Logging On to Improve Achievement: Evaluating the Relationship between Use of the Learning Management System, Student Characteristics, and Academic Achievement in a Hybrid Large Enrollment Undergraduate Course

by

JOHN C WHITMER
B.A. (Central Connecticut State University)
M.A. (University of California, Davis)

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF EDUCATION

in

Educational Leadership

in the

OFFICE OF GRADUATE STUDIES

of the

UNIVERSITY OF CALIFORNIA

DAVIS

Approved:

_______________________________
Dr. Paul Porter, Chair

_______________________________
Dr. Jamal Abedi

_______________________________
Dr. Janet Gong

Committee in Charge
2012
This work is licensed under a Creative Commons Attribution-ShareAlike 3.0 Unported License.
ABSTRACT

Student persistence in higher education is a critical national issue with implications for economic development and personal fulfillment. Mirroring national trends, approximately 50% of students entering the California State University (CSU) system graduate within six years, with much lower rates for students from under-represented minority groups. Academic technologies such as the Learning Management System (LMS) have been suggested as means to implement instructional reforms to improve persistence at a large scale. Data from academic technology use has the potential to assess the effectiveness of these reforms.

This study expands current research in student persistence through a quantitative study of student LMS use, background demographic characteristics, and achievement in a large enrollment (n = 377) course that was offered at CSU Chico in Fall 2010. This course was conducted in a hybrid format in which LMS activities replaced one of the weekly meeting sessions. The study analyzed four variables that indicated the instructional purpose of LMS use (administration, assessment, content, engagement). In addition, nine variables with conventional demographic characteristics were analyzed. Student course grade was used as the dependent variable. Correlation analysis was used to investigate the relationship between each variable and student grade. Regression models were calculated to investigate the relationship between the combined variables and student grade.

Separate LMS use variables explained over four times the variation in the final grade compared to the student characteristic variables. Combined LMS use resulted in a model that explained 25% of the variation, which was increased to 35% with the addition of student characteristic variables. For at-risk students, LMS use had a 25% smaller effect.
These results suggest that the LMS is a potentially valuable resource to carry out instructional reforms to improve persistence. Further, data from the LMS can be used as a meaningful indicator of student educational effort. At the course level LMS use appears to be a much better predictor of student achievement than traditional demographic variables. However, lower results for at-risk students invites further research to explain this finding and explore means to remedy it.
DEDICATION

I dedicate this study to my wife Synda, whose confidence in my abilities, encouragement to keep going, and support in more ways than I can count made it possible to complete my graduate studies.
ACKNOWLEDGEMENTS

I am grateful to many people for their support of this research project, through their personal support to keep up the work, through their academic support with developing the ideas and concepts it expresses, and in the organizational support to get the data and think through the meanings and implications. Dissertations may have single authors, but they are collaborative efforts.

First and foremost I thank my wife Synda who spent countless days "leaving me alone" and taking care of our family obligations to provide me the dedicated time I needed to complete my studies. I appreciate your patience, resilience, and determination to help me get this done. I promise to walk at the ceremony and complete this process with all the attending rites of passage.

It is poetic that I began my graduate coursework for the CANDEL program in Dr. Porter's leadership course, and complete the program with his guidance. His help thinking through my problem and the broader implications of the research was crucial to this study. I am also thankful to the other members of my dissertation committee, Dr. Jamal Abedi and Dr. Janet Gong, who have been thoughtful, courteous, and responsive to my questions throughout the dissertation process. Their expertise and reflections shaped this research in many ways.

My colleagues Kathy Fernandes and William A. Allen at Chico State were essential to not only get access to the data set, but also to attain organizational approvals and think through the best variables given campus procedures. They have also been generous with their time and offered insights that can only come from being situated within an organizational context. Given that this project was one of those "other projects as assigned", I'm especially grateful for the time they spent to make this project a reality.
The issues in the dissertation were developed through many courses in the CANDEL program and in my career. I am thankful for the flexibility built into the program that allowed me to address the concepts in the program through projects that were meaningful to my professional interests and personal inclinations. This is a programmatic focus of CANDEL, but there were several faculty whose input on early projects was instrumental in developing this dissertation, including: Dr. Gloria Rodriguez, Dr. Carlos Ayala, Dr. Viki Montera, Dr. Michal Kurlaender, and Dr. Philip Young. I also would like to specially thank Dr. Paul Heckman and Dr. Paul Porter for their vision and inspiration for the program. CANDEL provides an opportunity for deep thinking, or "noodling", that is rare in professional programs.

Finally, I would like to thank Suzette Smiley-Jewel, for her help applying Occham's razor where it was needed and helping me expand where it was needed.

So, Synda, what should we do on Sunday?
# TABLE OF CONTENTS

## CHAPTER 1 INTRODUCTION
- Problem Statement ................................................................. 1
- Higher education student persistence ........................................ 4
- Learning Management System (LMS) adoption and educational technology research ........................................ 5
- Changing demographics of higher education .......................... 7

## CHAPTER 2 LITERATURE REVIEW
- Area 1: Higher Education Student Persistence .......................... 23
  - Overview ............................................................................. 23
  - Distinguishing student persistence from institutional retention .......................................................... 24
  - Tinto’s interactionalist theory of persistence .............................................................. 25
  - Conditional influences of race, socio-economic status and gender on persistence .................. 26
  - URM student persistence .................................................. 27
  - Socio-economic status (SES) and persistence .......................................................... 29
  - Gender and persistence ..................................................... 31
  - Contemporary research methodology: multivariate analyses and variable interactions .......... 33
  - Classroom instruction and persistence ........................................ 33
  - High enrollment courses and academic technology ................................................. 34
  - Course grades as indicators of persistence ................................................. 35
- Area 2: The No Significant Difference Debate and Internet Technology Research .................. 35
  - Overview ............................................................................. 35
  - The no significant difference debate ........................................... 36
  - Contemporary research on Internet technology in education ........................................ 38
  - URM student achievement in technology-enhanced courses ........................................ 40
- Area 3: Learner Analytics .......................................................... 42
  - Overview ............................................................................. 42
  - Academic analytics: analyzing campus data for institutional improvement .................. 42
  - Learner analytics: bringing analysis to learning contexts ............................................. 44

## CHAPTER 3 ............................................................................. 51

## RESEARCH METHODOLOGY .................................................... 51
- Overview ............................................................................. 51
- Research Design ..................................................................... 51
- Course selection ..................................................................... 52
- Protection of human subjects .................................................. 53
- Data Sources ......................................................................... 53
- LMS use: WebCT™ web server log ........................................... 54
CHAPTER 5 .............................................................................................................................................. 102
DISCUSSION ............................................................................................................................................. 102
Overview ........................................................................................................................................................... 102
Conclusions for Instructional Practices ........................................................................................................... 103
Diverse student populations ......................................................................................................................... 103
LMS use is a systematic indicator of academic achievement ...................................................................... 104
Students consistently use the LMS across categories ............................................................................... 106
Student characteristics are low predictors of achievement ................................................................. 107
Combined LMS and student characteristic data makes small advance ............................................... 108
LMS use data is proxy for student effort ................................................................................................. 109
LMS use and expanded instructional method effectiveness ......................................................................... 110
LMS use is less effective for at-risk students ............................................................................................... 111
Demographic data in the Postmodern age ............................................................................................... 112
Conclusions about Data for Learning Analytics ..................................................................................... 113
Filtering LMS log file for educationally relevant activity .......................................................................... 113
Data range and limits of dichotomous and categorical variables .............................................................. 115
Contributions to Previous Learner Analytic Research Studies ................................................................... 115
Methodological Limitations and Potential Revisions ............................................................................... 117
Implications for Practitioners and Suggestions for Future Research ..................................................... 117
Implications for practitioners ..................................................................................................................... 118
Suggestions for future research ................................................................................................................ 118
References ................................................................................................................................................ 120
LIST OF TABLES

Table 1: Change in Post-Secondary institutional enrollment by race/ethnicity ........................................ 7
Table 2: LMS Use Categories and Corresponding LMS Tools ................................................................. 11
Table 3: Student Characteristic Variables and Previous Studies ............................................................ 14
Table 4: Key Terms used in Study ............................................................................................................ 16
Table 5: Empirical Study Precedents in Learner Analytics: Contexts and Findings .............................. 46
Table 6: Source Databases used for Data Set ......................................................................................... 53
Table 7: LMS Use Tools and Activity Variables Included in Research .................................................. 56
Table 8: Count of Majors in each College Variable ............................................................................. 59
Table 9: Student characteristic Variables and Correlation Coefficient Methods ............................. 63
Table 10: Distribution of Participant Gender Race, Socio-Economic Status, and Age .................... 70
Table 11: Course Participant Demographics: Additional Family and Educational Background .......... 72
Table 12: Cross Tabulations of Demographic Variables ....................................................................... 74
Table 13: Additional Cross Tabulations of Student Variables: Pell, Race, URM ............................. 75
Table 14: Additional Cross Tabulations of Student Variables: URM and Major College .................. 76
Table 15: Intercorrelation Analysis of Student Characteristic Independent Variables ..................... 78
Table 16: Variables Omitted from Analysis for Missing Data ............................................................... 79
Table 17: Descriptive Data on LMS Use: Summary and by Tool ......................................................... 84
Table 18: Descriptive Data on LMS Use: Categories Used for Inferential Analysis .......................... 86
Table 19: Factor Analysis Results of LMS Use Categories ................................................................. 87
Table 20: Correlation results of LMS use with course grade ............................................................... 90
Table 21: Correlation results of Student Characteristics with Final Course Grade ............................ 92
Table 22: Multivariate Regression Results of Restricted Model 1: Final Grade on LMS Use .... 95
Table 23: Multivariate Regression Results of Restricted Model 2: Final Grade on LMS Use and Student Characteristics

Table 24: Multivariate Regression results of Complete Model: Final Grade on LMS Use, Student Background Characteristics, and Student Background Characteristic Interactions

Table 25: Comparison of Partial and Complete Regression Model Results

Table 26: Comparison of Regression Models Results by Population Sub-Sample and Selected Completed Regression Results
LIST OF FIGURES

Figure 1: California State University six-year graduation rates (1975-2001) .............................. 5
Figure 2: Frequency of undergraduate student usage of the LMS ................................................. 6
Figure 3: Tinto’s interactionalist conceptual model of student dropout ...................................... 26
Figure 4: High School GPA Histogram ........................................................................................ 73
Figure 5: Filtered vs. Unfiltered LMS Use by Category and Tool ............................................... 82
Figure 6: Comparison of Histograms for Total LMS Hits vs. Total LMS Dwell Time ............... 83
Figure 7: Comparison of Histograms for Administrative Activity Hits vs. Administrative Activity Dwell Time ........................................................... 84
Figure 8: Frequency Distribution of Course Grade ...................................................................... 88
Figure 9: Scatterplot of Assessment Activity Hits and Course Grade ......................................... 91
Figure 10: Scatterplot of HS GPA and Course Grade .................................................................. 93
CHAPTER 1
INTRODUCTION

Problem Statement

President Barak Obama launched a national “College Completion Agenda” in 2009 that set a goal for the United States to regain its position as the nation with the highest proportion of citizens with college degrees by 2020 (President Obama, 2009). The basis for his agenda is that, currently, only approximately 50% of the students who attend college in the United States graduate within six years. Moreover, students from under-represented minority (URM) groups graduate at much lower rates (Engle & Theokas, 2010).

One of the major barriers for increasing graduation rates is successfully getting students through large enrollment introductory courses (Bailey & Alfonso, 2005; Gainen & Willemsen, 1995). These courses are often taught in a didactic teaching style and taken by a large proportion of the students attending a college or university: in one study, the 25 highest-enrolled courses generated almost 50% of community college enrollments and one-third of enrollments in four-year institutions (Twigg, 2005).

The large numbers of students in introductory classes can make it difficult for faculty to effectively connect with the students. The sheer size of the classes can limit the instructor’s ability to conduct interactive activities such as group discussions, project-based learning or personalized feedback. It is also difficult to conduct administrative activities such as making announcements and distributing course notes. However, with declining state budget allocations, reducing course enrollment is not an option.

Because scholars have identified didactic teaching as a contributing factor to low pass rates in large enrollment courses (Harrison, 1989; Seymour & Hewitt, 1997), many universities
have implemented Internet-based technologies to foster student-instructor interaction online through discussion forums and in-class email (Arasasingham, Martorell, & McIntire, 2011; Deslauriers, Schelew, & Wieman, 2011; Wallin, 1997). One such technology is the Learning Management System (LMS), an application that includes many common instructional functions within an online “course”. Using the LMS, faculty can provide announcements about a course, post learning materials, conduct assessments, create discussions forums and lead chat sessions. The benefit of the LMS is that large numbers of students can participate in these activities and the LMS can be accessed at any time.

