kim et al., 2016; Purarjomandlangrudi et al., 2016; Tempelaar et al., 2015). The results of this research expanded upon the current knowledge base surrounding the predictability of student performance by student LMS interactions via trace data by contributing findings from an underresearched population of full-time, virtual, middle school students.

Literature Review Related to Key Variables

eLearning

It is widely confirmed in the literature that eLearning is a growing choice for education from early learning to college and beyond (Greene & Hale, 2017; Gros & Garcia-Penalvo, 2016; Harris-Packer & Segol, 2015; Morgan, 2015; Woodworth et al., 2015). Additionally, it is a mode of learning that has been recognized as effective and efficient; the consistent narrative is that eLearning provides flexible and adaptive learning opportunities that are not widely available through traditional, brick-and-mortar environments (Gros & Garcia-Penalvo, 2016; Martin et al., 2016). The student profile originally associated with eLearning, or distance learning as it is occasionally synonymously referred to, beginning in the 1990s has shifted from working adults and occupational training to K-12 students, confirming the need for more research in the lower primary and secondary grades (Gros & Garcia-Penalvo, 2016). eLearning courses can be presented in either an asynchronous or synchronous format; these are defined as the students and teachers being online at varying times or at the same time, respectively, and contribute to the adaptive nature of learning online (Woodworth et al., 2015).

eLearning has been historically used to encompass distance learning, blended learning, computer-supported collaborative learning, and MOOC (Barbour, 2013; Gros & Garcia-Penalvo, 2016). Ultimately, it refers to education supplemented in some form by technologically delivered materials or courseware (Gros & Garcia-Penalvo, 2016). eLearning is still evolving and adopting new methodologies to increase engagement and performance outcomes as well as close the inherent gap created from the distance between educators and students (Reyna, 2016; Siemens, 2014). As such, efficacy research surrounding eLearning versus traditional, brick-and-mortar student performance has increased in recent years as the availability of eLearning programs has spread (Carpenter, Kafer, Reeser, & Shafer, 2015; Harris-Packer & Segol, 2015; Noesgaard & Ørngreen, 2015).

There are mixed results regarding eLearning program performance and efficacy; varying especially in different grade bands, student populations, subjects, and even geographic locations (Carpenter et al., 2015; Harris-Packer & Segol, 2015; Pazzaglia et al., 2016; Woodworth et al., 2015). Harris and Nikitenko (2014) found that undergraduate quantitative skills students who took their course online outperformed geographically matched students in the same program who took the course face-to-face. Researchers are still actively exploring what contributes to the differences in student performance enrolled eLearning programs (Woodworth et al., 2015). There is a lack of rigor and consistency with which eLearning programs are implemented and monitored (Greene & Hale, 2017; Morgan, 2015; Woodworth et al., 2015). Additionally, researchers have purported that a specific type of learner may be particularly successful in the online environment. Pardo et al. (2017) declared self-regulated learning research, or the extent that students are engaged in the online learning process through LMS trace data, was directly related to student performance; the findings provided valuable insight into student learning and eLearning program quality based on self-regulated learning student behavior. You (2015) also confirmed that self-regulated learning behavior is critical to online student success. Morgan (2015) similarly asserted that students with strong literacy and technology skills are often the students to succeed in the eLearning environment. The literature has demonstrated that there are leads as to which students perform best in the higher education eLearning setting, but little consistent insight into what types of students are most successful in the K-12 eLearning environment.

Much of the current literature surrounds higher education, leaving eLearning at the K-12 level underresearched. Particularly little attention has been given to student activity that would otherwise be observed in a traditional face to face classroom, despite rapidly growing eLearning program enrollment and links to student outcomes (Morgan, 2015). My study contributed to this body of literature by examining student LMS interactions as a reflection of student behavior that could contribute to student performance in the eLearning environment. Majority of the literature presented in this chapter surrounds the topic of eLearning performance, student behavior via LMS interactions, and expanding upon existing knowledge through the unexplored gap regarding the K-12 population.

State virtual schools or charters. The focus of this study was on the growing area of K-12, full-time, eLearning. Full-time, K-12, eLearning programs are sometimes referred to as cyber schools, virtual schools, or virtual charters (Molnar et al., 2017). K-12 enrollment in 2013-14 stood at 1.8 million students with 310,000 being full-time eLearners (International Association for K-12 Online Learning, 2013). However, in 2014-15, enrollment increased to approximately 2.7 million students enrolled in an online K-12 program that supplemented their brick-and-mortar education, with 278,511 students being enrolled full-time in 2015-16; the current enrollment in full-time K-12 programs is unknown (Barbour, 2017; Gemin, Pape, Vashaw, & Watson, 2015; Molnar et al., 2017). More specific to the population focused on in this study, state virtual schools are a subset of the aforementioned enrollment figures, and the largest provider of full-time, K-12, online learning (Gemin & Pape, 2017).

State virtual schools are more formally defined as accredited public institutions, who receive federal funding, and present material in a course-based format through computers or other technology devices connected to the Internet (Morgan, 2015). These schools are geared toward students that receive their entire prescribed course instruction online through a school sanctioned by the state, district, or individual charter (Barbour, 2017). However, reasons for K-12 eLearning student enrollment range from temporary credit recovery to long term curriculum accommodations, individualized instructions, and gifted accelerations, flexible scheduling, an escape from bullying, and student health concerns. For example, students in rural areas are given the opportunity to explore curriculum topics that may not otherwise be available to them at their brick-and-mortar school due to lack of resources and/or content expert staff (Michinov, Brunot, Le Bohec, Juhel, & Delaval, 2011; Morgan, 2015). Additionally, traveling families, child prodigies, or teenagers already in the workforce find eLearning programs most suitable for their lifestyle needs. Furthermore, student mobility or credit recovery can motivate a student to enroll in a K-12 eLearning program temporarily leading to an inherently different, highly mobile, demographic makeup compared to traditional brick-and-mortar institutions with more stable populations (Woodworth et al., 2015).

Virtual K-12 programs often have a higher mobility rate which leads to a unique demographic makeup that differs significantly from a traditional brick-and-mortar school and subsequently impacts student outcomes (Choi et al., 2017; Woodworth et al., 2015). There is very little peer-reviewed research examining the performance of full-time K-12 eLearners (Barbour, 2017). Carpenter et al. (2015) observed the performance of online K-12 schools versus that of brick-and-mortar in the state of Colorado and found statistically significant state assessment performance differences in favor of brick-and-mortar schools in nearly all of the cases. However, when the virtual schools were compared to demographically similar brick-and-mortar schools, only some statistically significant differences appeared. In regards to mathematics specifically, Choi et al., (2017) reported that virtual K-12 schools have typically demonstrated weak results in comparison to brick-and-mortar programs, with the performance gap widening further in the secondary grades. Harris-Packer and Segol (2015) also noted the generally lower performance of virtual K-12 programs, but instead found that there were some virtual programs which performed equal to or above their traditional counterparts. Due to the rate of enrollment for full-time K-12 eLearning programs, more research regarding factors impacting student performance is vital to understanding the intricacies specific to the unique environment and variability in learners.

Learning Analytics

Learning analytics is a developing academic field (Perrotta & Williamson, 2018). There are various definitions for learning analytics as it applies to education. Originally called "academic analytics", Pardo et al. (2017) stated that learning analytics seeks to improve the quality of eLearning through analyzing data reflective of student learning processes captured through the LMS. It is a field that draws from educational data mining, psychology, computer science, and adaptive learning (Chatti, Lukarov, Thus, Muslim, & Yousef, 2014). Siemens (2010) defined and shaped learning analytics into a research field early on in his blog as “the use of intelligent data, learner-produced data, and analysis models to discover information and social connections, and to predict and advise on learning” (p. 1). However, the most accepted definition currently used by The Society for Learning Analytics Research was constructed by Siemens and describes learning analytics as the “measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs” (Seimens & Long, 2011, p. 30). The ultimate goal of learning analytics is to provide data-driven student monitoring, prediction, intervention, feedback, adaptation, and actionable recommendations for educators and learners alike (Chatti et al., 2014).

Learning analytics is still a new and developing field in eLearning education that requires more research across populations and conditions (Czerkawski, 2016; Gunn, 2014; Siemens & Long, 2011; Strang, 2016a). It is a field that has the possibility to address the foremost issues facing teaching and learning in the online environment. Martin and Ndoye (2016) asserted that despite its novelty, researchers advocate strongly for scholarly learning analytics work that identifies students that are at risk and predicts student outcomes. Accordingly, Miller et al. (2015) stated that learning analytics had been used recently in peer-reviewed work to refer to analyzing student-level data with the specific intent to predict performance. However, Tempelaar, Rientes, and Nguyen (2017) stressed that learning analytics might be limited in function unless more substantial underlying characteristics of student activity, such as motivation and learning style, can be ascertained.

Similarly, Gasevic, Dawson, Rogers, and Gasevic (2016) found that it is imperative to determine if and how students are using an LMS prior to creating predictive models to better understand genuine student learning and subsequent outcomes. Perrotta and Williamson (2018) stated that unless longitudinal student data are examined, learning analytics can be presumptuous in that it produces data-driven metrics which lead to actionable interventions based on “temporary approximations” of student behavior (p. 7). Therefore, a distinct drawback to learning analytics is that it often requires a large amount of data and effort to be effective and may be unpractical and uneconomical for educators to employ without significant training (Choi, Lam, Li, & Wong, 2018).

Despite these cautions, I found that a large body of research provided empirical evidence affirming the potential of learning analytics in predicting education outcomes which align with the purpose and direction of this study (Conde & Hernandez-Garcia, 2015; Gasevic et al., 2016; Reyes, 2015). There are a variety of studies that have demonstrated learning analytics’ potential to benefit students in various education applications. Lu et al. (2017) conducted a study that determined students who received interventions based on learning analytics had improved learning outcomes and engagement over those students who received interventions based solely on teacher observation. Koc (2016) found that through the learning analytics approach, student participation in a blended course (as defined by virtual lecture attendance and discussion forum submission) were positively related to both final project and exam scores.

Learning analytics is an approach to further understand student LMS interactions captured in the online learning environment and determine student behaviors that have an impact on performance. There is policy support and desire for more research regarding K-12 learning analytics (Perrotta & Williamson, 2018). Yeung et al. (2017) examined student interactions with an eLearning mathematics tablet application to find that those who spent more time on the learning activities and made more attempts, also performed better than those students who had relatively low level of time and attempts in the application. Similarly, Greller, Santally, Boojhawon, Rajabalee, and Kevin (2017) found that after clustering students on regularity of involvement in the LMS (based on login to platform, submissions, self-assessment tests, forum participation, and resource access), a direct relationship was apparent between the level of regular activity in the LMS and student performance; this was reflected in exam and course grades. Although Mwalumbwe and Mtebe (2017) found that student discussion post and exercise submissions to the LMS, were strongly related to student performance, they discovered that time and frequency spent in the LMS were not. In the same study, the researchers also found that student discussion posts were the most predictive factor of student performance out of the student LMS interactions examined (Mwalumbwe & Mtebe, 2017). The contrast in learning analytics findings regarding student LMS interactions is ostensible; while there are distinct connections, the variability in what student LMS interactions are most salient in student performance is highly contingent upon multiple factors ranging from what LMS is being used, what content students are learning, and how students are accessing materials (Greller et al., 2017).

Lastly, an important issue to discuss surrounding the rise of learning analytics and the use of student data for education research are the ethics of student privacy. The overarching categories of student data used in learning analytics are demographic, behavioral, performance, financial and longitudinal (Adejo & Connolly, 2017). My study used three out of the five data categories mentioned: demographic, behavioral, and performance. The area of student privacy is not addressed frequently in learning analytics. However, the use of private data and information is a concern of students learning in the online environment (Adejo & Connolly, 2017). It is imperative that while student data are obtained for appropriate use in learning analytics, student information and privacy is protected from harm. The measures taken in this study to protect the student data used and student rights to privacy is addressed in subsequent chapters.

Learning Management System (LMS)

An LMS is a "computer-based system designed to assist instructors and learners in the management and administration of course dissemination and participation, particularly distributing course content and tracking student performance" (Martin et al., 2016, p. 44; Ullman & Rabinowitz, 2004). Gros and Garcia-Penalvo (2016) noted that the LMS is associated with the beginning of widespread eLearning adoption. However, adoption and integration of the LMS alone is extensive throughout all areas of education (Mwalumbew & Mtebe, 2017). In K-12 eLearning there are various LMS in use, such as Google Classroom, EdModo, Blackboard, and Pearson SuccessNext. While I examined data from a propriety LMS in this study, it shares many of the same features and flexibility that the aforementioned LMS possess. Teachers typically can personally connect with students, families, and other educators through the system. Additionally, they can modify materials and engage students with wikis and discussion forums (Martin et al., 2016). Recently, with the growth of eLearning, the LMS has had a lasting impact on the teaching-learning process. It is known that student learning and performance is contingent upon use of the LMS and its tools (Winnie, 2011).

These systems have been widely applied across grade bands and learning programs (Cerezo, Sanchez-Santillan, Paule-Ruiz, & Nunez, 2016). Learning analytics research is often dependent upon data sourced from the LMS regarding student interactions within it (Choi et al., 2018; Gasevic et al., 2016). However, not all students are successful in an LMS; depending on the construction of the system, it may require more skill, involvement, and energy from the student to gather information and learn content than they would otherwise exert in a traditional classroom (Cerezo et al., 2016). Student control over using the LMS and its associated tools should be more closely examined to better predict student performance. Differences in the construction of the variety of LMS employed today makes generalizing research findings difficult (Choi et al., 2018). Researchers request more work to be done in the field to elucidate student behavior patterns in the LMS to better understand the virtual learning environment and improve student outcomes (Cerezo et al., 2016; Gasevic et al., 2016).