To date, a number of empirical studies have been completed to define the relationship between student LMS use and academic achievement (J. P. Campbell, 2007; Macfadyen & Dawson, 2010; Morris, Finnegan, & Wu, 2005; Rafaeli & Ravid, 1997). The studies use statistical techniques on large data sets compiled by educational technologies to conduct educational research (M. Brown, 2011). These studies, known as “learning analytics”, select a subset of available LMS use measures (e.g., posting a discussion message, total LMS hits) as their independent variables and regress these variables against course grade. The relationships have been used to build predictive models, that in some cases, can identify low-performing students at-risk of failing a course.

These predictive models advance upon traditional demographic data measures by providing real-time data that can immediately affect students. Predictions of low achievement provided to students within a course can motivate them to change their behaviors and improve their course grade (Arnold, 2010; Fritz, 2011). LMS data sources are also easier to access for faculty and educational technologists than student records with confidential personal information.
However, the data has not been analyzed to shed light on what LMS use signifies in terms of the broader educational purpose underlying student activity. Is LMS use related to the broader pedagogical functions that underlie detailed technical functions and features? Further, only one study has analyzed whether student characteristics affect the magnitude of these predictions (J. P. Campbell, 2007). Despite studies demonstrating lower achievement by URM (URM) students in online courses compared to white and Asian students (Lamkins, 2004; Welsh, 2007), the impact of race/ethnicity and other factors placing students at-risk of not succeeding in an online course has not been explored.

From a methodological perspective, learning analytics discard a large amount of LMS use data. As a result, these studies do not analyze the complete record of how students are using this technology. Moreover, there is scarce conceptual justification provided for the measures that are included in the analysis. Parsimony and statistical analysis require a small number of variables, but there is no established approach to including some variables and excluding others.

This study conducts an ex post facto analysis with data from an LMS log file and student records with the purpose of addressing the following questions: Is frequency of LMS use related to student academic achievement? Furthermore, does it matter, in terms of student achievement, what the broader pedagogical functions underlying LMS use are? For example, is using the LMS for activities that promote engagement (e.g., discussion, email) more highly related to improved achievement than administrative activities (e.g., calendar, announcements)? Is LMS use a stronger predictor of academic achievement than student background characteristics (e.g., race/ethnicity, socio-economic status) and current enrollment information (e.g., first-time student, college major)? And perhaps most important, is use of the LMS as effective for at-risk
students (as predicted by race/ethnicity, socio-economic status, and gender) as it is for students whose backgrounds create fewer barriers to academic success?

This study addresses these questions by examining a large introductory college course that was redesigned to integrate technology in order to improve course effectiveness and increase enrollment. The results from this study have implications for scholars of higher education persistence, technology evaluation, and learning analytics. Educational leaders with responsibilities in student success or persistence programs, academic technology staff and managers and policymakers may also find this study relevant to their concerns.

**Higher education student persistence.**

In recent years, there has been concern that higher education institutions in the United States graduate enough students to meet workforce requirements, as some estimates predict that 20 of the top 30 new jobs will require a college education (Engle & Theokas, 2010; Johnson & Sengupta, 2009). There is a clear relationship between educational attainment and economic value. The latest Bureau of Labor Statistics figures reported that people with a bachelor’s degree earn $1,038 weekly compared to $626 for people with a high school diploma and no college experience (2010). People with a bachelor’s degree have nearly a 5% lower unemployment rate than people with a high school diploma. In the current economic recession, economic attainment is more critical than ever.

In California, increasing the student graduation rate has been a long-standing concern. This is especially true of the California State University (CSU) and the California Community Colleges systems (Johnson & Sengupta, 2009; Shulock, 2007). The estimated six-year graduation rate for degree-seeking first-time freshman in the 2004 CSU cohort across all 23 campuses is 52.4% (California State University Analytic Studies, 2011a). As illustrated in
Figure 1, the CSU graduation rate has increased over the past 25 years, however, the rate of change will not be sufficient to meet student demand for higher education and California’s economic needs (Johnson & Sengupta, 2009). Figure 1 indicates the gap in the graduation rate between first-time freshman and transfer students and also points to a need to focus on first-time students (California State University Analytic Studies, 2008). Another area needing focus is the impact of a student’s racial/ethnic background on their educational attainment: the graduation rate varies dramatically (between 38.3% and 58.8%) when this background is considered (California State University Analytic Studies, 2011a).

Figure 1. California State University six-year graduation rates (1975-2001). Reprinted from “Closing the Gap: Meeting California’s Need for College Graduates,” by the Public Policy Institute of California, 2009.

Learning Management System (LMS) adoption and educational technology research.

Today’s mega-enrollment introductory classes present challenges for educators trying to conduct interactive activities due to the sheer number of students. One remedy to increase interaction in these courses is through the use of “high-impact activities”, which are frequently conducted via educational technologies (American Association of State Colleges and Universities, 2010; M. Taylor, 2010; United States Department of Education, 2010). High-
impact activities are teaching and learning practices, such as collaborative assignments, undergraduate research, common intellectual experiences, and others that have been “widely tested and have been shown to be beneficial for college students from many backgrounds” (Kuh, 2008). Although these practices are frequently implemented across an institution, they also take place within courses (Pascarella & Terenzini, 2005).

The LMS is the most frequently used academic technology in higher education and is commonly used to conduct high-impact activities. The 2010 Campus Computing Survey reported that 93% of higher education institutions have deployed an LMS (Green, 2010). The Educause Consortium for Applied Research (ECAR) Survey of Undergraduates and Information Technology (S. D. Smith & Caruson, 2010) provides additional information about LMS use within institutions. As indicated in Figure 2, almost two-thirds of students reported using the LMS for one or more courses, and one-third of the students reported using the application every day.

Figure 2. Frequency of undergraduate student usage of the LMS. Reprinted from “The ECAR Study of Undergraduate Students and Information Technology,” by S.D. Smith and J.B. Caruson, 2010, p. 15. Copyright 2010 by the EDUCAUSE Center for Applied Research.
Although the LMS is used frequently, students vary in their perception of the application’s value. In the ECAR study, 51% of student respondents reported a positive or very positive experience using the LMS (p.14). However, the student responses were found to positively correlate with the amount of LMS use; that is, the more a student used the LMS, the more positive was their attitude toward it.

**Changing demographics of higher education.**

Over the past 40 years, the demographic profile of college students has changed dramatically. The National Center for Education Statistics (NCES) reports that between 1976-2010 the percentage of URM students enrolled in degree-granting institutions rose 300%, compared to a 91% increase in overall enrollments (2012). The rate within this population grouping ranged from 194% (Black students) to 614% (Hispanic students), as detailed in Table 1. In terms of total numbers, White and Asian/Pacific Islander students still represented almost two-thirds of enrolled students in 2010, but URM populations are becoming a significant presence in higher education enrollment. Students from these backgrounds enter college with multiple barriers to college achievement including cultural assimilation, social support, and campus racial climate.

Table 1

*Change in Post-Secondary institutional enrollment by race/ethnicity. Table adapted from data in NCES Digest of Educational Statistics (2011)*

<table>
<thead>
<tr>
<th>Racial/Ethnic Group</th>
<th>Fall enrollment (in thousands)</th>
<th>Percent Increase</th>
<th>Percent of Enrollment (2010)</th>
</tr>
</thead>
<tbody>
<tr>
<td>All Students</td>
<td>10,986</td>
<td>21,016</td>
<td>91%</td>
</tr>
<tr>
<td>White</td>
<td>9,076</td>
<td>12,723</td>
<td>40%</td>
</tr>
</tbody>
</table>
NCES predicts that this trend will continue, with projected increases in the 2009-2020 period of 1% for White students, 25% for Black students, and a 46% increase for Hispanic students (2011). In addition to race/ethnicity, differential growth in enrollment is also predicted by gender. In the same period, an increase of 8% is predicted for men and 16% increase for women (NCES, 2011). The increased participation of URM students, combined with declining state support for higher education and a difficult economic climate, make it likely that student socio-economic status will become an increasingly relevant issue as it affects the ability of students to reach their higher education goals.

**Statement of need.**

Previous studies have examined the relationship between the frequency of student LMS use and academic achievement (J. P. Campbell, 2007; Macfadyen & Dawson, 2010; Morris et al., 2005; Rafaeli & Ravid, 1997). This research has resulted in the creation of predictive models that can provide students, faculty, and advisors with an “early warning” that the student is likely to fail or receive a low course grade (Arnold, 2010; Fritz, 2011; Macfadyen & Dawson, 2010). Alerts are triggered by a low frequency of LMS use relative to other students (e.g., a low number of logins or a low number of discussions messages read). However, it is important to recognize that the predictive models have limited explanatory value. They indicate that a student with relatively less frequency using a specific LMS feature is likely to receive a low grade, but they

<table>
<thead>
<tr>
<th>Race/Ethnicity</th>
<th>2010</th>
<th>2011</th>
<th>Increase</th>
<th>Gender Male 2009-2010</th>
<th>Gender Female 2009-2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian/Pacific Islander</td>
<td>198</td>
<td>1,282</td>
<td>548%</td>
<td>6%</td>
<td>300%</td>
</tr>
<tr>
<td>URM Students</td>
<td>1,493</td>
<td>5,977</td>
<td>300%</td>
<td>28%</td>
<td>28%</td>
</tr>
<tr>
<td>American Indian</td>
<td>76</td>
<td>196</td>
<td>158%</td>
<td>1%</td>
<td>1%</td>
</tr>
<tr>
<td>Black</td>
<td>1,033</td>
<td>3,039</td>
<td>194%</td>
<td>14%</td>
<td>14%</td>
</tr>
<tr>
<td>Hispanic</td>
<td>384</td>
<td>2,741</td>
<td>614%</td>
<td>13%</td>
<td>13%</td>
</tr>
</tbody>
</table>
do not consider the pedagogical function of that feature or how student characteristics might change the relationship between LMS use and academic achievement.

The current research project aims to increase the explanatory value and accuracy of LMS data, with the overall goal of improving graduation rates. By grouping LMS use measures (e.g., posting to a discussion forum) into broader pedagogical categories (e.g., engagement), this research helps to uncover the deeper ways that student use of the LMS is affecting achievement. For example, do students who use the LMS frequently for engagement-related activities have higher grades than students who use them less frequently? These findings could indicate effective uses of the LMS and suggest future research topics to better understand the relationship of courses using these activities with achievement. By including student characteristics, this research reveals to what degree a student’s background and current college enrollment information predicts their course achievement.

**Learning analytics.**

As databases and web servers collect increasing amounts of information from consumers, companies are analyzing this data to conduct new forms of market research and product development (Economist, 2010; Economist Intelligence Unit, 2011). The term “analytics” has been coined to describe analysis based in user data for market research and overall decision making (LaValle, Hopkins, Lesser, Shockley, & Kruschwitz, 2010). Analytics can be applied to academic data sources to analyze how learning behaviors impact student achievement.

Enterprise Resource Planning (ERP) systems at university campuses are centralized databases with student characteristics, course participation information, and academic achievement information. In 2005, the term “academic analytics” was coined to describe the use of ERP data sources for institutional decision-making (Goldstein & Katz). A subfield called
“learner analytics” has evolved that considers the LMS and other academic technologies as potential sources of data (M. Brown, 2011).

While academic analytics is used to conduct broad measures of institutional effectiveness and to identify areas that need improvement (J. P. Campbell, DeBlois, & Oblinger, 2007; Norris, Baer, & Leonard, 2008), learner analytics focuses on the learning context itself and efforts to improve the learning environment (Siemens, 2011). For example, one academic analytics study found that students taking online introductory math and English courses had lower achievement rates than similar students taking in-person versions of the same courses (Xu & Jaggars, 2011). These findings indicated a problem, but they did not specify what learning practices were missing or ineffective. Learner analytics could be applied to the detailed actions of students taking the online courses and provide insight into which of their activities were ineffective. With this knowledge, it is possible to redesign course activities for improved student achievement.

Academic technology leaders perceive strong interest in learning analytics from their institutions. In a 2012 study of 339 higher education institutions, 28% of the respondents reported that analytics was a priority for the entire institution and an additional 69% responded that analytics was a major priority for at least some of the institution (Bichsel, 2012). In addition, 86% of the respondents predicted that in two years analytics would be a larger institutional priority than it is now.

Articles and reports in professional journals provide guidelines for strategic organizational and management processes that support the development of learning analytics (Bichsel, 2012; Goldstein & Katz, 2005; Pirani & Albrecht, 2005; Ravishanker, 2010). However, these guidelines do not help leaders and staff charged with implementing an analytics initiative make decisions about which data sources are most effective and how common data
sources, like the LMS, can be analyzed to address strategic institutional questions such as student achievement. This study seeks to contribute to the literature supporting development of analytics through discussion of the data sources and processes required to conduct this case study and the identification of gaps and potential improvements for future analytics efforts.

**Student LMS use categories.**

This study investigates student LMS use and student characteristics in order to increase our understanding of how these factors affect student academic achievement in a redesigned large enrollment undergraduate course.