Despite the rapid growth of K-12 eLearning, research using LMS data are lacking with K-12 eLearners, especially when compared to the amount of literature on higher education students (Lowes et al., 2015; Morgan, 2015). Lowes et al. (2015) stated that eLearning courses generate generous amounts of student LMS trace data from student actions in the virtual classroom that can be used to glean insight into how student online behavior impacts success. Unlike the physical classroom, student learning behaviors cannot directly be observed in an online classroom. The LMS has enabled the use of student learning data to enhance the education process, especially related to student LMS interaction trace data (Agundo-Peregrina et al., 2014; Khalil & Ebner, 2015). Ultimately, the usage of LMS generated student interaction data, especially in K-12, is underutilized in eLearning research which could lead to performance predictive insights crucial to lifting student outcomes (Agundo-Peregrina et al., 2014).

Proxy Variables and Trace Data

ELearning courses generate student LMS interactions, or trace data, based on student action in the online environment that can be used to assemble insight into how student online behavior influences success (Lowes et al., 2015). Trace data can be considered evidence, or representations, of student learning and interaction behaviors in the eLearning classroom. In essence, trace data can act as a proxy for otherwise unobservable student behavior in the virtual classroom. Bainbridge et al. (2015) stated that trace data captured by the LMS reflects student performance, participation, and system use. Student trace data are commonly used in learning analytics research to approximate student classroom behavior in the virtual environment (Kim et al., 2016; Lara et al., 2014). These “proxy variables” are often used in social science research when the direct observation or measurement of a conceptual variable is unfeasible (Kim et al., 2016; Wickens, 1972). Jo et al. (2015) discussed candidates of proxy variables in the LMS that reflect larger constructs of successful learning behaviors and characteristics of students; for example, trace data such as, total student login duration, regularity, and frequency could be reflective of student persistence and engagement.

Trace data, or pieces of data captured by an LMS to represent a specific behavior that is otherwise unobservable in the virtual classroom, demonstrate value in eLearning and learning analytics research because they operate not just as predictors, but indicators to educators about student learning behavior (Kim et al., 2016). Kim et al. (2016) were able to accurately predict student performance >70% of the time at the beginning of the course, and >90% of the time by the end of the course, using student LMS interactions as a proxy for measuring classroom learning behaviors such as, active participation, engagement, and interaction. The researchers concluded that student trace data could be used as a proxy for theoretical and empirically based student learning behaviors to create a prediction model for eLearning student performance.

Currently, there is no consensus on what variables able to be extracted from the LMS are the most salient in predicting student outcomes (Agundo-Peregrina et al., 2014; Khalil & Ebner, 2016; Miller et al. 2015). However, there is a consistent subset of trace data utilized in peer-review studies that are related to student performance and demonstrated they could predict at-risk students and student outcomes. Temporal data, such as time spent in the LMS or frequency of logins is a measure often used to define student LMS activity and approximate student engagement (Carver, Mukherjee, & Lucio, 2017; Jo et al., 2015; Martin et al., 2016). Trace Data, or interactions between the student and the LMS, are a popular metric used in learning analytics research. Khalil and Ebner (2015, p.1329) outlined trace data as, “mouse clicks, number of accessed resources, number of finished assignments, videos accessed, documents accessed, files downloaded, questions asked, discussions involved, and social network activities.” These metrics are considered “basic,” or fundamental low-level student LMS interactions (Agundo-Peregrina et al., 2014). Yeung et al. (2017) referred to them as “fine-grain” behaviors. Rientes and Toetenel (2016) asserted that simple metrics such as the number of page clicks will not bring much insight to learning analytics on their own. However, Percell (2016) stated that while page views, or clickstream data, may not be demonstrative of genuine learning, it can be used to provide a clearer picture of what student behavior is impacting learning and performance. You (2015) concurred that it is important when examining certain frequency behaviors, that they are representative of deliberate student learning in order to draw meaningful conclusions. Overall, researchers agree that trace data are a low-level measure of student LMS interaction; however, through more rigorous research, it can be used as a powerful informant relating to student performance in the eLearning classroom. More recent research and literature regarding the specific trace data used in this study will be discussed in depth in the subsequent sections.

Lesson Page Views and Number of Resources Accessed

The types of student LMS interaction variables used in learning analytics research has varied based on what is available for extraction and what the researcher deems important; this is considered a limitation in learning analytics research (Lowes, Lin, & Kinghorn, 2015). In the current study, I used lesson content page views and the total number of times the resources are accessed in a mathematics course available through a proprietary LMS for K-12 eLearning students. These are confirmed student LMS interactions, or trace data, in the current body of research (Khalil & Ebner, 2015). Because it is difficult for teachers to observe how students learn and behave in the virtual classroom, it is vital to examine these particular student LMS interactions to better understand, and ultimately predict student performance (Cerezo et al., 2016).

A large body of current research assessed the relationship between similar student LMS interactions to the two chosen for the study herein. Lara et al. (2014) found that the frequency with which undergraduate students visited course content and accessed resources in the virtual classroom, was positively correlated with their perseverance in an online course. Pardo et al. (2017) utilized LMS captured learner activities, or trace data, such as resources accessed and video playback, to determine how they contribute to the variance within student final course scores. The number of times a resource page was accessed was significantly related to the 145 undergraduate student final course scores. Lowes et al. (2015) found that when they examined frequency course behaviors such as the number of logins and the number of days in the LMS for 12 online high school courses with a total of 798 students, all were associated with higher final course scores. Uniquely, Cavanaugh et al. (2016) found that there appeared to be an intermediate login frequency and duration which resulted in the most desirable course performances. Too little or too much time in the LMS demonstrated a negative relationship with student performance.

However, not all empirical research produced consistent significant findings related to student LMS interactions. Martin et al. (2016) stated: “user trace log data may not always reflect intentional decisions to click or not click certain areas of the site.” (p. 54). Strang (2016b) did not find significant correlation or predictability between student trace data (lesson views and login frequency) and student course grades in the Moodle LMS. These results were delivered with caution because it was unknown how the trace data were captured and subsequently calculated in Moodle via analytics algorithms and may have impacted the interpretation of the findings.

Similarly, Templaar et al. (2015) did not find that basic LMS trace data were able to significantly predict student outcomes on their own when examining a blended undergraduate course. They found formative assessment data to be most predictive. However, if real-time interventions are needed, and formative assessment data are unavailable, student LMS interactions were considered valuable feedback as a second option. More research in other LMS software and with other populations is required to solidify these findings (Strang, 2016a). Further recent research and literature related to ascertaining the relationship between trace data variables and student performance are discussed in the subsequent sections of this chapter.

Student Demographics

Demographics are commonly included as predictors in online education research (Strang, 2017). Student demographic variables have been shown to have a considerable impact on learning mathematics (Tempelaar et al., 2015; Yarbrough et al., 2016). Specific to the area of mathematics, Yarbrough et al. (2016) delineated a multitude of studies that identified gender differences in academic performance related to the subject area. Additionally, the FARM status of a student is a direct reflection of socioeconomic status; it also has been demonstrated to impact students academically due to the restraints it imposes upon student needs (Khalil & Ebner, 2011; Marchetti, Wilson, & Dunham, 2016).

Although learning analytics is still in its infancy, as evidenced by current research, it typically includes demographic variables alongside student LMS interaction data as it is germane to student performance (Koc, 2017). Shrader, Wu, Owens, and Santa Ana (2016) successfully included student demographic characteristics age, sex, education level, and employment status in their mixed methods study of student LMS activity and satisfaction in a MOOC course to show that older students used the system more actively while gender did not affect usage. Miller et al. (2015) incorporated gender, age, major of study, and student trace data from the LMS to develop a hierarchical regression model that successfully predicted student outcomes. Additionally, Tempelaar et al. (2017) included undergraduate student gender and mathematics major in combination with trace data and self-reported measures to build a model with strong predictive power of student performance. Researchers concur that when simple student LMS interaction data are combined with student demographic characteristics, stronger, more consistent predictions can be made about student outcomes.

Current Findings in Literature

Correlation

Correlation methods used in an exploratory fashion in learning analytics are intended to elucidate relationships meant for future deeper analysis (Luo, Pan, Choi, & Strobel, 2018). Due to the variation in trace data available in the array of current LMS, there is no agreement on which are most predictive of student performance. The correlation method is a simple way to approach this problem and serves as a foundation for more robust statistical analyses to build upon. Typically, in learning analytics research, when a significant relationship is discovered between student LMS interactions and performance, a regression model would be run later to confirm the findings for predictive claims (Agundo-Peregrina et al., 2014).

Phan, McNeil, and Robin (2016) utilized a correlation in an exploratory fashion to evaluate the research question of what relationship exists between student course performance in a MOOC and patterns of LMS participation. The researchers' findings were significant; students who were reported as being more active, posted to discussions, responded to discussion posts, and submitted an assignment, were more likely to receive a higher score in the course than those who were not as active. Similarly, Luo et al. (2018) examined the relationship between student "chronotypes" (their preferred time to access an asynchronous LMS) and their activity level on subsequent student performance. The researchers discovered students with different preferred access times still accessed the Blackboard Learn LMS with a similar frequency, thusly was not a significant variable. For example, if a student preferred to log on in the morning and another preferred to log on in the evening, both students were similarly active in usage. It was found that the overall level of student LMS activity affected student grades, not chronotype; students who are more active in the LMS achieve higher grades. However, the sample was limited to 88 participants from two undergraduate history courses. A more diverse sample could lead to differing conclusions.

Agundo-Peregrina et al. (2014) was a seminal study in learning analytics that is often referenced by current literature. The authors examined four interaction classifications of students enrolled in online courses in the Moodle LMS and students who were enrolled in face-to-face, online supported courses. The researchers evaluated the bivariate relationship between student performance and student-teacher interaction, student-student interaction, and student-content interaction, and student-system interaction. Each type of interaction was found to have a moderate to strong correlation with student performance in the online courses but not in the face-to-face, online supported courses. Student-teacher and student-content interactions were the most strongly related to student course performance. These findings have implications for learning analytics of student LMS trace data only applying to fully online courses.

Not all correlative research has found significant relationships. While most have identified some positive relationships, Strang (2016a) conducted a study that did not find any significant relationship between undergraduate business student LMS interactions and performance consistent with other literature. In examining the relationship between student reading lessons and final course grade, no significant correlation was found. However, an interesting negative correlation was found between LMS logins to Moodle and course grades. This may indicate a certain pattern of underlying behavior that requires the student to log in more frequently but still perform lower as a result.

Clustering

Aside from statistical analyses, clustering of student behavior is a popular choice in learning analytics research. It allows for researchers to place students into distinct groups based on levels of student LMS interaction to determine if differences exist among the groups' academic performance (Li & Tsai, 2017; Mwalumbwe & Mtebe, 2017; Yeung et al. 2017). Cerezo et al. (2016) clustered students based on temporal data in the LMS, student discussion forum activity, and procrastination. The researchers created four groups: non-task oriented and low procrastination, task-oriented and low procrastination, task-oriented and medium procrastination, and nontask oriented and high procrastination. Expectedly, students who spent more time in LMS tasks (task oriented) and turned in materials on time (low procrastination) performed better than those who did not.

Li and Tsai (2017) clustered students based on online classroom material viewing behavior in a blended computer science students to assess if any significant differences existed between the clusters' performance. The researchers asserted "different engagement levels and behavior patterns may, in turn, affect [student] learning performance" (p. 287). Three groups emerged in their study: consistent use, less use, and slide-intensive use students. Students access a variety of materials in the LMS to assist their completion of tasks and subsequently learn new skills; "Viewing online learning materials is the most frequently performed online learning activities" (Li & Tsai, 2017, p.295). A nonparametric test confirmed that students who more actively viewed course materials consistently also had higher homework scores and final course scores.

Perrotta and Williamson (2018) criticized that the clustering method used in learning analytics can be reductionist. Despite researchers' best efforts, the cluster partitions may not truly reflect the overall structure of the data and true student behavior, thusly producing artificial groupings of students (Perrotta & Williamson, 2018). However, clustering students based on their levels of interactions with the LMS has proved to be an important methodology in learning analytics research that allows not only researchers but educators to more readily identify which groups of students exhibit levels of student LMS interaction that lead to successful performance (Yeung et al., 2017). Due to the popularity of this statistical method, more research is required to ascertain the strengths and weaknesses of its use in education and learning analytics (Perrotta & Williamson, 2018).

Prediction and Modeling

Xing et al. (2015) set forth that learning analytics research usually employs predictive regressions and modeling. It is the central goal of education research to be able to better identify students, through evidence-based prediction, who require intervention and support before performance declines. This goal is evident in the large body of peer-reviewed research examining the predictive capabilities of the learning analytics approach (Choi et al., 2018).

Dvorak and Jia (2016) conducted a regression on student timeliness, regularity, and intensity of eLearning work in two undergraduate courses. They found that every hour a student completed their work before the deadline was associated with a .116 grade point increase. However, while initially timeliness, regularity, and intensity were significant predictors, they became insignificant when the researchers controlled for prior GPA; meaning, prior academic performance was a more significant predictor of current student performance than student work habits. However, when Strang (2017) examined student course logins, assignment activity, and lesson reading activity, a positive predictive relationship with student grades emerged, but student age, gender, and culture did not significantly contribute to the stepwise regression.