One method to evaluate the impact of the LMS on student achievement is by categorizing student LMS use by pedagogical functions (Janossy, 2008). When a student uses the LMS, each action that they take (e.g., reading an announcement, posting a discussion message, submitting an assessment) is recorded in the website log file. Dawson and McWilliam (2008) suggest four categories to classify LMS use: (a) administration; (b) assessment; (c) content; and (d) engagement. These categories correspond to broadly understood definitions of LMS use by academic technologists in practice.

For this study, student LMS use was grouped by categories based on the tool used for each LMS action as defined in Table 2. The detailed list of possible actions within each tool is included in the research methods section of this study.

Table 2

*LMS Use Categories and Corresponding LMS Tools*

<table>
<thead>
<tr>
<th>LMS Use Category</th>
<th>Definition</th>
<th>LMS Tools within Category</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administration Activity Hits</td>
<td>Student LMS website hits related to course planning and administrative activities.</td>
<td>Announcement, Calendar</td>
</tr>
<tr>
<td>Assessment Activity</td>
<td>Student LMS website hits related to assessment</td>
<td>Assessment</td>
</tr>
</tbody>
</table>
Preliminary analysis of the data indicated that the frequency of student LMS use was consistent across all categories, which creates a limitation to statistical analysis of these categories. However, it is still important to include these categories in the analysis, given the findings from previous research and literature in educational technology that have highlighted the importance of studying the instructional method underlying technology use (Clark, 1983; Kozma, 1994b; Russell, 2001). LMS use outside of these tools was not considered in the study. Also excluded were duplicate entries; there are some actions within the tools that create duplicate log file entries for the same action (e.g., discussion forum post, discussion home viewed). This process and excluded actions are described in more detail in the methods section of this proposal.

**Student background characteristics.**

Students do not enter the university as a blank slate, but instead begin their studies with cultural identities, prior academic formation, and personal motivations that impact their future academic achievement. Academic institutions track some of these characteristics through application processes and subsequent reporting. An established body of previous research has examined the relationship between demographic variables (e.g. race/ethnicity, age, gender) and persistence (Astin, 1975; Peltier, Laden, & Matranga, 1999).
In these studies, student racial/ethnic background was found to have a persistent significant relationship with college completion. However, research using other demographic variables (e.g. gender) has mixed outcomes: in some studies it is found to be significant, and in others it is not significant (Peltier et al., 1999). With the increasing participation of URM students in higher education, Reason (2003) advocated to increase studies on demographic variables and their interactions. Reason found that although individual variables may not be related to achievement, the interactions of multiple variables (e.g. race/ethnicity and socio-economic status) showed promise for future research, as the interactions more closely mapped to the complex identities of students currently entering higher education institutions than isolated variables.

Few studies have examined these variables in the context of student achievement in individual courses. A 1978 literature review found few significant relationships between these variables and academic achievement (Margrain). Recent studies of online courses have found that current college achievement is the strongest predictor of future achievement (Brallier, Palm, & Gilbert, 2007; Bukralia, 2009). Although this finding is useful to create accurate predictions, it is a tautology – students doing well are likely to continue doing well. This variable does not reveal why students are achieving these grades. There are many potential issues that can effect student achievement in a single course, which makes it difficult to identify factors that systematically effect all students in the course. However, this is the context in which the most immediate increases in student achievement can be achieved, making it an ideal context for research that aims to improve student persistence.
This research project builds on these previous study findings by analyzing nine variables that indicate the relationship between a student’s demographic background and current college enrollment information and their final course grade. The nine variables are identified in Table 3.

Table 3

*Student Characteristic Variables and Previous Studies*

<table>
<thead>
<tr>
<th>#</th>
<th>Variable</th>
<th>Definition</th>
<th>J. P. Campbell (2007)</th>
<th>Barber &amp; Sharkley (2012)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Gender</td>
<td>Gender.</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>2.</td>
<td>URM Student</td>
<td>Student racial/ethnic group under-represented in higher education student enrollment.</td>
<td>Y (race)</td>
<td>Y (race)</td>
</tr>
<tr>
<td>3.</td>
<td>Pell-Eligible</td>
<td>Student qualification for federal Pell grant. Proxy for low-income.</td>
<td>N</td>
<td>Y (Pell received)</td>
</tr>
<tr>
<td>4.</td>
<td>High School GPA</td>
<td>High school grade point average.</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>5.</td>
<td>First in Family to Attend College</td>
<td>Student parental educational achievement does not exceed high school.</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>6.</td>
<td>Major-College</td>
<td>University college to which student major belongs.</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>7.</td>
<td>Enrollment Status</td>
<td>Student enrollment status (e.g., first-time freshman, continuing student, etc.).</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>8.</td>
<td>URM and Pell-Eligible (interaction)</td>
<td>Interaction between URM status and Pell-Eligible. Value is true if both variables are true.</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>9.</td>
<td>URM and Gender (interaction)</td>
<td>Interaction between URM status and gender. Value is true if URM status is true and gender is male.</td>
<td>N</td>
<td>N</td>
</tr>
</tbody>
</table>

Table 3 cross-references the variables to their inclusion in the research design of two previous studies that investigated the relationship between student characteristics, LMS use, and course grade. Because these two studies relied on previously collected data, some of the variables in the table were not included in the final analysis of these studies due to unavailable of
the data due to campus recording procedures. Nevertheless, their inclusion in the research design of these previous studies demonstrates their perceived importance by scholars.

Several of the variables listed in Table 3 were created through transformations of source data. The new variables provide a better distribution for statistical analysis. For example, in the case of race/ethnicity, descriptive data provided detailed distributions of student identification using the eight standard IPEDS categories, which were transformed into the dichotomous variable “URM student” for conceptual and data considerations. The main concept of interest is whether a student is from a race/ethnicity that is under-represented in higher education and is, therefore, less likely to persist through graduation. Transformation was further driven by the source race/ethnicity data having a low number of students in several of the existing categories (e.g., American Indian, African American). Aggregating race/ethnicity into a dichotomous variable allowed a sufficient number of students in each category to use the variable for statistical analysis.

Within the variable set, “first in family to attend college” indicates the educational background of the student’s father and mother. If both had a high school degree (or less) as their highest level of educational attainment, this variable was set to “yes”. If either parent attended college or progressed further in their education, this variable was set to “no”. “Pell-eligible” indicates whether a student is low-income or not. It was available as a variable in the university data set, but this value was calculated in consideration of multiple student characteristics, including household income, size of household, and dependent/independent status. It is a consolidated statistic that uses federal government indicators to determine economic need.

Using the college to which a student’s major belongs as a variable reduced the 188 majors offered at Chico State to eight college categories. Students without a declared major
were indicated with a unique major code. While the “major-college” variable indicates student motivation and intent upon entering college, the “enrollment status” variable indicates if a student is a first-time freshman, a continuing college student, a student who has left and returned, or if the student transferred from another college or university. “Enrollment status” provides the most robust characterization of the student’s current status in college from available data.

This study does not include several variables that describe student academic achievement prior to college enrollment, such as SAT scores. Although these are important variables, a preliminary analysis of the data set found enough missing values (36-38%) that they were prevented from being included in the study. Likewise, a low distribution of values prevented the inclusion of student age (81% of students were 18-20 years old). The availability of this data is related to student enrollment status (e.g., no current college GPA for first-time freshman, no high school testing information for transfer students). The remaining variables describe important factors identified in the literature to influence student achievement.

**Dictionary of key terms.**

Table 4 provides definitions of key terms used in this study. The data source for each term is provided where appropriate.

<table>
<thead>
<tr>
<th>#</th>
<th>Term</th>
<th>Definition</th>
<th>Examples</th>
<th>Data Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Persistence</td>
<td>Student achievement of academic goals in higher education.</td>
<td>Graduation, employment skills, other</td>
<td>N/A - conceptual term</td>
</tr>
<tr>
<td>2</td>
<td>Academic</td>
<td>Measure of how well a student mastered course material through course participation.</td>
<td>Final course grade</td>
<td>Peoplesoft database</td>
</tr>
<tr>
<td></td>
<td>achievement</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>LMS Use</td>
<td>Each action a student</td>
<td>Discussion message</td>
<td>LMS Activity Log</td>
</tr>
</tbody>
</table>
conducted in the LMS, recorded in a web server log file. Each LMS tool has multiple corresponding actions.

4. Categories of LMS Use

| Conceptual categories that correspond to the general intent of using the LMS. Each category contains several tools and their corresponding actions. |
| Administration, assessment, content, engagement |
| LMS Activity Log file (via transformation) |

5. Student Characteristics

| Variables in the student record that describe a student’s background. |
| Gender, high school GPA |
| ERS database |

**Research Questions**

The following questions guide this research study:

1. What is the relationship between a student's LMS use and academic achievement? How does this relationship vary by the pedagogical purpose underlying their LMS use? (correlation)

2. What is the relationship between a student's background characteristics and academic achievement? (correlation)

3. How does analyzing a student's background characteristics change the predictive relationship between their combined LMS use data and academic achievement? (multivariate regression)

3a. What is the relationship between a student's combined LMS use and academic achievement? (multivariate regression, restricted model 1)

3b. What is the relationship between a student's combined LMS use, combined student background characteristics, and academic achievement? (multivariate regression, restricted model 2)

3c. What is the relationship between a student's combined LMS use, combined student background characteristics, interactions between demographic variables (URM and Pell-
eligible, URM and gender), and academic achievement? (multivariate regression, complete model)

4. How does the relationship between a student's combined LMS use, combined student background characteristics, and academic achievement vary by students predicted to be at-risk for low achievement by their URM status and Pell grant eligibility? (multivariate regression by population sub-samples)

Significance

This study has potential significance for student persistence programs, academic technology departments, and institutional policy. Staff and administrators responsible for implementing technology on colleges may find this study especially relevant because it empirically tests an assumption that underlies many of their efforts: increased use of technology is related to improved student achievement. While the overarching goal of this research is to improve graduation rates in higher education, it is apparent that this is a complicated and challenging issue that one study alone cannot achieve. However, this study advances that goal by identifying how specific uses of the LMS contribute to student achievement. This study contributes to traditional student achievement research by comparing the relationship with achievement provided by student use of the LMS to the relationship with traditional student demographic and enrollment variables. Of utmost importance, the study identifies differences in achievement using the LMS by populations known to have lower rates of graduation. Because this study investigates a lower-division, general education course, the findings have potential applicability to a large proportion of enrolled students.

Institutions make large investments in the use of the LMS and other academic technologies. This study has relevance for institutional policy-makers by demonstrating the
benefits achieved by their direct investments and the faculty and student time spent using this
technology. Administrators are provided with a better understanding of the likely impact of
those investments and benchmarks that can be used as comparisons for future research.

Finally, this study supports data driven decision-making by academic technology
managers. Data from the LMS is readily available, whereas student demographic information is
difficult to obtain due to technical limitations and policy issues surrounding the privacy of
demographic data. If LMS use data is demonstrated to predict student achievement, managers
can begin to analyze this data to support their decisions and leadership activities to benefit the
highest goal of academic administration: improving student success.

**Context**

This study was conducted on one section of an “Introduction to Religious Studies” course
conducted at the California State University, Chico (CSU Chico), in the 2010 Fall Semester.
CSU Chico is a mid-sized campus (14,640 full-time enrolled students in 2010-2011) in the
California State University System (California State University Analytic Studies, 2011b). The
campus has a large service area that extends over 13 counties in the sparsely populated northwest
corner of California. To serve both this remote population and the ‘digital native’ undergraduate
students that live in Chico, the university has made significant investments in academic
technology infrastructure and faculty professional development. As part of this effort, the
campus began a program to redesign existing courses called “Academy e-Learning”. In this
yearlong program, a faculty team redesigns an existing course in order to improve student
achievement and reduce costs.

Introduction to Religious Studies was one of six courses redesigned in the 2009-2010
Academy e-Learning program. This high-demand undergraduate course fulfills a CSU Chico
general education requirement. The redesign aimed to increase the total number of students in each course section. Multiple activities were developed including recorded video lectures, online text-based learning materials, short graded assessments, and others. All of these activities were conducted within the LMS. Some course materials were made available to students only in this manner. The in-person meeting frequency was changed from twice weekly to once weekly with online activities replacing the other course meeting. The redesigned course was offered for the first time in Fall 2010 in a section with 373 students. Student usage of the LMS in this course was extremely high: it had 249,490 hits to the course website, with each hit representing an action within the course (e.g., reading an announcement, submitting an assignment). This was the highest gross usage (e.g., not averaged by student) of the LMS by any course offered at Chico State that semester.

In a program evaluation, the course instructors compared traditional sections to the redesigned course section to evaluate the effectiveness of the redesign (Vela, 2011). Two conventional sections of the course (with 146 and 99 students each) were also conducted in Fall 2010 by the same two-person faculty team. There was no indication of any difference between the sections in the course catalog, resulting in a random distribution of student enrollment between the courses. At the end of the semester, there was a 10% average increase in the final course exam scores of the redesigned course compared to the traditional course.