You (2015) investigated the significant behavioral indicators of learning in LMS data and their effects on course achievement of 530 undergraduate students enrolled in an online elective course. The researcher appropriately performed a hierarchical linear regression analysis on LMS interaction measures of student studying, total viewing time, sessions, late submissions, reading course information packets, and messages created on a discussion board as predictors of final course score while controlling for where the student was in their program of study. All but total viewing time and messages created on the discussion board were significantly related to student final course score. The regression model with the four significant regressors accounted for 58.1% of the variance within final course scores. After controlling for what year the students were in their program of study, student studying, sessions, late submission, and reading the course information packets significantly predicted student course performance.

Choi et al. (2018) uniquely did not use student LMS interaction data but determined that linear regression better predicted student performance in a freshman undergraduate course than logistic regression through the use of student response clickers in a face-to-face classroom and student summative assessments and demographic information. This is empirically consistent as linear regression better accounts for scores whereas logistic regression typically limits prediction down to pass or fail. This was a unique approach because the majority of learning analytics research utilizes data from the LMS, not physical mechanisms such as student response “clickers.”

Carver et al. (2017) examined if time spent within the graduate online classroom predicted student course score. Logistic regression was used to elucidate if student LMS time activity, defined by total time spent in the course, course modules, document repository, and synchronous online sessions, significantly predicted the students’ earning an A in the course. When the students’ specific program type was controlled for, the logistic regression model significantly predicted students who earned an A in the course. However, time spent in synchronous sessions was the only significant predictor found.

Similar to the aforementioned Agundo-Peregrina et al. (2014) study, Joksimovic et al. (2015) set out to clarify the complex relationship between online interactions and student outcomes as there is no clear distinction on what interactions are most effective. The researchers found that student to system interactions had a positive effect on student performance in their hierarchical linear mixed model regression. Additionally, student to content interactions was negatively correlated with student performance. These findings corroborate the assertion that student LMS interactions in the online education setting have an essential impact on student performance (Joksimovic et al., 2015).

Koc (2017) utilized structural equation modeling to predict student final project and exam scores by discussion forum posts and online lecture attendance in a virtual undergraduate computer programming course. Discussion forum posts were found to have the strongest relationship with both final project and final exam scores. This may be because of the increased opportunity for student-to-student interaction. Both student LMS interactions used explained 35% of the variance in final project score. It was found in the model that while it was a good fit for the data, discussion forum posts had an unexpected direct effect on final course score while online lecture attendance had an indirect effect via final project score. This could suggest that more engaging learning opportunities lead to higher performance for students enrolled online.

Akcapinar et al. (2015) asserted that predictive modeling is a popular statistical approach to utilizing student data in the educational setting. They conducted a study that tested three separate prediction models based on 10 different student LMS interactions reflective of usage to determine which best predicted student course grades. A classification model used accurately predicted student passing 91.8% of the time and student failing 81.5% of the time based on student LMS interactions. Similarly, Stapel, Zheng, and Pinkwart (2016) constructed a model that accurately predicted student pass/fail 73.5% of the time. However, it was based on learning objectives performance measures in a supplementary mathematics eLearning tool rather than student LMS interaction data.

Perrotta and Williamson (2018) raised concerns that while models and algorithms are created to illuminate patterns and coordination of student LMS interactions and performance from complex networks of student data, they are only a "temporary approximation" of a student at any one point in time (p. 7). Thus, they require more longitudinal consideration and research to determine their efficacy in education and learning analytics. Predictive models which are based on student LMS interactions have widespread potential from early-risk identification to their probability of failing a course. Research employing this methodology has important implications for student self-monitoring, interactive activity dashboards, and educator support in adapting materials to best suit student needs (Akcapinar et al., 2015).

Summary and Conclusion

With the growth of K-12 eLearning as an educational choice and learning analytics as an approach to better support eLearning students and instruction, the continuation of research exploring student LMS interactions that are effective predictors of performance is imperative. The main goal of the recent literature discussed in this chapter and the study herein is to provide insight into eLearning students via learning analytics to better inform instruction, identify, monitor, and intervene with at-risk students, and predict student performance. There is an evident gap in eLearning research with a learning analytics approach focusing on K-12 student's performance, specifically mathematics, which this study spanned.

There is a distinct gap between eLearning performers K-12 student performance and their brick-and-mortar peers. Through data-driven, evidence-based learning analytics research, more understanding can be provided into the field as to what student LMS interactions can foster higher performance in K-12 eLearning students. While there still are inconsistencies regarding which student LMS interactions are most predictive of student performance in the online classroom, the learning analytics approach has consistently demonstrated its ability to elucidate the presence (or absence) of relationships between them. The statistical method of regression is predominantly represented in the literature as the most appropriate approach to isolating what student LMS interactions are the most related and predictive of student performance. As a historical and contextual review of the literature was presented in this chapter, the regression statistical method is discussed more in depth in Chapter 3 because it is specifically applied to this study's research question.

Chapter 3: Research Method

Introduction

The purpose of this quantitative study was to fill the present gap in K-12, full-time, virtual education research regarding which LMS interactions are predictive of student performance. I accomplished this through elucidating whether a significant predictive relationship was present between student LMS interactions and student performance after controlling for student demographics previously seen to impact mathematics performance. Student LMS interactions were defined by student clicks through lesson content pages (Predictor Variable; PV_1) and the number of times resources were accessed in the course (Predictor Variable; PV_2) in eLearning Mathematics 7B accessed through a proprietary LMS. Student demographics included as controls (or covariates) were gender (Covariate; CV_1) and FARM status (Covariate; CV_2). Lastly, student final course score in Mathematics 7B (Dependent Variable; DV_1) represented student performance and was consistent with the outcome variable assessed in a majority of learning analytics eLearning education research. In this study, I used archived seventh grade Mathematics B course and student data captured by the proprietary LMS and Google page tracking to determine whether student LMS interactions significantly predict final course score. In this chapter, I will present the threats to external and internal validity and discuss the ethical considerations regarding student data privacy. I intended to provide the results from this study to the virtual educators teaching at the locations examined herein and user interface developers to better inform instruction, intervention, and student at-risk monitoring systems within the LMS.

Research Design and Rationale

With this quantitative study, I hoped to provide insight into the relationship between student LMS interactions and student final course score for students who are enrolled full-time in a public eLearning K-12 program. As enrollment for these programs has increased, the results of this research lends information toward what trace data are most prominent in predicting student success. There was no potential time constraint on data collection as archived data already captured and stored within the LMS and Google page tracking were used.

I extracted the data for student demographics (i.e., gender and FARM status), student clicks on lesson content pages, student frequency in accessing resources, and final student Mathematics 7B course score from archived data stored in the proprietary LMS and Google page tracking. The research design was descriptive in nature because participant data were only collected once and there was no manipulation of any variables. I used a correlational design, which explores the degree with which two or more variables are related (see Creswell, 2013). In this study, I focused on the predictive relationship of the variables being examined. This design choice was most appropriate because with the research questions I sought to determine whether student lesson content page clicks or the number of times resources were accessed can predict student performance in a seventh grade mathematics course, respectively. Correlational designs are often the first exploratory step to determine whether further research is warranted to define the relationship between the two student LMS interactions examined (Creswell, 2013).

Eventually, when applicable, relationships discovered in correlational designs lead to experimental designs to demonstrate causality (Warner, 2013).

The variable framework displayed in Figure 3 reveals the variables (i.e., one dependent variable, two covariates, and two predictor variables) involved in this study. A hierarchical regression statistical analysis was used to test the research questions. I investigated whether student lesson content page clicks (PV_1) or student access to resources (PV_2) has a significant predictive relationship with final course score (DV_1) after controlling for the effects of student gender (CV_1) and FARM status (CV_2).

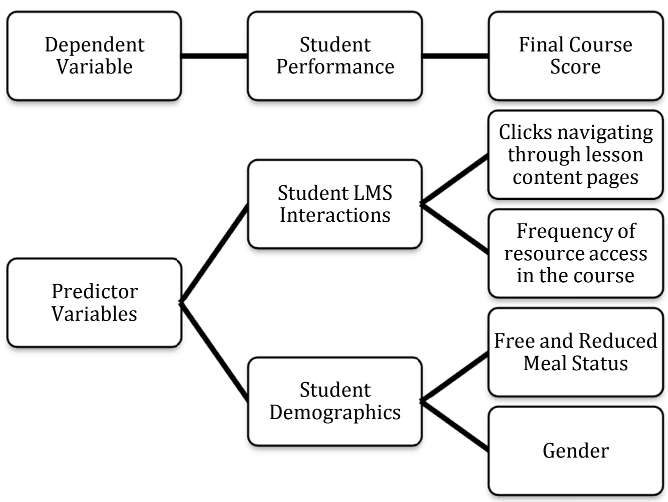


Figure 3. The variable framework of the current study.

Variables

The variables involved in this study are described in Table 1. The table notes the variable type and where it was sourced from. Additionally, a description of what the variables specifically represent is listed in the last column.

Table 1

Archived Variables

Variable	Variable Type	Data Source	Description
PageClicks	Continuous (predictor to be tested in RQ1)	Google Page Tracking	The number of times a student clicks onto a lesson content page
Resources	Continuous (predictor to be tested in RQ2)	Google Page Tracking	The number of times a student clicks on the “backpack” icon to access resources
Gender	Categorical (covariate)	LMS	Male or female
FARM	Categorical (covariate)	LMS	Receives free and reduced lunch or does not receive free and reduced lunch
Math7BScore	Continuous (Criterion variable)	LMS	0–100 final Mathematics 7B course score from 2017–18

Methodology

In this quantitative study, I used a hierarchical regression statistical analysis. The hierarchical statistical method supported the research questions and aligned with the research purpose to determine whether a predictive relationship exists between either of the two student LMS interactions examined and Mathematics 7B course scores while also considering the effect of covariates of student gender and FARM status. Regression is regarded as a prominent statistical tool in research. Warner (2013) cautioned against using regression methodology that groups all variables in a block so that the contributions of each predictor variable on the outcome are indiscernible. The two hierarchical regressions I used were able to isolate the unique impact each predictor variable has on the dependent variable and answer the two research questions posed herein. This method fit within the correlational design choice and was best aligned to test the nature and predictive relationship of the predictors (i.e., student clicks navigating lesson content pages, the frequency of student access to resources, gender, and FARM status) and the dependent variable (i.e., final Mathematics 7B course score). Warner (2013) also stated that hierarchical regression has the potential to reveal more insight into the individual contribution of each regressor to the overall variance in the dependent variable.

Population

The general population for this research study was seventh grade mathematics students enrolled in full-time, public, virtual education programs throughout the United States who used the proprietary LMS and curriculum discussed herein. The target population was those enrolled in the seventh grade mathematics course in two Midwestern states, full-time, K-12, virtual schools who used the proprietary LMS and curriculum. One was considered a large virtual charter, and one was a smaller virtual public school. Both were fully online schools that served students in kindergarten through 12th grade and used the same proprietary LMS and curriculum. The study sample was the entire target population of students enrolled in the Mathematics 7B (a second-semester course) during the spring of 2017–2018 at the two schools mentioned previously. The target population measured approximately 200 students. I obtained the list of these students from the proprietary LMS used by the two virtual schools after all personally identifiable information had been removed to maintain student privacy. At the time of the study, I was employed at the corporate office that supported the full-time, virtual, K-12 programs mentioned earlier and interacted with the LMS data in my daily work.

Sampling Procedure

I did not use a sampling procedure because archived data were available for the entire target population of seventh grade mathematics students enrolled at two Midwestern, full-time, virtual schools that used the proprietary LMS and curriculum. This was based upon data availability and data collected as part of my role at the time of the study as a senior research analyst at the company that owns and operates the proprietary curriculum and LMS. I retained a homogenous group of students from the target population to avoid any outliers or undue influence on the analyses. The inclusion and exclusion criteria for this study were:

Seventh grade students only, those enrolled in Mathematics 7B,

- On-time enrollees (i.e., students who began Mathematics 7B on/or by January 8th–10th, 2018), and
- Students who completed the course obtaining a final grade.

Power Analysis

I downloaded G*Power statistical software, Version 3.1, which is available for free download through the Heinrich-Heine University of Dusseldorf, for my personal computer to determine the necessary sample size required for the study. The *F* Test, linear multiple regression: fixed model, R^2 increase was used to define the sample size for this study. According to the analysis, with a single test predictor and two control predictors (i.e., covariates) for each of the two research questions, a minimum of 55 students would be needed to achieve the power of .80 with the medium effect size (see Appendix A). The approximate 200 students enrolled in Mathematics 7B course during the second half of the 2017–2018 school year exceeded the minimum required participant threshold as designated by the power analysis.

Procedures

Recruitment. I did not use recruitment procedures in this study. The schools and students participating in the study had their data collected and maintained as part of their everyday agreement and relationship with the corporate office managing and operating their proprietary LMS and curriculum. The trace and performance data on students who were enrolled in Mathematics 7B during the 2017–2018 school year for this study were archived within the LMS. The course was approximately 20 weeks long and occurs during the second semester of the school year, starting January 8th–10th and concluding May 22nd–25th. Additionally, demographic data were obtained from students upon their enrollment and stored within the LMS. Google page tracking data were continuously stored while the student was enrolled in the course.

Informed consent. I sent a letter of data use agreement and cooperation to the director of research, assessment, and accountability of the corporation that supported the data analysis for the public virtual programs being used in this study. The letter of cooperation outlined the purpose of the study, the learning analytics approach, sharing of findings, and the confidentiality of student private information. A letter of data use was not necessary from the specific school locations because their data were part of my and the organization’s everyday work in supporting the operations of the schools through the management of the proprietary LMS and curriculum. I obtained the signatures on the letter of data use agreement and cooperation prior to institutional review board (IRB) approval in accordance with the Family Educational Rights and Privacy Act of 1974 (U.S. Department of Education, 2015). The letter permitted me to access the proprietary LMS and Google page tracking analytics to extract the data for this research study through my current role and permissions as a senior research analyst.

Data collection. I sourced the data for this study from the proprietary LMS provided by my role and permissions at the time of the study within the Department of Research, Assessment, and Accountability that supported the schools who use the proprietary LMS and curriculum being examined herein. The statistical package that I used to securely store and analyze the data provided was International Business Machine’s (IBM) Statistical Package for the Social Sciences (SPSS; IBM, 2016). This software was present on my password-protected laptop.

Instrumentation. Archival data containing all variables were obtained from Google page tracking and the proprietary LMS. The LMS data will include student course performance, student gender, and student FARM status. Google page tracking data included student clicks on lesson content pages in Mathematics 7B, and the number of times resources were accessed in the Mathematics 7B course.

Data Analysis. The data for the target population were obtained through Microsoft Excel worksheet exports from Google page tracking as well as the proprietary LMS. The data were stripped of any identifying information to ensure student anonymity. No further data cleaning was required. The data were imported into SPSS for coding and analysis. The only variables that needed transforming were the dichotomous gender and FARM variables. These variables were coded for statistical analyses. Gender was dichotomously coded as follows: 0 indicates a male student, 1 indicates a female student. Similarly, FARM was coded as follows: 0 indicates a non-FARM student, 1 indicates a FARM student. This coding method allowed for the dichotomous variables to be quantitatively assessed. Student click data and course scores were continuous in nature and needed no further modifications for analysis.

To answer the research questions in this study, two hierarchical regressions were run to statistically analyze the data in SPSS; one hierarchical regression for each research question. Each research question was run independently of one another to isolate the effect of the test predictor on the outcome while controlling for the covariates involved. The research questions and hypotheses are restated here. The steps that were taken for data analysis of each research question are described further below.

Research Question 1: Can student clicks navigating the LMS course content page(s) predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

$H_0$1: Student clicks navigating the LMS course content page(s) cannot significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

$H_1$1: Student clicks navigating the LMS course content page(s) can significantly predict student performance in a full-time, virtual, Grade 7

Mathematics course after student demographic variables are controlled for.

Research Question 2: Can the number of times resources were accessed within the course modules predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

$H_0 2$: The number of times resources were accessed within the course modules cannot significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

$H_1 2$: The number of times resources were accessed within the course modules can significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

The covariates involved in the research questions and hypotheses were student gender and FARM status. These covariates were chosen for inclusion in the regression models because current research in the field of learning analytics, specifically focused on mathematics performance, have included these student demographics as they have previously demonstrated impact on performance (Khalil & Ebner, 2011; Marchetti et al., 2016; Miller et al., 2015; Tempelaar et al., 2017). To better understand the effects of the test predictor in each research question, it was important to control for possible covariates already known to impact student performance.

RQ1 was entered into SPSS as a two-block hierarchical regression. The first block had the dependent variable, student course score in Mathematics 7B, as the outcome, and the two covariates, FARM and student gender as predictors. The second block then included the test predictor of student clicks on lesson content pages for Mathematics 7B. The final regression model (in Block 2) had three total predictors, one test predictor, and two covariates. This block-wise model allowed for the effect of the test predictor, student clicks on lesson content pages, to be ascertained in the second block, while already controlling for the effects of the covariates on student course score in the first block of the regression model. The significance of R-square increase from Block 1 to Block 2 will address the first research question.

Similarly, RQ2 was run through SPSS as a hierarchical regression with two blocks. The first block had the dependent variable, student course score in Mathematics 7B, as the outcome, and the two covariates, FARM and student gender, as predictors. The second block then included the test predictor of how many times the student accessed resources in the Mathematics 7B course. The final model (Block 2) had three total predictors, one test predictor, and two covariates. Mirroring the first research question, this block-wise method enabled me to isolate the effect of the test predictor, how many times a student accessed resources, on student course scores while controlling for the effect of the covariates in the first block of the regression model. The research question was addressed via the significance of the R-square increase from Block 1 to Block 2 of the regression model.

For each of the research questions, the Type I error was set as .05. If the *p* value for the R-square change from Block 1 and Block 2 is above the alpha level of significance, then I would fail to reject the null hypothesis. Following the explanation of data analysis, threats to the validity of the study must be considered. The possible threats to internal, external, and construct validity within the research will be discussed further below.

Threats to Validity

I will discuss the external, internal, and construct validity of this research study in this section. The integrity of research studies is contingent upon the tools used to measure the data and the nature of the data itself. Statistical analysis is important to test the data and assist researchers in determining the validity of the study; however, statistical procedures are but a minor part of the research. If the data, design, or measurement in the study are unsound, the results will be unusable (Frankfort-Nachmias & Leon-Guerrero, 2014). Thusly, it is imperative to discuss external, internal, and construct statistical validity as it specifically relates to this research study.

External

Warner (2013) defined external validity as the ability for research results to be generalized to participants, settings, and resources in the general population outside of what is included in the study's sample. Additionally, Warner stated that a study that is "closely analogous to or resembles, the real-world situations that the researcher wants to learn about" will achieve external validity (p .70). Nonexperimental studies normally have sufficient external validity because they take place in the natural, uncontrived environment, measuring real-world events or behaviors (Warner, 2013). The study herein involves data measured objectively by and within the eLearning classroom during the natural course of the Math 7B class. Students are often unacquainted with the data being captured by the LMS regarding their clicks on the screen. Because students do not focus on how the LMS tracks their click behavior, this invites the notion that the data captured regarding student clicks is as close to a naturally occurring measurement as possible.

It is important to consider that generalizations cannot be made across constructs. External validity would be decreased if more general conclusions were to be drawn from the single construct measured. For example, the researcher cannot draw conclusions regarding the relationship between the frequency of all student click behavior in the virtual classroom and student performance, if only specific lesson content page clicks are being measured and tested. Similarly, we cannot draw conclusions about the predictive relationship of student lesson content page clicks in all subject courses if only mathematics is being explored. Student click behavior may vary depending on the content and subject being presented in the virtual classroom. Lastly, there are inherent differences in LMS. It is important to consider this when interpreting findings from this research study.

Internal

Internal validity refers to how well the results of a study can be used to suggest causal relationships related to the participants, settings, and materials involved; this is sometimes accomplished at the expense of external validity (Warner, 2013). This study was a nonexperimental correlational design and did not make causal inferences regarding student LMS interactions and final mathematics course score. Nonexperimental research studies, typically have weak internal validity due to the inability to rule out any confounding variables (Warner, 2013).

There are a few possible aspects of this study that can impact internal validity beyond those inherent in the research design. For example, information a student's history of online learning or exposure to learning via technology connected to the Internet has not been collected and cannot be accounted for. Additionally, extraneous variables within the students' environments that impact their use of the computer and ability to focus on their learning cannot be controlled for either. It is not possible to rule out all confounding variables or unaccounted for correlative effects. The current study attempts to control for variables cited in recent peer-reviewed literature known to demonstrate impact on student mathematics performance (gender and FARM status). However, not all factors can be accounted for in the students learning environment (i.e., where they access the computer to log into the virtual classroom and the number of possible distractions.). It is possible that a student accessed their course from an Internet café, library, or other public environment that may impact the student's ability to focus on the course, this way impacting their interaction with the LMS. This can cause the researcher to potentially falsely accept or reject the null based on extraneous circumstances impacting the data.

Student history with online learning can influence their use of the LMS and can subsequently be reflected in the data analyzed for the current research. If a student is a frequent and savvy online user, they may be more comfortable navigating the LMS which would change their mouse click behavior. Alternatively, the data would look different for a student who is a first time LMS user. Again, these circumstances can cause the null to be falsely accepted or rejected. Lastly, it is important to consider my role as researcher, who collected the data for this research study. This role carries the possibility of exhibiting bias when analyzing results especially given my connection to the data at the time of this writing (Robson & McCartan, 2016).

Construct Validity

The intent of construct validity in nonexperimental research studies is to ensure the variables measured reflect the larger constructs or phenomena at the focus of the study. The potential issue with this is that there is no certain way to guarantee that the variables chosen are actually measuring the construct for the study (Warner, 2013). It is imperative to the construct validity of a study to demonstrate that the variables being measured are aligned with the operationalization of the constructs and theoretical frameworks founded in the literature, for the results to support predictive findings and accept, or reject, the hypotheses. This ensures that the findings of a study provide meaningful contributions to the field regarding the participants, settings, and materials being examined (Creswell, 2013). The student LMS interactions are pulled out of Google page tracking and are assumed to reflect student click behavior which may impact performance. However, a potential weakness in construct validity exists due to the possibility of mouse clicks not being truly reflective of student learning behavior. This means that even though a student clicked into a lesson page, or the classroom resources, they may not have interacted with the items or read the content. The field of learning analytics acknowledges this as a commonly occurring threat to validity in research studies examining virtual student behavior (Li & Tsai, 2017; Stapel et al., 2016).

Ethical Procedures

Ethical considerations are shaped by the nature of the research design. This research study was a nonexperimental quantitative correlational study that used archived data from two separate and qualitatively different full-time, virtual, K-12 public schools. While the course being examined is the same, the possible differences in school climate and teaching should be considered. A distinct effort was made to have a representative sample of the overall seventh grade full-time eLearning population; it is possible there is an ethical concern of true representation. Ultimately, the chief ethical concern in this study was regarding my role of researcher conflicting with data being sourced from my place of occupation.

Ethical considerations regarding student data and privacy must be addressed. Outside of protecting student data on a password-protected laptop, participants remained anonymous and confidential; no personally identifiable information was required to complete the study. Results from this research study will not be shared with any parties outside of those agreed upon in the data agreement and letter of cooperation. These documents are stored on my laptop, separate from the data, for a minimum of 5 years following the completion of this study to maintain the integrity of the agreement. Finally, the materials, participants, and settings in the research study were aligned with the IRB stipulations at Walden University (IRB Approval number for this study can be found in Appendix C).

Summary

In this nonexperimental quantitative research study, I sought to determine if a predictive relationship existed between student LMS interactions and student course score for full-time virtual students in Mathematics 7B. Appropriate data sharing agreements and letter of cooperation were obtained from the director of the department that supports virtual schools who use the proprietary LMS and curriculum involved in this study. Through my role at the company at the time of this writing, access and permission to the archived data were already granted through everyday work. The archived data were accessed from the proprietary LMS, through my current permissions and role as senior research analyst, for two mid-western, virtual, K-12 schools. The archived data were de-identified of any private, personally identifiable, student information before it was stored and subsequently analyzed in SPSS. A hierarchical linear regression was used to statistically test if there is a significant predictive relationship between student lesson page views, their access to the virtual backpack for course resources, student gender and FARM status, and their final course score in Mathematics 7B. The results the regressions used to test the research questions will be discussed in Chapter 4.

Chapter 4: Results

Introduction

The purpose of this quantitative study was to address the current research gap in learning analytics research concerning the full-time, K-12, eLearning population. Ultimately, I sought to determine whether two types of student LMS interactions could predict mathematics course performance: student clicks on course content pages and the number of times resources were accessed. In this study, I analyzed archived student data regarding Mathematics 7B course scores (DV_1), lesson content page clicks (PV_1), resources accessed (PV_2), and the covariates of student gender (CV_1) and FARM status (CV_2).

I developed the following research questions to guide this study:

Research Question 1: Can student clicks navigating the LMS course content page(s) predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

Research Question 2: Can the number of times resources were accessed within the course modules predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

I tested a null hypothesis for each research question:

$H_0$1: Student clicks navigating the LMS course content page(s) cannot significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

$H_0 2$: The number of times resources were accessed within the course modules cannot significantly predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for.

In this research study, I drew from Spring 2018 (second semester) archived data from a Mathematics 7B course taken at two, full-time, K-12, virtual public schools located in Midwestern states. Both schools used the same organization's curriculum and proprietary LMS to deliver the same Mathematics 7B course to their students. The statistical analysis was conducted using two, separate, hierarchical multiple regressions with each research question being independently analyzed in its own model. I entered the first research question into the hierarchical regression (i.e., Model 1) such that the student final course score was the outcome (DV_1), the covariates of student gender and FARM status (CV_1 and CV_2) entered in Block 1, and the predictor student lesson page clicks (PV_1) was entered into Block 2. The second research question was entered similarly (i.e., Model 2), except the predictor of the number of times a student accessed resources (PV_2) was entered into Block 2. Hierarchical regression, a form or multiple regression, is considered a mainstay in social science research (Laureate, 2016a). Hierarchical regression, as Warner (2013) noted, is most appropriate to isolate the ability of a predictor to predict an outcome while controlling for other covariate effects. In the rest of this chapter, I will go on to discuss the data collection, describe the nature of the data, and evaluate the statistical findings of the aforementioned research questions and quantitative analysis.

Data Collection

My collection of data did not deviate from the initial plan outlined in Chapter 3. Data collection consisted of extracting archived student data from the proprietary LMS and Google page tracking for students who were enrolled from January 8th–10th to May 22nd–25th, 2018. The data extracted for the hierarchical regression analyses were:

- Seventh grade students enrolled from January 8th–10th and did not withdraw until May 22nd–25th in Mathematics 7B,
- Student gender,
- Student FARM status,
- Student access to course resources,
- Student clicks on course content pages, and
- Student final course score in Mathematics 7B.

I used the entire target population of seventh grade mathematics students enrolled in two, full-time, K-12, virtual public schools located Midwestern states. One school was a large virtual charter school, and the other was a smaller virtual public school. After examining the sampling frame and exclusionary criteria, there were a total of 238 seventh grade students enrolled in Mathematics 7B during the time frame of the second semester and who successfully received a final grade. There were 101 male students and 137 females. Additionally, there were 125 non-FARM students and 113 FARM students. This reasonably equal distribution of the covariates contributed to the representativeness of the participants being reflective of the larger population of full-time, eLearning, seventh grade mathematics students.