However, the redesigned course also had an increased number of students with drop, withdraw, or fail grade (7% and 11%, respectively). This is a surprising result because the increase in final exam grade would seem to lead to less low grades. One possible explanation is that most students performed better, but a smaller number of students had lower achievement in the redesigned course. Evaluations indicated that difficulty using the new technology was a
problem for some students (Vela, 2011). This dichotomous outcome presents an opportunity to investigate the relationship between LMS use, student characteristics, and academic achievement.

CSU Chico staff has run server queries on student and faculty usage to determine LMS adoption since the implementation of the WebCT Vista™ product in 2003. These queries collect detailed data on the frequency of course logins, the tools used within the course (e.g., discussion, mail, assignment), the actions conducted within those tools (e.g., posting a discussion forum, replying to a discussion forum), and the time spent on each action. The Registrar’s office maintains databases with records for each student, including demographic, financial, and academic background information. In 2011, the CSU Chico Data Warehouse integrated WebCT Vista LMS data for 2003-2010.

Campus administrators approved this project after considering the technical requirements and student privacy concerns. The researcher worked for CSU Chico Academic Technology Services during the initial phase of this study, which facilitated access to this data.

Limitations

The student characteristics in this study include common standardized measures recorded by campuses. However, it is possible that other characteristics not available in campus databases are important indicators of student academic achievement. These potential indicators include learning styles, student motivation, and many others (Zacharis, 2011). These alternative characteristics could provide increased accuracy in the predictive model and could be included in subsequent research.

This study investigates the frequency of student LMS use. However, the quality of student LMS use is not investigated. A thoughtful discussion posting is not distinguished from a
quick reply, nor is a thorough reading of online material distinguished from a quick review. This issue has been previously identified as a limitation in learner analytics research that relies on LMS log files (Buckingham Shum & Ferguson, 2011; Siemens & Long, 2011). Methods for future research are in development, including social learning analysis, LMS activity patterns, linguistic analysis, and others (Siemens et al., 2011). This study provides baseline data that future studies can be measured against to demonstrate advances in these methods.
CHAPTER 2
LITERATURE REVIEW

Area 1: Higher Education Student Persistence

Overview.

Research to understand why some college students successfully complete their degree and others fail to do so has been conducted for over 70 years (Braxton, 2000). In the 1960’s, the focus of this research changed from identifying the background characteristics of students likely to complete their degree before they entered the university to evaluating efforts to increase the achievement of students once they have been admitted (Pantages & Creedon, 1978). This research has been encapsulated under the broad categories of "retention" and "persistence". The current study uses the broader category of student persistence.

Tinto advanced this approach in a new theoretical model of student departure (1975, 1988, 1993). Initial formulations of this theory were critiqued for only applying to traditional students in four-year residential institutions (Metz, 2004; Milem & Berger, 1997). Students from under-represented racial/ethnic backgrounds and from low socio-economic status (SES) families persistently graduate at lower rates than students from majority racial / ethnic groups on campuses and higher-SES families. Although persistence research has focused on quantifying the differences between populations, some of the underlying factors leading to these differences have been identified. Research findings suggest that studies should investigate multiple variables simultaneously as well as the interactions between these variables to most accurately understand the persistence of students whose background makes them less likely to persist in their higher education goals.
**Distinguishing student persistence from institutional retention.**

Understanding why some students graduate, and other students do not graduate, is one of the largest questions in higher education research in terms of the importance and the complexity of the issue. Consequently, this is a prolific area in higher education research. Several literature reviews have been published on the topic (Pantages & Creedon, 1978; Pascarella & Terenzini, 1991, 2005; Peltier et al., 1999; Reason, 2003; Tinto, 2006). Two terms frequently used to define this body of literature are "persistence" and "retention". Although sometimes used interchangeably, there are important differences between these terms as clarified in the following passage by Reason (2009):

Retention is an organizational phenomenon — colleges and universities retain students. Institutional retention rates, the percentage of students in a specific cohort who are retained, are often presented as measures of institutional quality. Persistence, on the other hand, is an individual phenomenon — students persist to a goal. (p. 660)

By taking an organizational perspective, studies of retention limit measures of success to annual retention rates and graduation rates. Persistence takes a broader student perspective and considers outcomes such as students meeting personal goals without graduating, temporarily un-enrolling or "stopping out", or transfer to another institution. Persistence also tends to consider a broader set of student characteristics that affect a student's higher education decisions, such as family support, personal motivation and other (Allen, 1999; Asera, 1998; Mattern & Shaw, 2010).

Persistence is more directly related to student success in the broader sense. However, it can be difficult to acquire data about individuals, especially in a large-scale study, although this type of research has been performed (Adelman, 2006). Retention figures are more readily
accessible than measures of individual-level goal achievement and are often used as proxy measures for student persistence.

**Tinto’s interactionalist theory of persistence.**

In 1975, Tinto identified the principle elements of his new theory in relationship to two limitations in existing literature. First, the literature failed to make a distinction between voluntary withdrawal and academic dismissal. Instead, the literature assumed that students who did not complete a degree were unsuccessful academically. Tinto asserted that many students were not forced out through academic failure, but instead chose to not persist in their studies. Second, the literature did not focus on whether a major factor behind students’ decisions to not persist was a lack of social and cultural integration into their university.

To address these limitations, Tinto developed a model that included student characteristics, personal commitments, and the academic and social systems of the university. He later called this an “interactive model of student departure” (1993, p. 113). In Tinto’s model, the interaction between these elements determined whether a student would be more or less likely to persist in their career to graduation. Figure 3 depicts these interactions.
In the interactive model of student departure, Tinto asserted that students enter the university with a set of characteristics that increase or decrease their commitments to their professional goals and the institution. The strength of these commitments are either strengthened or weakened through the student’s experiences at the institution, which ultimately impacts the likelihood that the student persists to graduation or leaves the institution. The model explains why students from similar backgrounds have different rates of persistence in the same institution. However, the largest impact of the model was its emphasis on institutional practices to improve student persistence. It implied that universities could increase persistence by increasing student interaction with faculty and social communities, which would then lead to increased commitment between students and the institution.

In the initial formulation of his model, Tinto encouraged an increase in the on-campus social activities and informal interactions between faculty and students. However, other scholars noted that these practices were only available to traditional students at residential colleges and criticized Tinto’s model as not being applicable to community colleges, commuter institutions, and non-traditional students (Karp, Hughes, & O'Gara, 2010; Liu & Liu, 1999; Longwell-Grice & Longwell-Grice, 2007).

**Conditional influences of race, socio-economic status and gender on persistence.**

It has been well documented that students from certain family backgrounds are much less likely than other students to persist in meeting their higher education goals (J. D. Campbell, 2008; Morrison, 2010; Reason, 2003). Race, socio-economic status, and gender are fundamental constructs used in the social sciences to understand difference and inequality (Andersen &
Collins, 2006; Grusky, 1994). In their review of persistence studies, Pascarella and Terenzini (2005) concluded that these constructs had a "conditioning" effect on efforts to improve persistence, and they called for future research to "... explore whether the impact of any particular experience differs in magnitude for different kinds of students" (p. 153-154). Subsequent studies have focused on documenting whether these backgrounds have statistically significant relationships to persistence and specifying the magnitude of that relationship. There is less attention given to the reasons why this relationship occurs, although there are some factors specified in these studies.

**URM student persistence.**

Students from URM backgrounds consistently graduate at lower rates than students from white or Asian/Pacific Islander backgrounds (Carol, 2012; Gloria, Castellanos, Lopez, & Rosales, 2005; González & Ballysingh, 2012; Rendon, Jalomo, & Nora, 2000). URM student persistence is one of the most thoroughly researched areas in persistence literature on differences by demographic groups. Factors that have been found to influence the success of under-represented students include cultural assimilation, social connections and support, and racial climate.

Tinto's model for successful persistence is based in a process of student acculturation into a new college identity (1993). This new identity may appear race-neutral to white students, but for URM students this identity is racialized as a dominant white identity and represents assimilation into the dominant culture (Gloria et al., 2005). Many students experience a "culture shock" when they enter campus, and successful persistence requires navigating this new culture successfully (Rendon, 1994). Several scholars have raised this issue as a critical problem for under-represented students (Gloria & Castellanos, 2003; Gloria et al., 2005; Rendon et al., 2000).
There are few identities for URM students that allow for the maintenance of their prior racial/ethnic identity and their new college identity. This creates a tension between students' home values and that of the new community. These new values have been found to contribute to student persistence in higher education (Walpole, 2003); that is, if students do not assimilate these new values, they are less likely to succeed in courses and achieve their higher education goals. This process of assimilation leads to conflicts in the affiliations and identities of students from URM backgrounds (Hurtado, 1994).

Increasing the difficulty of this cultural adaptation is family social support, which is frequently lacking for URM students. Several studies have found this support to be a differentiating factor in persistence for students from URM racial/ethnic backgrounds (L. L. Brown & Robinson Kurpius, 1997; Gloria, Kurpius, Hamilton, & Willson, 1999; Hall, 1999; Palmer, Davis, & Maramba, 2011). Many families have not had personal experiences attending higher education institutions, leading to difficulty understanding the experience and needs of their children as they participate in a university education.

Campus racial climates have also been found to affect the persistence of URM students (Harper & Hurtado, 2007; Fegain, 1992). Racial climate is the overall sense of comfort that students perceive based on their racial background. URM students are less likely than white students to find the college environment generally supportive (Hurtado & Carter, 1997; Suarez-Balcazar, Orellana-Damacela, Portilla, Rowan, & Andrews-Guillen, 2003) and less likely to feel a sense of belonging (Johnson et al., 2007).

This climate sends a message to students of whether they are welcome in or excluded from the campus community (Reason, 2009). This message can be subtle or explicit; Hurtado (1994) found that 15% of Latino/a students were directly insulted or threatened based on their
ethnic background. Museus, Nichols, and Lambert (2008) found that the influences of this climate are often subtle, effecting goal commitment, social involvement, academic involvement, and institutional commitment, with strength and significance varying between ethnic groups.

One factor contributing to a campus racial climate is the presence or absence of peer mentors. Students have expressed desires for faculty and staff members from their own ethnic group (Gándara & Osugi, 1994). Peer mentors are believed to create stronger connections with faculty (Cammarota, 2007). Supportive relationships with faculty have been found to positively correlate with persistence for the entire student body (Pascarella & Terenzini, 2005), and this finding has also been confirmed with URM students (Gloria et al., 2005; Gloria et al., 1999; Schreiner, Noel, Anderson, & Cantwell, 2011).

The issues effecting URM student persistence lead to different empirical research findings on the variables related to student persistence. Allen (1999) found that high school rank, first-year college GPA, and self-reported desire to finish college accounted for 68% of the variation in retention of URM students from the first to the second year of college. For white or Asian/Pacific Islander students, high school rank, first-year college GPA, and parental education were significant predictors, yet only accounted for 38% of the variance in first-year retention. The different variables and magnitude of the prediction underscores the diverse experiences of students and the likelihood of a student succeeding limited by their background.

**Socio-economic status (SES) and persistence.**

Students from low SES backgrounds have been found to graduate at lower rates than high SES students in multiple studies, even after controlling for race/ethnicity and other demographic factors (ACT, 2004; Walpole, 2003). Scholars have made calls to increase research on the ways SES affects persistence (Berger, 2000; Milem & Berger, 1997). One approach has been to
consider SES beyond the immediate impact on financing and affordability of college by taking into account such factors as pre-college schooling, cultural capital, personal aspirations, type of institution attended, and family support (Barratt, 2011).

Social class has a strong influence on students before they enroll in the university. Students from lower-SES backgrounds usually attend K-12 schools that provide a lower level of preparation for a university education, and these students require a higher level of academic support to succeed academically in courses (Howard, 2000). This preparation also transcends content-level subject knowledge to encompass the broader cultural capital that has been found to influence the success of students in the university (Milem & Berger, 1997). Students from lower-SES backgrounds, like students from URM backgrounds, confront cultural differences in the university setting that they must adapt to in order to succeed and feel comfortable in this new environment.

Student aspirations for their higher education goals in terms of university choice, major, and terminal degree are also related to a student's SES status (Walpole, 2003). Choice of institution has one of the strongest impacts on persistence. Students from lower socio-economic classes are much more likely to attend for-profit universities than public or private non-profit schools, and students attending for-profit schools graduate at much lower rates than students attending other types of institutions (Bowen & Bok, 1998).

Parental support and aspirations for youth also affects students from low socio-economic classes. For these students, a high school diploma is the standard terminal degree, and students are considered successful if they are able to secure a full-time job after graduation. Students from high SES families are expected to attend a selective university and graduate into a promising career, if not attend graduate school (Hearn, 1984, 1990; McDonough, 1997, McDonough,
Korn, & Yamasaki, 1997). Lower parental expectations combined with less academic preparation and less time to dedicate to studies (due to financial needs and work commitments) make it very difficult for SES students to succeed in their higher education ambitions.

**Gender and persistence**

The relationship between gender and persistence is less clear than other social constructs. Women are disadvantaged in several ways in American society: they are paid less for performing the same work as men; they are stereotyped as less capable of performing math, science and other abstract reasoning; and they are discriminated against in other explicit and subtle ways. As gender is a primary individual identity marker, it is one of the initial demographic variables considered in persistence research (Pascarella & Terenzini, 2005; Reason, 2003; Tinto, 2006).

Women have made significant advances in higher education, and as indicated previously, make up the highest proportion of enrollment. In several studies gender has been found to have a significant relationship with persistence, with women more likely to graduate than men (Astin 1975, Astin Korn & Green 1987, Peltier, Laden & Matranga, 1999). However, gender differences have been initially present in the findings of other studies, but have disappeared when other factors are controlled for — especially race and SES (Reason, 2003; St. John, Hu, Simmons & Musoba, 2001).

Reason (2009) found that gender was significant in a simple statistical model, but significance disappeared in each of two step-wise regression analyses performed with the same data set. A multi-institutional study by St. John et al. (2001) found that gender was not significant when the model only included basic demographic variables (e.g., gender, age, race, financial dependence, family income, and SAT/merit-index). However, gender was significant
in a second model, which added first-semester college GPA. When variables related to the
institution (e.g., type of institution, degree program, and housing) were added to the revised
model, gender again did not achieve significance. St. John et al. concluded that the interaction
between variables created the significance of gender and that gender alone was not a significant
predictor of persistence.

Gender and choice of major, especially in science, technology, engineering and
mathematics (STEM) fields, has been investigated (Ohland et al., 2011; Shapiro & Sax, 2011).
Studies have found that women are discouraged in high school to take the needed math and
science courses to prepare them for these degrees. In addition, they have found that factors in
the "culture and pedagogy" (Shapiro & Sax, p.8) of STEM courses deter women from selecting
these majors and persisting to graduation. These factors include competitive relationships
between students and the impersonal large enrollment lectures that characterize lower-division
courses.

The interaction between gender and race has been significant in several studies
(Murtaugh, Burns & Schuster, 1999; Muñoz & Maldonado, 2012; Ohland et al., 2011). Much of
the increase in overall degree attainment rates over the past 20 years can be attributed to the
increased persistence of white women and women of color (National Center for Education
Statistics, 2011). However, this increase masks certain inequities, such as the decrease in
graduation rates of male URM students (King, 2006). Gender is a complex identity, and the
research demonstrates that the most meaningful findings are complex social issues that are best
understood by combining gender with other variables.
Contemporary research methodology: multivariate analyses and variable interactions.

As described in the previous sections, complex factors are involved in the lower persistence of students from under-represented racial/ethnic groups and lower socio-economic classes. Gender is a factor as well. Pascarella and Terenzi (1998) called for studies that examine multiple variables simultaneously. In addition, they suggested that studies look at the interactions between variables in order to increase our understanding of persistence (2005).

In his review of persistence literature, Reason (2009) found that because isolated demographic factors have large within-group variance, new methods have emerged that combine these factors. In a previous study, Reason (2003) called for changes to methodology: "[w]hile a study of retention should include race as a variable, the statistical analysis must be sophisticated enough to examine the interaction of race with other variables" (p.183). Using interactions has led to more accurate findings in recent research (Seidman, 2005).

Classroom instruction and persistence.

Braxton (2004) challenged the empirical support for Tinto’s model by deconstructing the model into 13 researchable propositions; Braxton defines persistence as an “ill-structured problem” (p. 2) that requires multiple theoretical approaches to address the complexity of the factors that influence persistence. He found that existing empirical studies only supported five of the propositions and that most of these studies were conducted at residential institutions. However, Braxton endorsed previous recommendations about the importance of instructional practices. He specifically identified “active learning”, which includes “… cooperative learning, debates, role playing, discussion, and pair and group work" (Braxton 2004, p.48) as beneficial educational activities. While challenging the all-encompassing nature of Tinto’s theory, Braxton asserted that research supports the use of classroom practices to increase student persistence.
In a multi-institutional study conducted in 2011, Pascarella, Salisbury and Blaich found that improved classroom instruction did lead to improved student retention. They also found a “significant role of learnable faculty instructional behaviors in student persistence” (p.6). Other studies have confirmed the role of faculty interaction and classroom instructional practices on improving student retention (Demaris & Kritsonis, 2008; Endo & Harpel, 1982).

**High enrollment courses and academic technology.**

One problem leading to decreased academic achievement in large enrollment courses is a lack of interaction between students and faculty (Gainen & Willemsen, 1995; Harrison, 1989; Seymour & Hewitt, 1997; Tobias, 1990). While many of the large enrollment courses studied for academic achievement cover science, technology, engineering and mathematics-related topics, this issue is not unique to these fields. In order to improve persistence while maintaining enrollment levels, pedagogical changes have been proposed for these courses, including the use of academic technologies (Wang, Fong, & Kwan, 2010). Multiple studies have found that increases in student outcomes are achieved by including technology elements into courses, especially those that provide increased interaction between students and faculty or provide immediate feedback (Arasasingham et al., 2011; Deslauriers et al., 2011; Terman, 1978; Wallin, 1997; Woelk, 2008). A limitation of these studies is that enrollment in the course is used as the independent variable and students are infrequently disaggregated by their background characteristics or actions within the course. When these variables are used, it is unclear if improved achievement is related to the students’ actual use of the technology application and which students benefit from the technology.
Course grades as indicators of persistence.

If graduation or withdrawal from the university was the only outcome variable used in persistence studies, only multi-year studies could be conducted. Although grades do not directly measure persistence, scholars have found that they are reliable indicators of the phenomenon. Tinto argued that grades are the “most significant factor in predicting persistence in college” (1975, p. 109). In fact, grades have been found to be the most significant predictor of persistence outcomes in multi-institutional studies and single institutional studies (Cabrera, Nora, & Castaeda, 1993; D. L. Smith, 1992), and course grades have been used more than any other measure as an indicator of academic success (Pascarella & Terenzini, 2005).

Area 2: The No Significant Difference Debate and Internet Technology Research

Overview.

The No Significant Difference debate between Clark and Kozma is a central issue in the history of educational technology research (Clark, 1983, 1994; Jonassen, Campbell, & Davidson, 1994; Kozma, 1991, 1994a, 1994b). The debate centers on whether technology functionality should be considered as an independent variable in educational research, or if only non-technical factors should be considered. Although Clark and Kozma differ on several issues in this debate, they concur that technology features alone do not result in changes to learning outcomes. Contemporary research has validated this assertion with new technologies (Means, 2010; Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011; U.S. Department of Education Office of Planning, 2009).
In a 1984 article credited with starting this debate, Clark called for an end to what was then known as comparative media studies. (At the time, video was the technology most frequently studied and, therefore, studies of technology were called “media studies”.) In these investigations, researchers studied a new media as the only independent variable and student achievement as the dependent variable. When reviewing these studies, Clark consistently found no statistically significant difference. He determined that these studies made an error in considering media as the independent variable. In a quote that summarizes his perspective on the role of media, he stated “[t]he best current evidence is that media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition.” (1984, p.445) Clark called for researchers to “refrain from producing additional studies exploring the relationship between media and learning unless a novel theory is suggested.” (p. 457).

Kozma answered Clark’s call with a new theoretical framework (Kozma, 1991, 1994a, 1994b). Kozma asserted that media should not be studied as an independent variable, but instead, media had characteristics that interacted with other processes of learning and knowledge creation:

Specifically, to understand the role of media in learning we must ground a theory of media in the cognitive and social processes by which knowledge is constructed, we must define media in ways that are compatible and complementary with these processes, we must conduct research on the mechanisms by which characteristics of media might interact with and influence these processes, and we must design our interventions in ways that embed media in these processes. (Kozma, 1994b, p. 4)
In Kozma’s theory, existing media studies were fallacious because they asserted that media could be studied as a variable isolated from the cognitive and social processes that were involved in learning. The interaction between these processes and media could lead to studies that incorporated the role of media in learning.

Clark opposed this new theory because it asserted that the characteristics of media were relevant to consider in educational research, which repeated a conceptual error in previous studies (1994). This exchange opened a debate between Clark and Kozma about media studies and the role of technology in educational research (Clark, 1994; Jonassen et al., 1994; Kozma, 1994a, 1994b). At the core of this debate was a disagreement about the epistemological status of technology as a necessary condition for learning versus a sufficient condition for learning. For Clark, technology must be a necessary, non-replaceable element in learning in order to merit research. Since a technology could be exchanged with another technology that provided similar functionality, it did not merit investigation (Clark, 1994).

Although this critique could be brought against any specific learning approach or technique, in other areas specific techniques were related to deeper concepts. Media studies research did not investigate the deeper conceptual value conveyed through technology, but suggested that video or other technologies had intrinsic educational value. As such, technology must be a necessary condition in order to be considered as a variable in research.

For Kozma, technology was a sufficient condition that made a substantial contribution to learning (Kozma, 1994a, p. 25). Although technology solutions could be exchanged for others, Kozma asserted that they contributed to learning by virtue of supporting certain learning methods or student characteristics. In his approach empirical research on specific technologies
was warranted in order to understand the relationship between methodologies and specific technologies.

Despite these differences, Clark and Kozma agreed on several issues (Shrock, 1994). For both scholars, existing media studies were invalid because they studied technology as a single independent variable. They advocated that learning methodology, the task being attempted, and learner characteristics should be the first items considered in educational research. In addition, Clark believed that media could provide economic value: “media can influence the cost or speed of learning” (Clark, 1994, p. 26). If media could reduce the cost for creating or reproducing learning, then it could provide a financial benefit. Combined with appropriate methodologies, Clark believed that distance learning technologies could be effective in increasing the geographical dissemination of education (2006).

Kozma asserted that it was important for educational researchers to become involved in the design of technology-mediated learning environments: "In what ways can we use the capabilities of media to influence learning for particular students, tasks, and situations?" (Kozma, 1994a, p. 26). The importance of this issue was increasing as technology came to play a larger role in the broader society and economy. Kozma’s question suggested a new approach to media studies research.

**Contemporary research on Internet technology in education.**

Russell’s *The No Significant Difference Phenomenon* (2001) provides an annotated bibliography of 355 comparative media studies conducted between 1928-1998. Despite focusing on contemporary technologies, these studies found no significant difference based on the usage of technology alone. A recent widely-cited meta-analysis sponsored by the U.S. Department of Education examined over 1,000 studies in distance learning to evaluate the difference in learning
outcomes between students learning using online technologies, blended online/in-person learning, and traditional in-person instruction (2009). This study also found no significant difference between online and in-person instruction, although it did find a small positive effect between blended compared to in-person instruction. However, this result was not related to the use of distance learning technology, as noted in the report:

“[a]nalysts noted that these blended conditions often included additional learning time and instructional elements not received by students in control conditions. This finding suggests that the positive effects associated with blended learning should not be attributed to the media, per se” (ix).

The positive results from distance learning technology came from providing supplemental instruction, not from inherent characteristics of technology features, as suggested by Clark and Kozma. Other meta-analyses of distance learning studies have come to similar conclusions (Bernard et al., 2004).

Russell (2001) agreed that while technology studied in isolation does not effect learning, there are potential gains to be found by changing instruction to integrate technology: “in going through the process of redesigning a course to adapt the content to the technology, [learning outcomes] can be improved.” (Russell, 2001, p. xiii). This approach reinforces Kozma’s call to consider the interaction between technology and other aspects of learning.

One of the most systematic attempts to redesign courses by integrating academic technology is Twigg’s Program in Course Redesign (2003a). Supported by an $8.8 million grant from the Pew Foundation, the Program in Course Redesign provided 30 grants of $200,000 each to colleges and universities around the country to redesign large enrollment introductory courses. By integrating technology, primarily the LMS, these redesign efforts sought to improve learning
outcomes and reduce instructional costs. The Program in Course Redesign required that entire courses were redesigned, but it did not prescribe the technology used, the instructional methods integrated, or the specific learning outcomes. As courses in various disciplines were redesigned, staffing changes (namely a reduction in faculty and other professional time funded) were made.

Formal program assessments were conducted on each course, and significant increases in learning outcomes in 22 of the 30 courses were found (Twigg, 2003b, 2003c, 2003d). Thus, the Program in Course Redesign provided an example of how course redesign could result in improved student success, and the program continues as a fee-based consulting service (The National Center for Academic Transformation, 2005). Without a control group, limited conclusions can be drawn about the role of technology in the redesigned course outcomes. It is possible that similar gains could have been produced through a redesign effort that did not use technology.