Results

With the null hypotheses, I tested whether student clicks on lesson content pages predicted their final course score in seventh grade Mathematics after controlling for student gender and FARM status. I also tested whether the number of times a student accessed resources predicted their final course score in seventh grade Mathematics after controlling for student gender and FARM status. In this study, I used hierarchical regressions to address the research questions.

Assumptions

The purpose of the hierarchical regression conducted in this research study was to determine whether the test predictor variables (i.e., student clicks on content pages and the number of times resources were accessed) could predict the dependent variable (i.e., student final course score in Mathematics 7B) while controlling for the covariates of gender and FARM status. There were a total of five assumptions that needed to be met in order to substantiate the findings of the hierarchical regression and increase the possible generalizability of the statistical outcomes: normality, linearity, multicollinearity, homoscedasticity, and undue influence. When assumptions of the regression model are violated, the reliability of the statistical analyses is suppressed (Warner, 2013). Regressions are robust against violations of its assumptions of heteroscedasticity and normality (Gelman & Hill, 2006; Osborne & Waters, 2002). However, it is vital to test for all assumptions to ensure results can be interpreted with integrity. The results of the assumption tests will be discussed in the following subsections as well as any implications they have upon data analysis and subsequent interpretation of findings (see Osborne & Waters, 2002; Warner, 2013).

Assumption 1: Normal distribution of residuals. Field (2017) stated that a P-P plot, or probability to probability plot, is most useful for evaluating this assumption in larger sample sizes. However, it is important to note that larger sample sizes adhere to the central limit theorem that states that the assumption of normality is inherently met despite what the sample data displays within probability to probability plots of the residuals for a given regression model (Field, 2017). Although the sample size in this study was considered large, I followed through with the test of normality. The plot of expected cumulative probability (i.e., normal) to the observed cumulative probability values for RQ1, as displayed in Figure 4, shows a slight deviation from the linear line; however, most points fall on or near the line with a clear trend. This means that residuals of the regression model for student course scores (DV_1) regressed onto lesson content page clicks (PV_1), student gender (CV_1), and FARM status (CV_2) adhere to the assumption of normality.

Secondly, the plotted residual values for RQ2 are displayed in Figure 5. A similar plot line appears with values slightly deviating from the line but generally displaying a straight trend. This means that the residuals of the regression model for student course scores (DV_1) regressed onto the number of times students accessed resources (PV_1), student gender (CV_1), and FARM status (CV_2) adhere to the assumption of normality. These test results, combined with the central limit theorem, confirmed that the assumption of normality was met for both RQ1 and RQ2.

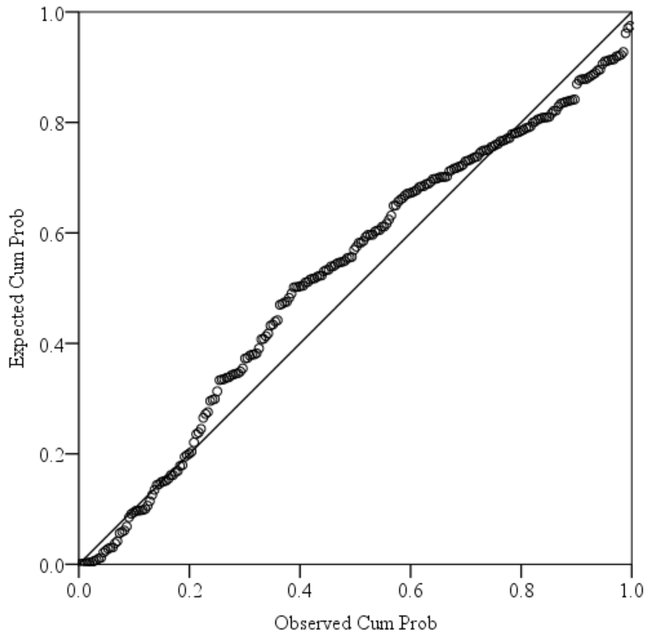


Figure 4. RQ1 P-P plot of Math 7B score.

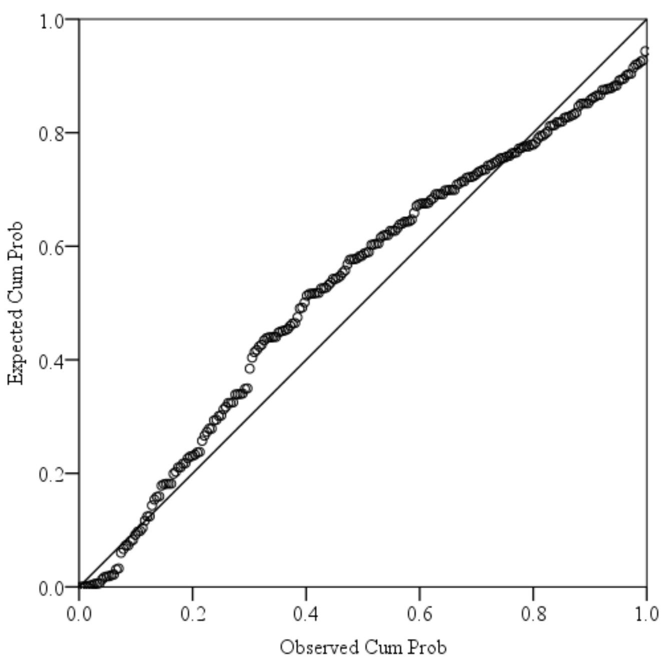


Figure 5. RQ2 P-P plot of Math 7B score.

Assumption 2: Linear relationship. The linearity assumption is the most pertinent to the interpretation of the regression model (Gelman & Hill, 2006). The relationships modeled in the hierarchical regressions should be linearly related with a straight-line relationship between the dependent variable and the predictor variables and covariates for regression results to be interpreted (Osborne & Waters, 2002). Once again, I interpreted P-P plots to determine whether the regression models met the linear relationship. Figure 4 displays the P-P plot for RQ1. This chart displays a majority of data points fall on or along the trend line indicating a linear relationship between student course scores (DV_1) and lesson content page clicks (PV_1), student gender (CV_1), and FARM status (CV_2). There was no clear horizontal trend or S-shape present, which would indicate a violation of linearity for the model.

The P-P plot for RQ2 demonstrates the same general linear trend that Figure 4 displayed. As seen in Figure 5, a majority of data points fall on or along the trend line indicating a linear relationship between student course scores (DV_1) and the number of times a student accessed resources (PV_2), student gender (CV_1), and FARM status (CV_2). While both P-P plots show a slight deviation from the trend line, it is not marked enough (i.e., horizontal or S-shaped) to violate the assumption of normality (see Osborne & Waters, 2002). Therefore, the results of the P-P plot evaluation of linearity indicate that both Model 1 and Model 2 met the assumption of linearity.

Assumption 3: Multicollinearity. The test of multicollinearity examines if any of the predictor or control variables are highly correlated with one another (Garson, 2012). There are two ways to verify the model does not display any multicollinearity: Pearson correlation matrix and variance inflation factor (VIF) values (Berry & Feldman, 1985; Laureate, 2016b). A Pearson bivariate correlation matrix is the most common way to assess if multicollinearity is present (Garson, 2012). A Pearson $r > 0.8$ or -0.8 between any of the predictor or control variables would indicate a strong correlation. As seen in Tables 2 and 3, none of the predictor or control variables for RQ1 or RQ2, respectively, display a strong relationship with one another.

The VIFs measure multicollinearity as well (Laureate, 2016b). The suggested VIF value is < 10.0. Generally, predictor variables within the model that have a significant relationship with one another will demonstrate a VIF ≥ 10 (Laureate, 2016b). The VIF measures for Model 1 were all < 1.1. The VIF measures for Model 2 were also < 1.1. These results are displayed with the regression coefficients of Tables 7 and 9 in the following section. With the results of both the VIF measures and Pearson correlations, the assumption of multicollinearity was met for both models.

Table 2

RQ1 Pearson Correlations

	Dependent variable	Covariate variables		Test predictor
	Math7Score	FARM	Gender	PageClicks
Math7BScore	1.000	-.283***	-.107*	.296***
FARM	-	1.000	-.035	-.063
Gender	-	-	1.000	-.057
PageClicks	-	-	-	1.000

*** *a* < .001 and * *a* < .05

Table 3

RQ2 Pearson Correlations

		Control variables		Predictor
	Math7Score	FARM	Gender	Resources

Math7BScore	1.000	-.283***	-.107*	.156***
FARM	-	1.000	-.035	-.057
Gender	-	-	1.000	.102
Resources	-	-	-	1.000

*** *a* < .001 and * *a* < .05

Assumption 4: Homoscedasticity of error variance. When data are homoscedastic, it has a normal distribution or errors among different levels of the independent variables (Berry & Feldman, 1985; Osborne & Waters, 2002). When there is slight heteroscedasticity, the impact on the significance tests of the regression model is minuscule; However, if the distribution of errors displays a marked “cone-shape” pattern, then findings are considered distorted and the chance for Type I errors are largely increased (Laureate, 2016b; Osborne & Waters, 2002). To examine the data for a normal distribution of errors, the scatterplot of the predictors and covariates in each research question must be interpreted (Laureate, 2016b).

RQ1 can be seen in Figure 6 as having a slight concentration of residuals for lesson content page clicks (PV_1), student gender (CV_1), and FARM status (CV_2) central and above the line, but no marked cone-shape trend which would indicate a serious violation of homoscedasticity. The slight heteroscedasticity will have little effect on the regression model’s significance tests (Osborne & Waters, 2002). The scatterplot of residuals distribution for RQ2 is displayed in Figure 7 and shows a homoscedastic distribution of errors for the number of times student accessed resources (PV_2), student gender (CV_1), and FARM status (CV_2). The absence of a marked cone-shape concentrated distribution in the scatterplot of residuals for RQ1 and 2 confirms this assumption was met for both models.

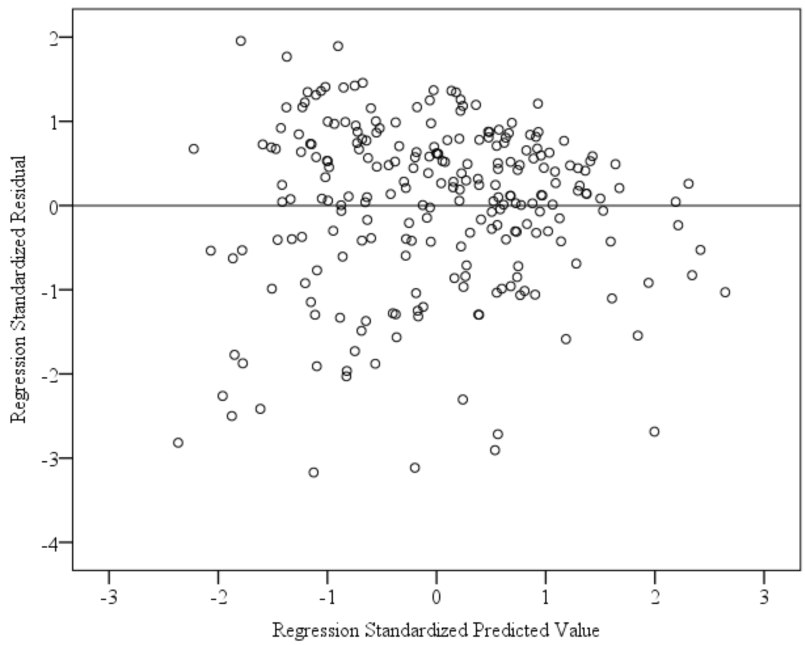


Figure 6. RQ 1 scatterplot of lesson page clicks.

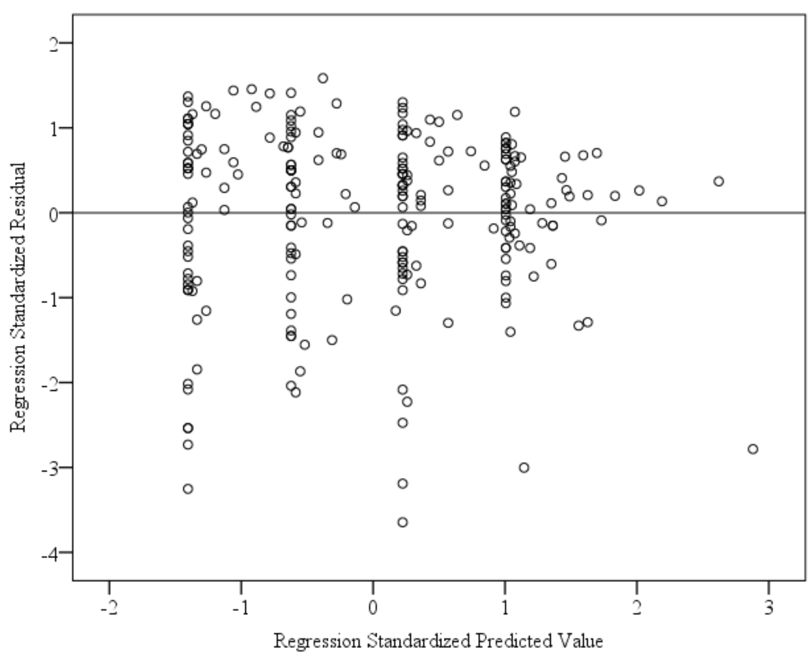


Figure 7. RQ 2 scatterplot of the number of times resources were accessed.