**URM student achievement in technology-enhanced courses.**

Scholars have found that significant differences in learning outcomes using technology are found when student characteristics are considered, especially in the case of URM student populations (Lamkins, 2004; Vanderpool, 2009). These students have lower academic achievement in online courses than well-represented students taking online courses, especially in the case of introductory courses (Bradford & Wyatt, 2010; Chen, 2006; Lamkins, 2004; Welsh, 2007). However, this achievement has been determined to be not due to inherent attributes of the technology, but rather to study skills and background knowledge required for successful achievement in online courses (Aljarrah, 2000; Done, 2009; Woodke, 2006; Z. Zhang, 2007). The achievement gap between URM students and majority students is higher in online courses
than it is in face-to-face courses because the independent student experience in online courses requires more academic preparation for successful participation.

In the Program in Course Redesign, the achievement gap between URM students and well-represented students persisted in redesigned courses (Twigg, 2005). While diverse in subject matter, 25 out of the 30 redesigned courses used a hybrid course format, which avoided some of the academic issues caused by the academic requirements of fully online courses. The Program in Course Redesign did provide some benefits for these students: the anonymity of online discussion forums, for example, made students feel more comfortable speaking up and asking questions. It was found that URM students who participated in discussion forums and logged in frequently had no disparity in learning outcomes.

In much of the research in this area, the average change in student achievement that results from a change in course design is all that is known. Whether students tried to use the technology and were unsuccessful, or whether they did not use available technology and had lower achievement cannot be determined from the research. Determining which of these situations is taking place may lead to very different conclusions and implications for practice.

If students are not using the technology, then improved technology skills and self-confidence could improve student achievement. If the technology is being used, but student achievement is lower than expected, then the technology activities need to be reviewed and potentially redesigned. These results have led one scholar to ask if the usage of information technology is a “missing link” in evaluating the impact of information technology (Devaraj & Kohli, 2003).
Area 3: Learner Analytics

Overview.

Learner analytics is a new area in educational research that is growing rapidly. Building upon the previous research field called academic analytics, learner analytics applies “business intelligence”, quantitative analysis of multiple data sources, to course data in order to improve retention and persistence (Goldstein & Katz, 2005). Only four empirical studies have been published to date that compare LMS use to academic achievement, although there are frequent mentions of the importance of this type of analysis (M. Brown, 2011; Taylor & Francis, 2010; WICHE Cooperative for Educational Technology, 2011). Existing studies have three major limitations that the current study seeks to address: (a) they do not consider how LMS use is related to deeper learning topics (e.g., pedagogical function, course content, etc.); (b) they eliminate large amounts of useful data from the LMS log file due to methodological concerns; and (c) they do not consider student characteristics.

Academic analytics: analyzing campus data for institutional improvement.

A foundation for learner analytics is a 2005 EDUCAUSE research report by Goldstein and Katz that surveyed 350 higher education institutions. In this report, the authors coined the term “academic analytics” for the application of business intelligence functions (e.g., predictive modeling, early alerts, dashboards, and others) to higher education. Goldstein & Katz associate academic analytics with the adoption of Enterprise Resource Planning (ERP) systems on college and university campuses. ERP systems centrally maintain information about student characteristics, course enrollment, grades and graduation information. While institutional research has been conducted on this data for many years, the ERP systems enabled potential
access to this data to the entire institution and an increase in the dissemination of reports beyond traditional sources and more complex reporting.

Empirical studies conducted by researchers have demonstrated relationships between student characteristics found in the data collected by ERP systems and student persistence (J. D. Campbell, 2008; Morrison, 2010; Reason, 2003). These studies have at times led to surprising results. In a study of student retention at the University of Arizona using 12 years of ERP data, J. D. Campbell sought to determine which student background variables (e.g., high school GPA, race/ethnicity, achievement scores) had the strongest relationship with persistence. He found that the variables predicted by theoretical models to have the highest relationship with persistence were different from those found in his analysis of the available data. For example, although high school GPA, SAT score, and gender were predicted to have strong relationships with persistence, other variables, such as state residency status, age, and family financial contribution that were not predicted to have this relationship did have a significant relationship.

As a result of these findings, J.D. Campbell recommends that future studies use “knowledge discovery” (p.127), which is otherwise known as “data mining”. Using this technique, researchers analyze a broader spectrum of data than what is typically included in hypothesis-driven research. Further, variables are seldom excluded before analysis is conducted. Knowledge is seen to emanate from the data, rather than constraining data through theoretical models.

Several studies have researched the relationship between student demographic data and student pass rates in online courses (Brallier et al., 2007; Bukralia, 2009). A persistent finding is that single demographic variables (e.g., race/ethnicity, economic status) do not have a significant relationship with student achievement. By contrast, multivariate analysis has found that
combinations of these variables have a significant relationship with student achievement.
Further, this research has found that variables more closely related to current schooling (e.g.,
degree seeking status, financial aid status, current grade point average) have the strongest
predictive value. These results suggest that variables describing the current conditions of the
student are more important than the student’s historical background. Although these findings are
not surprising from a conceptual perspective, the potential to provide these findings using
automated technologies that disseminate the results to a larger audience is a major breakthrough
in applying research to practice (J. P. Campbell, 2007; Hrabowski III, Suess, & Fritz, 2011;
Norris et al., 2008).

Beginning with Goldstein & Katz’s initial report, the LMS has been identified as a
potential future area for analytics: “As [learning] management systems attain the status and
stature of enterprise systems, they too will acquire and store volumes of student data that, when
combined with other information, can begin to help us more fully understand the student
experience around learning” (p.9). This call to extend academic analytics has more recently been
made by several other authors (M. Brown, 2011; Parry, 2011; Siemens & Long, 2011). The
inclusion of this data brings changes to the intended audience and scope of research, which has
led to the definition of a new field of study being called “learner analytics” (M. Brown, 2011).

Learner analytics: bringing analysis to learning contexts.

The first-known use of the term “learner analytics” was made in reference to corporate
training environments in 2004 (J. Berk). One of the most comprehensive definitions of the term
was made by the conference committee for the first Learner and Knowledge Analytics
conference in 2010: “Learning Analytics is 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.” (Siemens, 2011). Although this field has recently emerged, scholars have analyzed logs generated by Internet-mediated learning environments since the late 1990’s (Peled & Rashty, 1999; Rafaeli & Ravid, 1997). In two studies at the Hebrew University of Jerusalem, the authors analyzed an Internet-based application similar to an LMS that was used to post materials, facilitate online discussions, and conduct quizzes. In *Logging for Success*, Rafaeli and Ravid used student LMS activity as their independent variable and student grades as their dependent variable (1997). They found significant correlations between individual variables and student grades. A multivariate regression of significantly correlated variables yielded an $R^2$ value of .23, which is interpreted as a small effect size according to Cohen (1988). This study provided an early demonstration of the possibility of research using LMS use data, and the investigators recommended additional studies.

Recent professional and academic literature has recommended the use of learner analytics to improve student persistence (Brallier et al., 2007; Dawson, McWilliam, & Tan, 2008; Fritz, 2011; Jones, 2009; Prineas & Cini, 2011). Suggestions have been made that learner analytics holds large potential for improving persistence rates, as illustrated by Grajek (2011): “Analytics, and the data and research that fuel it, offer the potential to identify broken models and promising practices, to explain them, and to propagate those practices.” (p. 48). Enthusiasm for the potential of learner analytics suggests that there are many additional studies validating and extending the findings of Rafael & Ravid’s research. However, there has been a surprisingly small amount of research conducted since this initial study. Only three empirical studies after Rafael & Ravid have examined the relationship between LMS use and student grades. Table 5 describes the research contexts and findings in these studies.
Table 5

_Empirical Study Precedents in Learner Analytics: Contexts and Findings_

<table>
<thead>
<tr>
<th>Study</th>
<th>N (Sections, Students)</th>
<th>Course (Format, Subject(s))</th>
<th>Significant Correlated Variables</th>
<th>Significant Regression Variables</th>
<th>Regression R² Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rafaeli &amp; Ravid (1997)</td>
<td>3 sections, 179 students</td>
<td>Hybrid, Introduction to Business Information Systems</td>
<td>5</td>
<td>5</td>
<td>.24</td>
</tr>
<tr>
<td>Morris et. all (2005)</td>
<td>13 sections, 354 students</td>
<td>Online, English Comp II, U.S. History, Intro to Geology</td>
<td>8</td>
<td>3</td>
<td>.31</td>
</tr>
<tr>
<td>Macfadyen, &amp; Dawson (2010)</td>
<td>5 sections, 117 students</td>
<td>Online, Introduction to Biology</td>
<td>13</td>
<td>3</td>
<td>.33</td>
</tr>
<tr>
<td>Campbell J. P. (2007)</td>
<td>597 sections, 26,014 students</td>
<td>All sections offered for one semester; hybrid and online</td>
<td>10*</td>
<td>30**</td>
<td>.51</td>
</tr>
</tbody>
</table>

* LMS use variables only
** Student Characteristics and LMS use variables

Each of these studies follows a similar research method. First, variables for inclusion in the analysis are identified at the discretion of the researcher. These variables include the tools and actions that the researcher believes students use frequently. Variables that are believed to duplicate similar activities are eliminated; for example, discussion postings read and discussion postings complied. This process takes place before any data is analyzed, which could cause the mistake of excluding variables that have significant relationships with student achievement.

Next, each selected variable is correlated with the student course grade. The significant variables are combined into a multivariate regression equation, and an $R^2$ value is calculated that indicates the overall relationship between included variables and student final grade.
The findings from the first three studies listed in Table 5 are similar, with $R^2$ values indicating a small to medium effect size (Cohen, 1988). However, the learning environments of these courses were different in several respects. Rafaeli & Ravid’s study (1997) was a hybrid course, with online components supplementing face-to-face instruction. This course was also the first exposure of most students to LMS-like tools, an issue that may have created an obstacle to students effectively using the tools. Further, this study included three course sections with different student populations (undergraduates, MBA and executive MBA). All of these factors may have contributed to a smaller effect size than studies conducted in the present day or with a single course population.

The studies by Morris et al. and Macfadyen & Dawson had small effect sizes. Both of these studies were conducted with fully online general education courses. Morris’s study, although conducted with three different courses and with some different variables, had similar findings to Macfadyen & Dawson’s study, which analyzed a single course. Although a detailed description of the course activities and use of the LMS were not provided, it is likely that these courses had similar usage of the LMS. They also covered similar subject matter and had similar student populations. These factors may help to explain the comparable findings between these two studies.

A major difference in research context and design is found in J.P Campbell’s 2005 dissertation research study. This study was conducted across all course sections offered ($N = 597$) during a semester at Purdue University. Campbell included student characteristics related to Tinto’s persistence model in his analysis. Campbell’s findings had a higher number of significantly correlated variables (although he also included more variables in the design), with a .2% increase in the $R^2$ value compared to other studies, which is a medium effect size.
Campbell eliminated the courses with low levels of activity by excluding courses with less than two minutes of activity per week. Given the large number of observations in the study, Campbell was able to split his population into several samples and run different analysis techniques. LMS data overall had a small relationship with student achievement. Correlations found most individual LMS use variables had a correlation below \( r = .1 \). Variables were aggregated into a single measure of use to provide a more comprehensive measure of use. He found that four predictive variables were constant across multiple analysis methods: standardized test scores, current college grade point average, a total LMS use composite, and the course level (e.g., lower division, upper division). Multivariate regression was used to predict course grade, and logistic regression was used to predict students that needed help (defined as final grade below a C). The strongest LMS predictions were found with students in the lowest quartile of academic achievement, which is an opportune result as these are the students most in need of support.

One of Campbell’s research questions assessed the change in the predictive model created by adding LMS activity variables to a model created with student characteristics. He found an extremely small change: the \( R^2 \) value only increased by .02% (from an R2 value of .51) by adding the LMS data. Campbell suggested that this might be a result of the research design. LMS use varied between courses, which caused a large number of the LMS variables to have missing data. Furthermore, it was necessary to eliminate these records from the statistical analysis. Campbell suggested that future research be conducted with individual courses to address this issue. While Campbell has conducted additional research at Purdue University on prediction of achievement for individual courses, those results have not been published.
Campbell’s study findings suggest that combining student characteristics and LMS behavior is a promising approach to increase the predictive power of the relationship between these variables and student achievement. There is a serious lack of studies using this approach, creating a gap in current education technology research. This gap is important to address because Campbell’s research had a much higher $R^2$ value than other studies, and student characteristics provided most of the predictive value.

An additional limitation in previous research is the overall purpose for the studies. Past research has been conducted largely to demonstrate that variables in LMS activity logs have a statistically significant relationship to student achievement. The research does not seek to advance knowledge about how the LMS activity was related to the pedagogical function behind the online activities, the course content, or student characteristics (boyd & Crawford, 2011; Buckingham Shum & Ferguson, 2011; Siemens & Long, 2011). Morris (2005) poignantly frames this point:

In conclusion, this study would appear to be documenting the obvious — students who are more engaged with the content and discussions in an online course will persist and complete successfully. However, such studies provide a necessary basis for understanding the complex interactions between students, faculty, course materials and course structures. (p.229)

These findings in part have led several scholars to develop software applications that provide faculty access to LMS use logs for their courses (Mazza, 2007; Retalis, Petropoulou, & Lazakidou, 2011; H. Zhang, Almeroth, Bulger, & Mayer, 2010). Using these applications, faculty can identify students who have low levels of LMS use relative to other students in the course. Faculty can then contact students and make appropriate interventions. Although this is
an important application of LMS use reporting to improve student achievement, it is limited by a
dependence on individual faculty and student responses. Moreover, by focusing on individual
learners, these applications do not help us to evaluate if a course activity or overall design can be
improved for all learners in a course, which has a much larger potential impact to improve
student achievement.

The field of learning analytics is growing rapidly. The Journal of the Learning Sciences,
a prominent educational research journal, has issued a call for papers for a special issue on
learning analytics to be published in 2012 (Taylor & Francis, 2010). An annual academic
conference on Learning and Knowledge Analytics was started in 2011 to disseminate research
and advance knowledge in the field (Athabasca University, 2011). This study seeks to contribute
to this emerging body of research.
CHAPTER 3
RESEARCH METHODOLOGY

Overview

This section presents the methods that were used to perform the research in this study, including the rationale for selecting the Introduction to Religious Studies course as the object of research, the LMS and student characteristic data sources that were used, and the dependent and independent variables within those sources that were selected for analysis. The filtering and other methods used to transform raw data into a format appropriate for research are described, and the procedures used to ensure that the data was valid for research in consideration of missing data and data distributions are defined. The correlations and multivariate regressions conducted to test the research questions are also described in this chapter.

Research Design

This study used an observational research design, which is a study without experimental control over the assignment of participants. This design is well-suited for applied research studies in which random assignments of participants is not feasible or would be unethical (Agresti & Finlay, 2009). As in most studies with college students, enrollment in this course was determined by student choice and was not under the control of the researcher. A common issue encountered in social science research is concern about the generalizability of observational research design results beyond the immediate population (R. A. Berk & Freedman, 2010). The present study addresses these concerns, in part, by situating this study within previous research.
Course selection.

This study investigated the complete population \((n = 73)\) of a course section of “Religious Studies 180: Introduction to Comparative Religion” at CSU Chico that was taught in the Fall Semester of 2010. This course was chosen from a pool of thirteen courses that completed the “Academy eLearning” program at CSU Chico during 2009. Academy eLearning is a year-long professional development program by which teams of faculty work with Academic Technology Services department staff to redesign their courses to integrate instructional technology for improved student success. Faculty participate in a two-week workshop, then redesign their course in consultation with staff over an entire year. Courses that are redesigned in this program are likely to have a deep integration of technology into learning activities. These courses provide best-case examples and may not reflect average LMS use at the university.

Religious Studies 180 was redesigned from a conventional in-person course to a hybrid course format with one in-person weekly meeting and one meeting replaced with online activities. The learning format of the course was not denoted in the course schedule. Therefore, while this study used a convenience sample, the students enrolled in this course should not differ from the undergraduate population at CSU Chico taking introductory General Education courses.

Religious Studies 180 was selected because it had the largest student enrollment of any of the redesigned courses. This large enrollment was helpful for several reasons. First, with a large sample, the statistical power of the study findings would increase. This made it possible to conduct analysis with a larger number of student characteristic variables than would have been feasible with a smaller sample. The large enrollment was also ideal to generate LMS activity. Subsequent analysis of the LMS log for this course found that there were 249,490 LMS website hits in the course, the largest amount of LMS use of any course offered at Chico State during the 2010 Fall Semester.
Protection of human subjects.

The privacy and confidentiality of students was secured through adherence to UC Davis and Chico State Institutional Review Board (IRB) policies. IRB applications were submitted and approved at both institutions. The analysis of previously collected data was conducted for this study. Anonymized numbers were used to code individual students, and no other personal identifiable information was present in the data sets.

Data Sources

This study examined a data set that was extracted from the CSU Chico Data Warehouse and other campus databases that contained student LMS use, student characteristics, and final course grade information. The data sources were validated by campus staff familiar with each database to ensure that the data was accurately uploaded into the Data Warehouse. The data was extracted and then analyzed in the Stata statistical software application.

The data set joined tables from three data sources: WebCT™ Vista LMS data (WebCT), the Enrollment Reporting System (ERS), and the Peoplesoft Student Information System (Peoplesoft). These databases and their content are listed in table 6.

Table 6

Source Databases used for Data Set

<table>
<thead>
<tr>
<th>Database</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>WebCT Vista™ LMS Use (WebCT™)</td>
<td>LMS use from server log file</td>
</tr>
<tr>
<td>Enrollment Reporting System (ERS)</td>
<td>Student demographic information and academic history in standard CSU format</td>
</tr>
<tr>
<td>Peoplesoft Enterprise Resource Planning System (Peoplesoft)</td>
<td>Student demographic information and academic achievement in CSU Chico format</td>
</tr>
</tbody>
</table>
LMS use: WebCT™ web server log.

This database was extracted from the WebCT™ LMS web server log. WebCT™ creates a record for each student action within the LMS course, such as posting a discussion message or beginning a quiz. Each record includes fields for the student ID, tool used, activity conducted within that tool, length of activity, time of day, and other pertinent data.

Student characteristics: ERS and Peoplesoft.

Two sources of student demographic and achievement information were used in this study: ERS and Peoplesoft. The ERS database uses data definitions defined by the CSU Chancellor’s Office for submission of student enrollment information. Each campus in the CSU system uses the ERS database format to submit information about the backgrounds of students enrolled in the campus. In order to support future research by other CSU campuses, the ERS database was used as the primary source for each student characteristic variable. ERS did not provide Pell-grant eligibility and student final course grade. These variables were extracted from the Chico State Peoplesoft student information system.

Reduce Data

LMS log file records.

Variables to include in the analysis were selected from the LMS log file, which required extensive filtering. Records for users with a status of “enrolled student” were selected for inclusion in the data, thereby removing faculty usage of the LMS from analysis. Every technical request made by the WebCT™ program in response to a student action is recorded in the log file. In its original form, the log file provides an inaccurate representation of student LMS use. For example, if a student selects “read compiled discussion message”, this consolidates multiple
discussion messages into a single message, which results in a log file hit for each message in the message thread. This single LMS action by the student could result in 20 log file hits.

To control for this inaccuracy in the source data, filters were applied to the LMS data by two criteria: (a) dwell time and (b) actions performed. Log entries with a dwell time of less than five seconds were excluded from analysis to control for automated server-level events such as compiling a discussion message. At the opposite end of the scale, some actions could result in an inflated dwell time, such as students beginning an action and leaving their computer: to control for this issue, activities with greater than one hour of dwell time were excluded from analysis.

Some LMS activity variables also result in multiple log file entries for a single student LMS action. For example, in order to read an announcement, a student must also view the announcement forum. To control for these duplicate entries, the actions in the log file were reviewed and log file entries with the following actions were excluded from the data set: “topic-viewed”, "organizer-viewed", "discussion-home-viewed", "assignments-listed", "web-links-home-viewed", "folder-selected", "compile-lm", "assessment-view-list", "announcement-view-list”.

Several LMS use tools had less than one mean hit per student and, therefore, were not considered a meaningful part of the course. These tools were dropped from the data set. The tools dropped for this reason were: “student bookmarks”, “who is online”, “notes”, “chat”, “media-library”, “learning-objectives”, and “tracking”.

All remaining LMS log file entries were included in the data set. Table 7 contains a complete list of the included LMS tools and actions.
Table 7

*LMS Use Tools and Activity Variables Included in Research*

<table>
<thead>
<tr>
<th>#</th>
<th>LMS Use Category</th>
<th>LMS Tool Variable</th>
<th>LMS Activity Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Administration</td>
<td>Announcement</td>
<td>View-Announcement</td>
<td>The total number of times a student opens an announcement.</td>
<td></td>
</tr>
<tr>
<td>2. Administration</td>
<td>Announcement</td>
<td>View-Announcement-List</td>
<td>The total number of times a student opens the list of all announcements.</td>
<td></td>
</tr>
<tr>
<td>3. Administration</td>
<td>Calendar</td>
<td>Access</td>
<td>The total number of times a student accesses the calendar tool.</td>
<td></td>
</tr>
<tr>
<td>4. Administration</td>
<td>File-Manager</td>
<td>File-Uploaded</td>
<td>The total number of times a student uploaded a file to the course for use in any activity.</td>
<td></td>
</tr>
<tr>
<td>5. Assessment</td>
<td>Assignment</td>
<td>Assignment-Read</td>
<td>The total number of times a student reads (opens) an assessment activity.</td>
<td></td>
</tr>
<tr>
<td>6. Assessment</td>
<td>Assignment</td>
<td>Assignment-Listed</td>
<td>The total number of times a student lists the assignments for a course.</td>
<td></td>
</tr>
<tr>
<td>7. Assessment</td>
<td>Assignment</td>
<td>Assignment-Submitted</td>
<td>The total number of times a student submit an assignment item for a course.</td>
<td></td>
</tr>
<tr>
<td>8. Assessment</td>
<td>Assessment</td>
<td>Assessment-Started</td>
<td>The total number of times a student started an assessment for the course.</td>
<td></td>
</tr>
<tr>
<td>9. Assessment</td>
<td>Assessment</td>
<td>Assessment-Submitted</td>
<td>The total number of times a student submitted an assessment for the course.</td>
<td></td>
</tr>
<tr>
<td>10. Assessment</td>
<td>Assessment</td>
<td>Assessment-View-Attempts</td>
<td>The number of times a student attempts to view an assessment for the course.</td>
<td></td>
</tr>
<tr>
<td>11. Assessment</td>
<td>Assessment</td>
<td>Assessment-View-List</td>
<td>The number of times a student viewed the assessments for the course.</td>
<td></td>
</tr>
<tr>
<td>12. Assessment</td>
<td>My Grades</td>
<td>View-Grades</td>
<td>The number of times a student opens the MyGrades tool.</td>
<td></td>
</tr>
<tr>
<td>13. Content</td>
<td>Content</td>
<td>Content-Page-Viewed</td>
<td>The total number of times content pages were accessed.</td>
<td></td>
</tr>
<tr>
<td>14. Content</td>
<td>Web-Links</td>
<td>Url-Viewed</td>
<td>The total number of times an external Internet URL was viewed or accessed by a student.</td>
<td></td>
</tr>
<tr>
<td>15. Content</td>
<td>Web-Links</td>
<td>Web-Links-Home-</td>
<td>The total number of times an</td>
<td></td>
</tr>
</tbody>
</table>
Filter ERS and Peoplesoft records.

Next, student characteristics records were filtered to only include students participating in the course whom received a final grade. The final student grade is the dependent variable and without it, the student could not be included for analysis. Four students received a “Withdraw” grade for the course and were excluded from the analysis. Two additional students participating in the course did not have a final grade. It is likely that these students were teaching assistants or
student support for the course. The LMS records for these six students were removed from the data set.

Join data sets.

Next, the LMS log file records were joined to the student characteristics using the anonymized student identifier. This identifier was present in each data set. This resulted in a consolidated data set that appended student characteristics and course grade to each LMS log file record.

Transform and Reduce Data

Recode categorical variable data types into numerical variable data types.

In order to include variables in a regression equation, they must be stored in a numerical data form. An accepted method to include categorical variables in regression equations is to recode these variables into numerical variables by creating a new variable and substituting categorical values with numerical values (e.g. Male = 0, Female = 1). This process was performed for gender, Pell-eligible, URM student, first in family to attend college, major-college and enrollment status.

Create generated variables.

Several variables were generated based on values in the data set. URM was created based on values in the race/ethnicity source data set. If a student was classified as “White”, "Asian", "Native Hawaiian/Pacific Islander", "Two or more races/ethnicities", or "Decline to State", the value was set to “0” (e.g., “no”). All other values were set to “1” (e.g., “yes”). The race/ethnicities thereby classified as URM included: "Black/African-American", "American Indian/Alaskan Native", and "Hispanic/Latino". First in family to attend college was created
based on the mother and father education variables. If the highest level of education for both mother and father was high school graduate, this variable was set to “1” (e.g., “yes”). All other values were set to “0” (e.g., “no”).

The student major was recoded to the university college unit that the major belonged to. This change reduced the range of potential values for this categorical variable from 189 to 10, thereby increasing the number of observations in each value. This improved the statistical power of analysis of this variable. However, this technique also grouped students with different majors into the same category, reducing the content validity of the variable. The number of majors present in each college variable is listed in Table 8. Some Colleges, such as Agriculture, have a small number of majors, and the subject matter of the majors is similar (e.g. Animal Science, Soil Science, etc.). Other Colleges, such as Behavioral and Social Science, have a large number of majors and the subject matter of the majors is different in some cases (e.g. Child Development, Economics, Corrections). This range of different values affects the integrity of the variable for analysis. With the large number of majors and the size of the study sample, this grouping was maintained for analysis purposes.