Assumption 5: Undue influence. The assumption of undue influence examines if there are outliers significantly impacting the regression model (Laureate, 2016b). To further investigate, Cook's distance (*D)* is used to determine if any outliers are causing undue influence and should generally be removed (Osborne & Overbay, 2004). By accepted standard, a Cook's maximum distance statistic of < 1.0 indicates there is no undue influence on the model (Laureate, 2016b). Model 1 displayed a maximum Cook's $D = .091$, well below the accepted threshold of 1.0. Likewise, Model 2 returned a maximum Cook's $D = .335$. These results indicate that the assumption of undue influence has been met for both regression models.

Descriptive statistics. Descriptive statistics are important pieces of information to contextualize the data and subsequent statistical results (Ary, Jacobs, Irvine, & Walker, 2018). Table 4 summarizes the descriptive statistics of the dependent variable and the predictors: student clicks on lesson content pages, and the number of times the student accessed resources in the virtual classroom. The mean course score of the entire target population used herein (67.8%), number of page clicks on course content pages (1004.8), and the number of times resources were accessed (6.3) are displayed along with the standard deviations. Additionally, the frequency distribution of the covariates, student gender and FARM status, are shown in Table 5. This information displays an equitably distributed target population with slightly more female than male students, and relatively equal FARM and non-FARM student groups.

Table 4

Descriptive Statistics for RQ1 and RQ2

	M	*SD*
Math7BScore	67.8	.163
PageClicks	1004.8	332.6
Resources	6.3	13.019

Table 5

Frequencies for RQ1 and RQ2

	N	% of Target Population
Female	137	57.6
Male	101	42.4
FARM	113	47.5
Non-FARM	125	52.5

Research Question 1

A hierarchical regression was conducted to test if the predictor variable of student clicks on lesson content pages predicted final course score in Math 7B while controlling for the covariates of FARM status and gender. A significance level of .05 was used in the hierarchical multiple regression analysis. Two blocks were used in the regression model. Block 1 included the covariates student FARM status and gender. The amount of times students clicked on lesson content pages was then added to Block 2 as the predictor variable. If the p value of the predictor variable added to Block 2 was less than the set threshold of .05, then the relationship was deemed significant. Results of the hierarchical regression for RQ1 are shown in Tables 6 and 7.

The hierarchical regression model for RQ1 revealed that introducing the predictor variable of student clicks on lesson content pages to the regression model in Block 2 produced an R^2 change that was significant, $\Delta R^2 = .074$, $F(1, 234) = 20.843$, $p < .001$. Additionally, Block 2 was statistically significant, $F(3, 234) = 15.729$, $p < .001$. The addition of student clicks on lesson pages to the model increased the R^2 7.4%. Block 2 accounted for a total of 16.8% variance in student final course score when FARM status and gender were controlled for. Student clicks on lesson pages was a significant predictor, $B < .001$, $p < .001$ (as shown in Table 7). For every single click on a lesson content page, a student's final course score will increase by <.001 points on average. This is because student lesson content page clicks ($M = 1004.8$, $SD = 332.6$) vary widely in the target sample and students would have to make a considerable amount of clicks to see a difference in course scores. Based on the change in the R^2, the null hypothesis was rejected. This means that student clicks on lesson pages predicted final course score when FARM status and gender were controlled for.

Table 6

RQ1 Hierarchical Regression Model Summary and ANOVA Results

	Model Summary						ANOVA		
Block	R^2	ΔR^2	ΔF	*df* (1,2)	*p*	Durbin-Watson	*F*	*df* (1,2)	*p*
1	.094	.094	12.146	2,235	.000[a]		12.146	2,235	.000[a]
2	.168	.074	20.843	1,234	.000[b]	1.918	15.729	3,234	.000[b]

a. Predictors: (Constant), FARM, Gender
b. Predictors: (Constant), PageClicks, Covariates: FARM, Gender
c. Dependent Variable: Math7BScore

Table 7

RQ1 Regression Coefficients for Individual Predictor Variables

Block		*B*	*β*	*p*	*VIF*
1	FARM	-.093	-.287	.000***	1.001
	Gender	-.038	-.117	.061	1.001
2	FARM	-.087	-.269	.000	1.005
	Gender	-.033	-.101	.093	1.005
	PageClicks	<.001	.273	.000***	1.007

*** *a <.001*

Research Question 2

A hierarchical regression was conducted to test if the predictor variable of student access to resources predicted final course score in Math 7B while controlling for the covariates of FARM status and gender. A significance level of .05 was used in the hierarchical multiple regression analysis. Two blocks were used in the regression model. Block 1 included the covariates student FARM status and gender. The amount of times a student accessed resources was then added to Block 2 as the predictor variable. If the *p* value of the predictor variable added to Block 2 was less than the set threshold of .05, then the relationship was deemed significant. Results of the hierarchical regression for RQ 2 are shown in Tables 8 and 9.

The hierarchical regression model for RQ2 revealed that introducing the predictor variable of student access to resources to the regression model in Block 2 produced a significant R^2 change, $\Delta R^2 = .023$, $F(1, 234) = 6.168$, $p < .05$. Block 2 was also statistically significant, $F(3, 234) = 10.331$, $p < .001$. The addition of times resources were accessed increased the R^2 2.3%. Block 2 accounted for a total of 11.7% variance in student final course score when student FARM status and gender were controlled for. The number of times students accessed resources was a significant predictor (B = .002, $p <$ .05). For every single access to resources, a student's final course score will increase by .002 points, on average. This is because student access to resources (M = 6.3, SD = 13.019) vary widely in the target sample and students would have to access resources many more times to see a difference in course scores. Based on the change in the R^2, the null hypothesis was rejected. This directs that the number of times students accessed resources predicted final course score when FARM status and gender were controlled for.

Table 8

RQ2 Hierarchical Regression Model Summary and ANOVA Results

	Model Summary						ANOVA		
Block	R^2	ΔR^2	ΔF	*df* (1,2)	*p*	Durbin-Watson	*F*	*df* (1,2)	*p*
1	.094	.094	12.146	2,235	.000[a]		12.146	2,235	.000[a]
2	.117	.023	6.168	1,234	.014[b]	2.003	10.331	3,234	.000[b]

a. Predictors: (Constant), FARM, Gender
b. Predictors: (Constant), Resources, Covariates: FARM, Gender
c. Dependent Variable: Math7BScore

Table 9

RQ2 Regression Coefficients for Individual Predictor Variables

Block		*B*	*β*	*p*	*VIF*
1	FARM	-.093	-.287	.000***	1.001
	Gender	-.038	-.117	.061	1.001
2	FARM	-.090	-.279	.000***	1.004
	Gender	-.043	-.133	.033*	1.011
	Resources	.002	.154	.014*	1.013

**** a <.001, * a < .05*

Summary

The hierarchical regression statistical analyses examined archival data from two Midwestern full-time K-12 virtual schools. A total of 238 seventh grade mathematics students (the entire target population) had their performance and demographic data extracted from the proprietary LMS. Their access to resources and lesson page clicks were extracted from Google Page Tracker. All data were combined and analyzed using SPSS analytical software. The hierarchical regression test for RQ1 indicated student lesson page clicks predicted Math7B score after controlling for student demographics (FARM status and gender), $\Delta R^2 = .074$, $F(1, 234) = 20.843$, $p < .001$. The hierarchical regression test for RQ2 also revealed the number of times students accessed resources predicted Math7B score after controlling for student demographics, $\Delta R^2 = .023$, $F(1, 234) = 6.168$, $p < .05$. These results indicate, at the alpha level of .05, the null hypotheses for RQ1 and RQ2 were rejected. The limitations to these findings and implications for discussion are addressed in Chapter 5.

Chapter 5

Introduction

The purpose of this research study was to determine whether a predictive relationship existed between student LMS interactions and their performance in a mathematics course after demographic effects were controlled for. In this research study, I used a learning analytics perspective to examine whether student mouse clicks on lesson content pages or the number of times students accessed resources in the virtual classroom were able to predict student course scores after student gender and FARM status were controlled for. The sample of 238 students was comprised of the entire target population. The first null hypothesis was that student mouse clicks on lesson content pages (PV_1) would not significantly predict student course performance in Math 7B (DV_1) after the effects of student gender (CV_1) and FARM status (CV_2) were controlled for. The second null hypothesis was that the number of times a student accessed resources in the virtual classroom (PV_2) would not significantly predict student course performance in Math 7B after the effects of student gender and FARM status were controlled for. I used a hierarchical regression to test each of these hypotheses individually.

Interpretation of Findings

The findings of this research study fall in line with the results of similar studies in the field of eLearning that demonstrate student LMS interactions can predict student performance (Agundo-Peregrina et al., 2014; Goggins & Xing, 2016). However, the results of this study extended knowledge by examining the underresearched population of K-12, full-time, virtual students. The theoretical framework used for this study was comprised of the online learning as online participation theory (Hrastinski, 2009) and the three types of interactions model (Moore, 1989). The theoretical framework supported the overarching assertion discussed in the literature review of Chapter 2 that virtual student participation impacts performance. Lara et al. (2014), Rientes and Toetenel (2016), and Strang (2016a) cautioned against reducing student participation down to trace data such as mouse clicks. However, most studies concurred with the findings from this study that generally, the more "active" a student was in the virtual math course, the better their performance (Curtis & Werth, 2015; Pazzaglia et al., 2016). In this study, I addressed two research questions by using hierarchical regression statistical analysis:

Research Question 1: Can student clicks navigating the LMS course content page(s) predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

Research Question 2: Can the number of times resources were accessed within the course modules predict student performance in a full-time, virtual, Grade 7 Mathematics course after student demographic variables are controlled for?

With the results from the statistical analyses for RQ1, I concluded that student mouse clicks on course content pages predicted final course score in Math 7B after the effects of gender and FARM status were controlled for were $\Delta R^2 = .074$, $F(1,234) = 20.843$, $p < .001$. This allowed me to confidently reject the null hypothesis and accept the alternative hypothesis. Similarly, with the results from the hierarchical regression for RQ2, I concluded that the number of times a student accessed resources predicted final course score in Math 7B after the effects of gender and FARM status were controlled for were $\Delta R^2 = .023$, $F(1, 234) = 6.168$, $p < .05$. I rejected the null hypothesis and subsequently accepted the alternative hypothesis.

Limitations of the Study

There were a few limitations to this study, which may prevent a generalization of the findings. While a strength of the study was that the entire target population was used, the findings were limited to students enrolled full-time in a virtual K-12 school and not the larger population of online students enrolled part-time or in blended programs. The course under study was taken in two, separate, Midwestern state schools, and the natural variance in school climate or teachers could have impacted the data and findings. The data from this study were contingent upon the student working within the LMS. The results from this study could not consider those students who worked offline or outside of the LMS with supplemental/printed materials. Additionally, the data for this study were obtained from a proprietary LMS that has unique features preventing it from being compared to other LMSs more widely used throughout the eLearning environment (i.e., Blackboard, Moodle, etc.). Lastly, I accounted for a small percentage of factors that could explain or predict student final course scores in this study. While the predictors and covariates included predicted student course scores, there are a large selection of student factors that can also impact performance that were not considered herein. Additional covariates and participation data should be included in future research to strengthen the results found in this study.

Recommendations

Students are enrolling in virtual K-12 schools more and more every year (Curtis & Werth, 2015; Lowes et al., 2015). Moreover, mathematics performance is lacking nationwide but especially in virtual programs (Choi et al., 2017). The results of this study, the theoretical framework, and recent similar studies in the field acknowledge that student participation in the virtual environment impacts their performance (Curtis & Werth, 2015; Goggins & Xing, 2016; Pazzaglia et al., 2016). However, more research is required to confidently define what other student participation factors in the LMS predict student performance.

Based on the findings from this study, I offer the following recommendations:

- Include student demographics in future research for more insight on how teachers should prioritize their support of students.
- Include student LMS interaction data into a dashboard indicator for teachers to better support students with low activity in the online course.

These recommendations are a guide for positive social change but are not all encompassing. Some students who are not active in the virtual classroom will be high performers. Similarly, a student could qualify for FARM status, but it does not guarantee they will be a low math performer. Ultimately, the results of this study should be disseminated to virtual K-12 programs that track student LMS interactions. More research is required to develop a polished understanding of what student LMS interaction factors impact student mathematics performance.

Implications

There are implications of these results for the two Midwestern, full-time, K-12, virtual schools involved in this study. Primarily, the schools have the opportunity to request dashboard features and alerts that would use student LMS interaction data to better support and intervene with students who are not as active as others. The common theme seen in the literature review of Chapter 2 and the findings of this study was that student LMS interactions impact student performance. As the eLearning environment grows in popularity due to its accessible and flexible nature, research that closes the gap in knowledge of what student participation factors predict performance can greatly assist virtual schools in supporting students.

Conclusion

In this study, I examined whether student LMS interactions could predict student mathematics performance after controlling for student demographic variables known to impact student outcomes. The statistically significant results confirmed that student mouse clicks on lesson content pages predicted final course score in Math 7B after controlling for FARM status and gender. They also confirmed that the number of times students accessed resources predicted final course score in Math 7B after controlling for student demographics. The results of this study contribute to the current body of research illuminating what student LMS interactions predict student performance. The results also align with current theory that student participation in the virtual classroom is related to performance. The findings can be used to inform student support through click activity indicators and dashboards within the virtual classroom. However, the findings also confirmed that more research is required to refine what other student LMS interactions can act as a proxy for student participation in the virtual classroom and subsequently predict student performance.

References

Adejo, O., & Connolly, T. (2017). Learning analytics in a shared-network educational environment: Ethical issues and countermeasures. *Learning*, *8*(4). Retrieved from https://pdfs.semanticscholar.org/fff6/5ac4db436771f715f2170cbb444377fedcc7.pdf

Agudo-Peregrina, Á. F., Iglesias-Pradas, S., Conde-González, M. Á., & Hernández-García, Á. (2014). Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. *Computers in Human Behavior, 31*, 542-550. Retrieved from https://onlinelearningconsortium.org/wp-content/uploads /2016/02/1-s2.0-S074756321300188X-main.pdf

Akçapınar, G., Altun, A., & Aşkar, P. (2015). Modeling students' academic performance based on their interactions in an online learning environment. *İlköğretim Online, 14*(3). Retrieved from http://dergipark.ulakbim.gov.tr/ilkonline/article/ download/5000084958/5000119309

Ary, D., Jacobs, L. C., Irvine, C. K. S., & Walker, D. (2018). *Introduction to research in education*. Boston, MA: Cengage Learning.

Bainbridge, J., Melitski, J., Zahradnik, A., Lauría, E. J. M., Jayaprakash, S., & Baron, J. (2015). Using learning analytics to predict at-risk students in online graduate public affairs and administration education. *Journal of Public Affairs Education*, *21*(2), 247–262. Retrieved from http://www.jstor.org/stable/24369796

Barbour, M. K. (2013). The landscape of K-12 online learning: Examining what is known. *Handbook of Distance Education, 3*, 574–593. Retrieved from https://ies.ed.gov/ncee/ed labs/regions/central/pdf/REL_2017228.pdf

Barbour, M. K. (2017). K–12 online learning and school choice. *The Wiley Handbook of School Choice*, 421.

Berry, W. D., & Feldman, S. (1985). *Multiple regression in practice*. Thousand Oaks, CA: SAGE.

Carpenter, D., Kafer, K., Reeser, K., & Shafer, S. (2015). Evaluating the performance of online K-12 schools. *International Journal on E-Learning, 14*(4), 423-441.

Carver, L. B., Mukherjee, K., & Lucio, R. (2017). Relationship between grades earned and time in online courses. *Online Learning*, *21*(4), 303-313.

Cavanaugh, C., Hargis, J., & Mayberry, J. (2016). Participation in the virtual environment of blended college courses: An activity study of student performance. *The International Review of Research in Open and Distributed Learning*, *17*(3). doi:10.19173/irrodl.v17i3.1811

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, *96*, 42–54. doi:10.1016/j.compedu.2016.02.006

Chatti, M. A., Lukarov, V., Thüs, H., Muslim, A., Yousef, A. M. F., Wahid, U.,… Schroeder, U. (2014). Learning analytics: Challenges and future research directions. *Eleed*, *10*(1).

Choi, J., Walters, A., & Hoge, P. (2017). Self-reflection and math performance in an online learning environment. *Online Learning, 21*(4), 79-102. doi:10.24059/olj.v21i4.1249

Choi, S. P., Lam, S. S., Li, K. C., & Wong, B. T. (2018). Learning analytics at low cost: At-risk student prediction with clicker data and systematic proactive interventions. *Journal of Educational Technology & Society*, *21*(2), 273–290.

Coker, H. (2015). Using data to see what students are doing: A critically enquiring approach to VLE development. *Journal of Perspectives in Applied Academic Practice, 3*(1). Retrieved from https://jpaap.napier.ac.uk/index.php/JPAAP/article /view/146/142

Conde, M. A., & Hernandez-Garcia, A. (2015). Learning analytics for educational decision making. *Computers in Human Behavior, 1*. doi:10.1016/j.chb.2014.12.034

Creswell, J. (2013). *Research design: Qualitative, quantitative, and mixed method approaches*. Thousand Oaks, CA: Sage Publications, Inc.

Curtis, H., & Werth, L. (2015). Fostering student success and engagement in a K-12 online school. *Journal of Online Learning Research, 1*(2), 163-190. Retrieved from https://www.learntechlib.org/p/150962/

Czerkawski, B. C. (2015). When learning analytics meets e-learning. *Online Journal of Distance Learning Administration*, *18*(2).

Dixson, M. D. (2015). Measuring student engagement in the online course: The online student engagement scale (OSE). *Online Learning, 19*(4). doi:10.24059/olj.v19i4.561

Ferguson, R. (2012). The state of learning analytics in 2012: A review and future challenges. Knowledge Media Institute, Technical Report KMI-2012-01. Retrieved from http://wiki.ewu.edu/oit_wiki/images/d/d8/Kmi-12-01.pdf

Field, A. (2017). *Discovering statistics using IBM SPSS Statistics.* Thousand Oaks, CA: SAGE.

Firat, M. (2016). Determining the effects of LMS learning behaviors on academic achievement in a learning analytic perspective. *Journal of Information Technology Education: Research, 15*, 75-87. doi:10.28945/3405

Frankfort-Nachmias, C., & Leon-Guerrero, A. (2014). *Social statistics for a diverse society*. Thousand Oaks, CA: Sage Publications.

Garson, G. D. (2012). *Testing statistical assumptions*. Asheboro, NC: Statistical Associates Publishing.

Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, *28*, 68-84.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. New York, NY: Cambridge University Press.

Gemin, B., & Pape, L. (2017). *Keeping pace with K-12 online learning, 2016*. Durango, CO: Evergreen Education Group.

Gemin, B., Pape, L., Vashaw, L., & Watson, J. (2015). Keeping pace with K-12 digital learning: An annual review of policy and practice. Durango, CO: Evergreen Education Group. Retrieved from http://www.kpk12.com

Goggins, S., & Xing, W. (2016). Building models explaining student participation behavior in asynchronous online discussion. *Computers & Education, 94*, 241-251. doi:10.1016/j.compedu.2015.11.002

Greene, K., & Hale, W. (2017). The state of 21st century learning in the K-12 world of the United States: Online and blended learning opportunities for American elementary and secondary students. *Journal of Educational Multimedia and Hypermedia*, *26*(2), 131–159.

Greller, W., Santally, M. I., Boojhawon, R., Rajabalee, Y., & Kevin, R. (2017). Using learning analytics to investigate student performance in blended learning courses. *Journal of Higher Education Development–ZFHE*, *12*(1).

Gros, B., & García-Peñalvo, F. J. (2016). Future trends in the design strategies and technological affordances of e-learning. *Learning, Design, and Technology: An International Compendium of Theory, Research, Practice, and Policy*, 1–23. doi:10.1007/978-3-319-17727-4_67-1

Gunn, C. (2014). Defining an agenda for learning analytics. In B. Hegarty, J. McDonald, & S. -K. Loke (Eds.), *Rhetoric and reality: Critical perspectives on educational technology* (pp. 683-687). Dunedin, NZ: Dunedin. Retrieved from http://ascilite.org/conferences/dunedin2014/files/concisepapers/257-Gunn.pdf

Harris, R. A., & Nikitenko, G. O. (2014). Comparing online with brick and mortar course learning outcomes: An analysis of quantitative methods curriculum in public administration. *Teaching Public Administration, 32*(1), 95-107.

Harris-Packer, J. D., & Ségol, G. (2015). An empirical evaluation of distance learning's effectiveness in the K-12 setting. *American Journal of Distance Education*, *29*(1), 4–17.

Hrastinski, S. (2009). A theory of online learning as online participation. *Computers & Education, 52*(1), 78-82. doi:10.1016/j.compedu.20 08.06.009

International Business Machines. (2016). IBM SPSS Software. Retrieved from http://www.ibm.com/analytics /us/en/technology/spss/

Iglesias-Pradas, S., Ruiz-De-Azcarate, C., & Agudo-Peregrina, An. F. (2015). Assessing the suitability of student interactions from Moodle data logs as predictors of cross-curricular competencies. *Computers in Human Behavior*, 81. doi:10.1016/j.chb.2014.09.065

International Association for K-12 Online Learning. (2013). Fast facts about online learning. Retrieved from: http://www.inacol.org/cms/wp-content/uploads/2013/11/iNACOL-Fast-Facts-About-Online-Learning-October-2013.pdf

Jo, I., Kim, D., & Yoon, M. (2015). Constructing proxy variables to measure adult learners' time management strategies in LMS. *Journal of Educational Technology & Society*, *18*(3), 214-225.

Joksimović, S., Gašević, D., Loughin, T. M., Kovanović, V., & Hatala, M. (2015). Learning at distance: Effects of interaction traces on academic achievement. *Computers & Education, 87*, 204–217. doi:10.1016/j.compedu.2015.07.002

Khalil, H., & Ebner, M. (2015). Learning analytics: Principles and constraints. In *Proceedings of World Conference on Educational Multimedia, Hypermedia and*

Telecommunications, 1326-1336. doi:10.13140/RG.2.1.1733.2083

Khalil, M., & Ebner, M. (2016). Learning analytics in MOOCs: Can data improve students retention and learning*?* In *EdMedia: World Conference on Educational Media and Technology*, 581-588. Association for the Advancement of Computing in Education. Retrieved from https://www.learntechlib.org/p/173003/.

Kim, D., Park, Y., Yoon, M., & Jo, I.-H. (2016). Toward evidence-based learning analytics: Using proxy variables to improve asynchronous online discussion environments. *The Internet and Higher Education*, *30*, 30–43.

Koç, M. (2017). Learning analytics of student participation and achievement in online distance education: A structural equation modeling. *Educational Sciences: Theory & Practice*, *17*(6).

Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area–Application to student data from Open University of Madrid, UDIMA. *Computers & Education*, *72*, 23–36.

Laureate Education (Producer). (2016a). *Multiple regression.* Baltimore, MD: Author.

Laureate Education (Producer). (2016b). *Regression diagnostics and model evaluation.* Baltimore, MD: Author.

Li, L.-Y., & Tsai, C.-C. (2017). Accessing online learning material: Quantitative behavior patterns and their effects on motivation and learning performance. *Computers & Education*, *114*, 286–297. doi:10.1016/j.compedu.2017.07.007

Liu, F., & Cavanaugh, C. (2012). Factors influencing student academic performance in online high school algebra. *Open Learning: The Journal of Open, Distance and e-Learning, 27*(2), 149-167. doi:10.1080/02680513.2012.678613

Lowes, S. (2014). A brief look at the methodologies used in the research on online teaching and learning. In *Handbook of research on K-12 online and blended learning* (pp. 83–104). Pittsburgh, PA: ETC Press. Retrieved from https://dl.acm.org/citation.cfm?id=2811041

Lowes, S., Lin, P., & Kinghorn, B. (2015). Exploring the link between online behaviours and course performance in asynchronous online high school courses. *Journal of Learning Analytics, 2*(2), 169–194. Retrieved from http://epress.lib.uts.edu.au/journals/index.php/JLA/article/viewFile/4264/5087

Lu, O. H., Huang, J. C., Huang, A. Y., & Yang, S. J. (2017). Applying learning analytics for improving students' engagement and learning outcomes in an MOOCs enabled collaborative programming course. *Interactive Learning Environments*, *25*(2), 220–234.

Luo, Y., Pan, R., Choi, J. H., & Strobel, J. (2018). Effects of chronotypes on students' choice, participation, and performance in online learning. *Journal of Educational Computing Research*, *55*(8), 1069–1087. doi:10.1177/073563311 7697729

Marchetti, R., Wilson, R. H., & Dunham, M. (2016). Academic achievement and extracurricular school activities of at-risk high school students. *Educational Research Quarterly, 39*(4), 3-20.

Martin, C. K., Nacu, D., & Pinkard, N. (2016). Revealing opportunities for 21st century learning: An approach to interpreting user trace log data. *Journal of Learning*

Analytics, *3*(2), 37–87. doi:10.18608/jla.2016.32.4

Martin, F., & Ndoye, A. (2016). Using learning analytics to assess student learning in online courses. *Journal of University Teaching & Learning Practice, 13*(3), 7. Retrieved from https://www.researchgate.net/profile/Florence_Martin4/publication/305851929_Using_Learning_Analytics_to_Assess_Student_Learning_in_Online_Courses/links/58923b71458515aeac945b54/Using-Learning-Analytics-to-Assess-Student-Learning-in-Online-Courses.pdf

Michinov, N., Brunot, S., Le Bohec, O., Juhel, J., & Delaval, M. (2011). Procrastination, participation, and performance in online learning environments. *Computers & Education*, *56*(1), 243–252.

Miller, L. D., Soh, L., Samal, A., Kupzyk, K., & Nugent, G. (2015). A comparison of educational statistics and data mining approaches to identify characteristics that impact online learning. *Journal of Educational Data Mining, 7*(3), 117-150. Retrieved from https://eric.ed.gov/?id=EJ1115333

Molnar, A., Miron, G., Gulosino, C., Shank, C., Davidson, C., Barbour, M. K.,... Nitkin, D. (2017). *Virtual schools report 2017*. Boulder, CO: National Education Policy Center. Retrieved from http://nepc.colorado.edu/publication/virtual-schoolsannual-2017

Moore, M. G. (1989). Three types of interaction. *American Journal of Distance Education, 3*(2), 1-6. Retrieved from http://aris.teluq.uquebec.ca/portals/598/t3_moore1989.pdf

Mwalumbwe, I., & Mtebe, J. S. (2017). Using learning analytics to predict students' performance in Moodle learning management system: A case of Mbeya University of Science and Technology. *The Electronic Journal of Information Systems in Developing Countries*, *79*(1), 1–13.

National Center for Education Statistics. (n.d.). NAEP: 2015 mathematics & reading. Retrieved from https://www.nationsreportcard.gov/reading_ math _2015/#mathematics/context?grade=4

Noesgaard, S. S., & Ørngreen, R. (2015). The effectiveness of e-learning: An explorative and integrative review of the definitions, methodologies and factors that promote e-learning effectiveness. *Electronic Journal of E-Learning*, *13*(4), 278-290.

Osborne, J. W., & Overbay, A. (2004). The power of outliers (and why researchers should always check for them). *Practical Assessment, Research & Evaluation*, *9*(6), 1–12. Retrieved from https://www.researchgate.net/profile/Jason_Osborne2/ publication/242073851_The_Power_of_Outliers_and_Why_Researchers_Should_ Always_Check_for_Them/links/0c96053606c3cb1b2f000000.pdf

Pardo, A., Han, F., & Ellis, R. A. (2017). Combining university student self-regulated learning indicators and engagement with online learning events to predict academic performance. *IEEE Transactions on Learning Technologies*, *10*(1), 82–92.

Pazzaglia, A. M., Clements, M., Lavigne, H. J., & Stafford, E. T. (2016). *An analysis of student engagement patterns and online course outcomes in Wisconsin* (REL 2016–147). Washington, DC: U.S. Department of Education, Institute of

Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Midwest. Retrieved from http://ies.ed.gov/ncee/edlabs.

Percell, J. C. (2016). Data collaborative: A practical exploration of big data in course wikis. *Quarterly Review of Distance Education*, *17*(4), 63–71. Retrieved from https://eric.ed.gov/?id=EJ1142961

Perrotta, C., & Williamson, B. (2018). The social life of learning analytics: Cluster analysis and the 'performance' of algorithmic education. *Learning, Media and Technology*, *43*(1), 3–16.

Phan, T., McNeil, S. G., & Robin, B. R. (2016). Students' patterns of engagement and course performance in a Massive Open Online Course. *Computers & Education*, *95*, 36–44. doi:10.1016/j.compedu.2015.11.015

Poon, L. K., Kong, S. C., Yau, T. S., Wong, M., & Ling, M. H. (2017). Learning analytics for monitoring students participation online: Visualizing navigational patterns on learning management system. In *International Conference on Blended Learning*, 166-176. doi:10.1007/978-3-319 -59360-9_15

Purarjomandlangrudi, A., Chen, D., & Nguyen, A. (2016). Investigating the drivers of student interaction and engagement in online courses: A study of state-of-the-art. *Informatics in Education, 15*(2), 269-286. Retrieved from https://www.mii.lt/informatics_in_edu cation/pdf/infedu.2016.14.pdf

Robson, C., & McCartan, K. (2016). *Real world research*. West Sussex, UK: John Wiley & Sons.

Reyes, J. (2015). The skinny on big data in education: Learning analytics simplified. *Techtrends: Linking Research & Practice to Improve Learning*, *59*(2), 75-80. doi:10.1007/s11528-015-0842-1

Reyna, J. (2016). Implementing e-learning across the school of education: A case study. *International Journal on E-Learning*, *15*(1), 101–120. Retrieved from https://www.learntechlib.org/p/41947/

Rienties, B., & Toetenel, L. (2016). The impact of learning design on student behaviour, satisfaction and performance: A cross-institutional comparison across 151 modules. *Computers in Human Behavior*, *60*, 333. doi:10.1016/j.chb.2016.02.074

Siemens, G. (2010). What are learning analytics? Retrieved from http://www.elearnspace.org/blog/2010/08/25/what-are-learning-analytics/

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review, 46*(5), 30. Retrieved from https://eric.ed.gov/?id=EJ950794

Society for Learning Analytics Research. (2012). About. Retrieved from http://www.solaresearch.org/about/

Stapel, M., Zheng, Z., & Pinkwart, N. (2016). An ensemble method to predict student performance in an online math learning environment. In *EDM* (pp. 231–238). Retrieved from https://pdfs.semanticscholar.org/e1ca/981f55484ade30a8132d3a9492994f989319.pdf

Strang, K. D. (2016a). Do the critical success factors from learning analytics predict student outcomes? *Journal of Educational Technology Systems, 44*(3), 273-299. doi:10.1177/0047239515615850

Strang, K. D. (2016b). Predicting student satisfaction and outcomes in online courses using learning activity indicators. *Journal of Interactive Learning Research*, *27*(2), 125-152.

Strang, K. D. (2017). Beyond engagement analytics: Which online mixed-data factors predict student learning outcomes? *Education and Information Technologies; New York*, *22*(3), 917–937. doi:10.1007/s10639-016-9464-2

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, *47*, 157–167.

Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2017). Towards actionable learning analytics using dispositions. *IEEE Transactions on Learning Technologies, 10*(1), 6-16. doi:10.1109/TLT.2017.2662679

Ullman, C., & Rabinowitz, M. (2004). Course management systems and the reinvention of instruction. *The Journal*. Retrieved from https://thejournal.com/articles/2004/10/01/coursemanagement-systems-and-the-reinvention-of-instruction.aspx

U.S. Department of Education. (2015). Family Educational Rights and Privacy Act (FERPA). Retrieved from https://ed.gov/policy/gen/guid/fpco/ferpa/index.html

Verzani, J. (2004). *Using R for introductory statistics*. Boca Raton, FL: CRC Press.

Warner, R. M. (2013). *Applied statistics: From bivariate through multivariate techniques* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Wickens, M. R. (1972). A note on the use of proxy variables. *Econometrica: Journal of the Econometric Society*, 759-761. Retrieved from http://www.jstor.org/stable/pdf/ 19129 71.pdf?refreqid=excelsior%3Ab018 0794eb09a4501455ae2300fdf3ac

Winne, P. H. (2011). A cognitive and metacognitive analysis of self- regulated learning. In B. J. Zimmerman, & D. H. Schunk (Eds.), *Handbook of Self-regulation of Learning and Performance* (pp. 15-32). New York, NY: Routledge.

Winne, P. H. (2017). Leveraging big data to help each learner and accelerate learning science. *Teachers College Record*, *119*(3), 1–24.

Woodworth, J. L., Raymond, M. E., Chirbas, K., Gonzalez, M., Negassi, Y., Snow, W., & Van Donge, C. (2015). *Online charter school study*. Stanford, CA: Center for Research on Education Outcomes. Retrieved from https://credo.stanford.edu/pdfs/OnlineCharterS tudyFinal2015.pdf

Xing, W., Guo, R., Petakovic, E., & Goggins, S. (2015). Participation-based student final performance prediction model through interpretable genetic programming: Integrating learning analytics, educational data mining and theory. *Computers in Human Behavior, 47*, 168-181. doi:10.1016/j.chb.2014.09.034

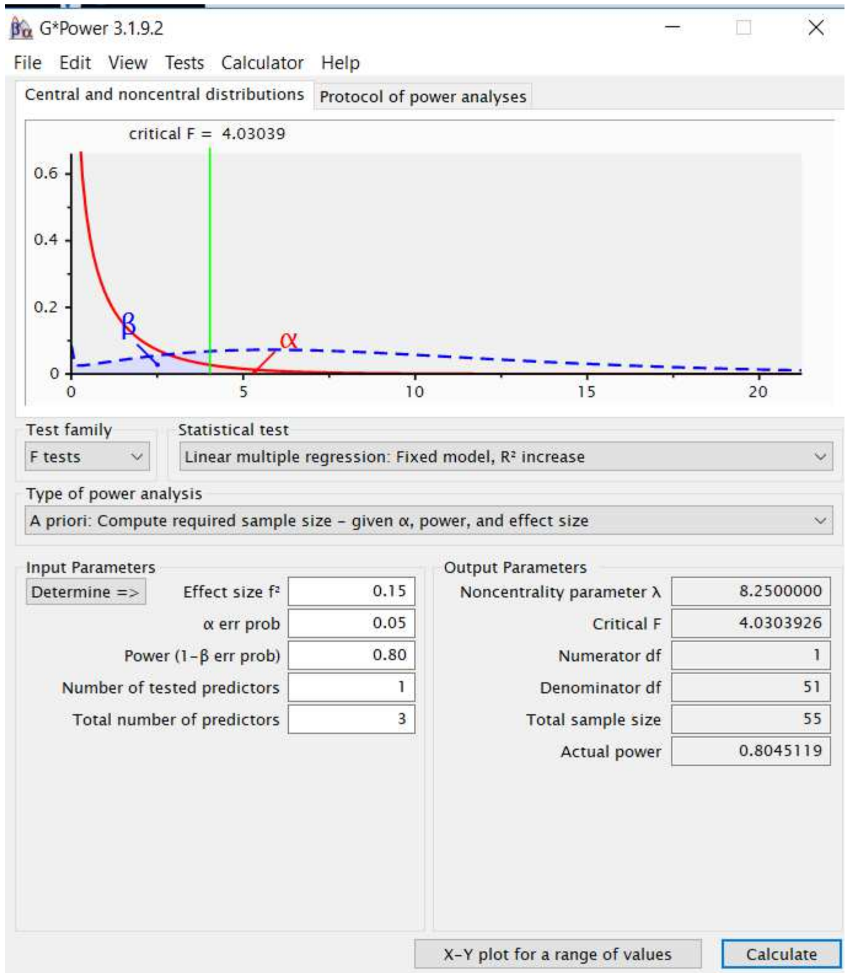
Yarbrough, J. L., Cannon, L., Bergman, S., Kidder-Ashley, P., & McCane-Bowling, S. (2017). Let the data speak: Gender differences in math curriculum–based measurement. *Journal of Psychoeducational Assessment, 35*(6), 568–580. doi:10.1177/0734282916649122

Yeung, C. Y., Shum, K. H., Hui, L. C. K., Chu, S. K. W., Chan, T. Y., Kuo, Y. N., & Ng,

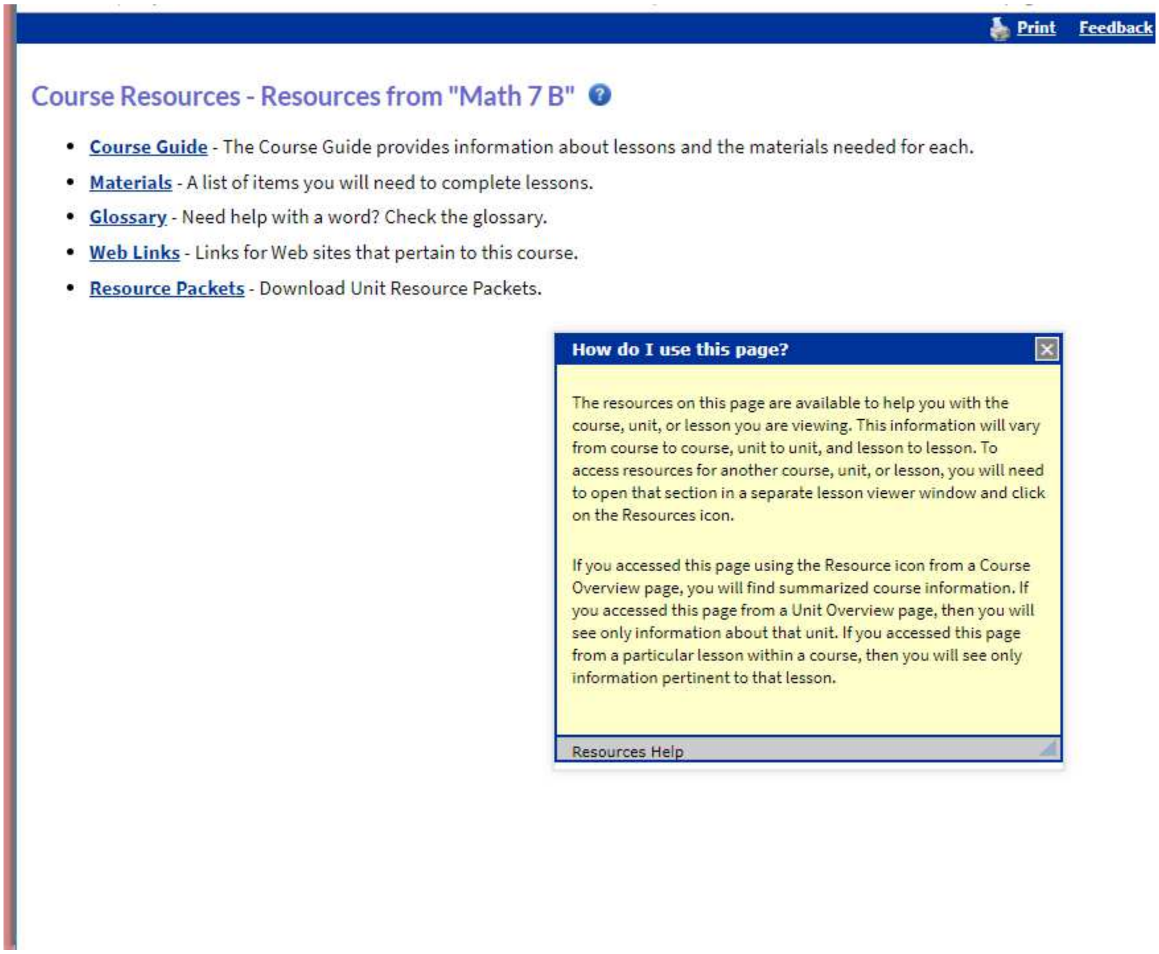
Y. L. (2017). Exploring characteristics of fine-grained behaviors of learning mathematics in tablet-based e-learning activities. *International Conference on eLearning, 5*. Retrieved from https://files.eric.ed.gov/fulltext/ED579374.pdf

You, J. W. (2015). Examining the effect of academic procrastination on achievement using LMS data in e-learning. *Journal of Educational Technology & Society, 18*(3), 64. Retrieved from http://www.jstor.org/stable/pdf/jeductechsoci.18.3.64.pdf

Appendix A: G*Power Analysis



Appendix B: Seventh Grade Mathematics Virtual Backpack



Appendix C: IRB Approval Number

The IRB approval number for this study is 10-09-18-0342797.