Table 8

<table>
<thead>
<tr>
<th>College</th>
<th>Count of Majors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>13</td>
</tr>
<tr>
<td>Behavioral and Social Sciences</td>
<td>34</td>
</tr>
<tr>
<td>Business</td>
<td>19</td>
</tr>
<tr>
<td>Communication and Education</td>
<td>24</td>
</tr>
<tr>
<td>Engineering, Computer Science and Construction</td>
<td></td>
</tr>
<tr>
<td>Management</td>
<td>22</td>
</tr>
<tr>
<td>Humanities and Fine Arts</td>
<td>43</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>27</td>
</tr>
<tr>
<td>Undeclared</td>
<td>1</td>
</tr>
<tr>
<td>Undergraduate Education</td>
<td>5</td>
</tr>
<tr>
<td>Total Number of Majors</td>
<td>189</td>
</tr>
</tbody>
</table>
Create interaction variables.

Two new variables were produced by the interaction of two original variables that indicate students at potentially higher risk of not succeeding in the course. URM and gender was used to create one variable, and URM and Pell-eligible student was used to generate the second variable. Dichotomous values were created, with a true value created only for the most at-risk population (URM and male, and URM and Pell-eligible).

Create aggregated LMS use variables.

LMS use per student was transformed into aggregate measures to answer the research questions. A new integer variable was created per LMS use category (e.g., administration, assessment, content, engagement) and this variable was set to “1” if the tool used in that log file entry belonged to that usage category (the categories and tool entries were defined in Chapter 1, Table 2).

Consolidate data into one record per student.

Next, the data set was consolidated into a single entry per student identifier. The student characteristics were identical in each record, and the first entry per student was used for these variables. The number of entries per LMS use category was counted, and the total was appended to the record in a new variable. In addition, to provide additional descriptive statistics, the total number of LMS log file entries, the total number of LMS log file entries per tool (e.g., discussion, mail, etc.) and the dwell time per tool and LMS use category was totaled and appended to each student record. This process resulted in the final data file that was used for the analysis.
Examine Data to Meet Analysis Assumptions

Missing data.

Each variable was reviewed for the presence of missing data. In the case of the LMS activity variables, a missing value represented no activity and was replaced with a “0”. For the student characteristic variables, missing data represents a value that was unknown in the source data. Statistical analysis cannot be conducted on missing data. The variables included in the research were analyzed for missing data before including them in the research analysis. That review process led to the elimination of several variables of interest due to missing data in greater than 10% of the observations. In the final data set, only high school GPA (21 observations, 5.63%) and first in family to attend college (2 observations, .54%) were missing data. Due to the small number of missing values, the mean across cases was used to impute the missing data values (Schumacker & Lomax, 2010). Because first in family to attend college is a dichotomous variable and the population mean was .28 (e.g., large majority with a value of “no”), both missing values were set to “no”.

Inspect student characteristics and final grade for dispersion.

Fundamentally, statistical methods analyze how differences between variables in a group of observations are related to one another. This requires that the variable values are different from one another between observations. The values for each variable in the analysis were inspected using tabulations to ensure that there was sufficient dispersion of the values in the variables.

Inspect and compare distributions of LMS use data.

The data analysis methods used for this research assume the normal distribution of variables within the population. Histograms are efficient means to test the normality of the
distribution and homogeneity of variance for continuous variables. These characteristics are assumptions that must underly regression analysis analysis (Huck, 2008). Student characteristics by definition cannot have extreme values because they are coded into categorical values or predefined ranges. The LMS use category website hits and dwell time variables were plotted with a histogram. These values were compared to one another to confirm that website hits was the best variable to use for the analysis based on the distribution of the data. The histograms were also inspected for outlying data values that fell outside a normal curve that was fitted to the data.

**Examination for independence of variables.**

The multivariate analysis methods assume that each variable has independent variation from other variables within an observation. If two (or more) variables had similar variance, either variable would have the same relationship with the dependent variable and would skew the analysis. To test for the independence of student characteristic variables, a correlation matrix was created, and variables with greater than .50 correlation were considered for removal. No correlations above this level were found, and all variables were maintained for analysis. To test for the independence of LMS use variables, a principal component analysis was performed.

**Data Analysis**

**Research question 1: correlation analysis of LMS use and final grade.**

The purpose of the first analysis method is to address the following research questions: (a) What is the relationship between a student's LMS use and their academic achievement? (b) Does this relationship vary by the pedagogical purpose underlying the student's LMS use? Correlation between each aggregated LMS use variable (total hits, administrative hits,
assessment hits, content hits, and discussion hits) and student final grade was conducted to test this question. The Pearson product-moment discrete nominal was used because the variables are quantitative and based in a raw (e.g., not transformed or ranked) score (Huck, 2008). The results are presented in a correlation matrix and differences between the correlation coefficients are calculated. Of primary interest is the relationship between the total hits variable and each of the categories of LMS use. Of secondary interest is the difference between the categories of LMS use. Statistical significance was established at the p < .05 level. Interpretation of effect sizes is a complex issue and scholars recommend considering the context in which the study is made (Pedhazur & Kerlinger, 1982). However, Cohen's (1988) definition is widely used to set general parameters on interpreting effect sizes. This definition defines effect sizes of .3 as small, .5 as medium, and .8 as large. Variables found to not achieve statistical significance were excluded from subsequent multivariate regression.

**Research question 2: correlation analysis of student background characteristics and final grade.**

The next analysis investigated the following research question: What is the relationship between a student's background characteristics and their academic achievement? A correlation analysis between each student characteristic variable with their final grade was conducted to test this question. Table 9 describes the correlation coefficients that were used for each of these variables, which were selected according to the independent variable data type (Huck, 2008).

<table>
<thead>
<tr>
<th>#</th>
<th>Independent Variable</th>
<th>Data Type</th>
<th>Correlation Coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Gender</td>
<td>Discrete, Dichotomous</td>
<td>Point-Biserial</td>
</tr>
<tr>
<td>2.</td>
<td>URM</td>
<td>Discrete, Dichotomous</td>
<td>Point-Biserial</td>
</tr>
</tbody>
</table>
For each of these variables, correlation with the final grade was calculated. The results are presented in a correlation matrix and rank order and differences between the correlation coefficients are calculated. The results of this analysis demonstrate the relative importance of each variable. The variables found to be statistically significant (p < .05) in this analysis were included in subsequent multivariate regression.

While these statistics are useful, they are limited in their explanatory power. Students do not use only one type of LMS activity, nor do they have a single defining background characteristic. Instead, they engage in complex practices and have diverse backgrounds. A multivariate regression analysis was used to uncover the relationship between these complex uses and backgrounds and student achievement.

**Research Question 3: multivariate regression analysis of LMS use, student background characteristics and final grade.**

The next set of methods investigated the following research question: How does analyzing a student's background characteristics change the predictive relationship between their...
combined LMS use data and academic achievement? This question was investigated in three phases: (a) What is the relationship between a student's combined LMS use and their academic achievement? (b) What is the relationship between a student's combined LMS use, combined background characteristics, and their academic achievement? and (c) What is the relationship between a student's combined LMS use, combined student background characteristics, interactions between demographic variables (URM and Pell-eligibility, URM and gender), and academic achievement?

First, multivariate regression was run of final grade on the combined LMS use variables. (restricted model 1) This analysis was followed with a regression equation that added the variables for student background characteristics (restricted model 2). The final part of this question added the interaction variables to the combined LMS use and student characteristic regression. This created the complete model with all variables included in the analysis.

**Research Question 4: multivariate regression analysis of LMS use and student characteristics by at-risk population sub-samples.**

The final research question divided the population into two sub-samples by the value for the URM and Pell-eligibility interaction variable. One sub-sample was created for students that had "yes" values for each variable (n = 80) and another sub-sample was created for students that had a "no" value for either variable (n = 293). This research question determined if the previous research results applied equally to students predicted to be at-risk of low achievement by their race and socio-economic status characteristics. The LMS variable regression (research question 3a) was run again on each of these populations and the results compared between these sub-samples and between each sub-sample and the entire population.

Next, a regression of LMS variables and a restricted set of student characteristics was run for each of these sub-samples. Because the population was sampled by URM and Pell-
eligibility, variables that included either of these terms were eliminated. This reduced the student characteristics included in the regression from nine to five variables. These results were compared between sub-samples to determine if the results applied equally to both. In addition, the results were compared to the complete model to determine which method had a higher degree of predictive accuracy.
RESEARCH FINDINGS

Overview

This chapter contains the results of the research. These results include descriptive data about the course participants' background and their use of the LMS. In addition, the results provide the inferential statistical findings about the relationship between these data and student course grade. A substantial number of students had criteria that placed them at-risk to receive low grades, such as Pell grant eligibility (56%), URM students (27%), and first in family to attend college (17%). Many of these criteria were higher in this course population than they were in the average CSU Chico undergraduate enrollment.

The LMS log file required extensive filtering to create an accurate data set that isolated educationally relevant student activities from the technical processes of the server. This filtering reduced the number of records from 249,490 to 68,069 (-78%). After this reduction, student average total LMS use was calculated at a mean value of 182 hits per student. The variables for different types of LMS use had high variation with a range from 16 mean hits per student for administrative activities to 91 mean hits per student for engagement activities. Student time spent on the course had a mean value of 5 hours according to the dwell time variable.

Course grade was the study’s dependent variable, and it had a good dispersion and distribution typical of a lower division course in most respects. However, 54 students failed the course (15%). This proportion of students was higher than two concurrent sections of the same course taught with the conventional course design and a lower student enrollment.
The null hypothesis was rejected for each of the research questions. Correlation analysis of the LMS use variables found that each variable was significantly related to course grade at the p < .0000 level, with moderate effect sizes (r = .35 to r = .47). Seven of the nine student characteristic variables were found to achieve significance at the p < .05 level, with a larger range in the effect size (r = -.11 to r = .31). The gender and major-college variables were not statistically significant.

The LMS variable correlations had a stronger significance and higher magnitude than the student characteristic correlations. The LMS activity variable with the smallest magnitude (administrative hits, r = .35) was larger than the largest magnitude student characteristic variable (high school grade point average, r = .31). Most likely, this difference in correlation strength is affected by features of the data, as categorical and dichotomous data has a low degree of variation and is not ordered in a meaningful way (e.g., small to large) (R. Taylor, 1990).

Multivariate regression found the combination of LMS activity to have a significant and small strength relationship with student grade (R² = .25). The complete model with all student characteristic variables yielded a slight increase in the predictive relationship with an R² value of .35. The interaction variables included in this model of URM-Pell eligible and URM-gender resulted in a very small (.07) increase in the R² value. Therefore, these results demonstrate that LMS variables have a predictive relationship to student achievement and are more effective predictors than standard demographic and enrollment variables typically used to identify students at-risk of low achievement.

To further investigate the potential effect of demographic variables on LMS analysis results, the population was split into sub-samples by the URM-Pell eligible variable. The LMS variable regression and a modified LMS and student characteristic regression were run on each
sub-sample. This analysis found that there was a 25% decrease in the $R^2$ value for at-risk students. Despite the lower predictive magnitude of student characteristic results compared to LMS variables, important differences between students based on these criteria persisted, and they are useful for predictive analysis of student achievement.

**Course Participant Descriptive Data Results**

Potential errors in an observational study include a population that is skewed on a certain characteristic (e.g., low participation by a racial/ethnic group, high participation by students with high SAT scores, etc.), low representation within a certain category, and/or missing data on a variable of interest. To ensure that the population in this study did not have any of these errors and was representative of the CSU Chico undergraduate population, comprehensive descriptive statistics of the population were calculated and any deviations from expected distributions were noted.

**Baseline demographic data.**

As indicated in Table 10 below, the course was imbalanced by gender, with 231 (62%) female students. The racial/ethnic composition of the course was diverse and relatively well distributed given the overall composition of Chico State, with 188 (50%) White students, 92 (25%) Hispanic/Latino students, and smaller numbers of students from other racial/ethnic groups. This distribution resulted in 109 (29%) students from URM backgrounds. There was also high enrollment by students from low socio-economic status backgrounds, with 163 (44%) of the students eligible for Pell grants based on family financial need. Overall, the course contained a substantial number of students from at-risk demographic groups as determined by race, socio-economic status, and gender.
Comparisons of the course population to the entire CSU Chico student population were calculated and yielded noteworthy differences. The campus population figures were calculated for students with less than 70 units completed in order to provide an accurate comparison for a lower-division undergraduate course. Pell-eligibility was the largest difference, with more students (+12%) in the course coming from a low-income background than the average population. The course also had more female students (+11%) and less White students (-9%) than the overall campus population. The lower number of White students was made up for by a higher number Hispanic/Latino and Asian students. The participation of URM students was similar to their presence on the campus, as both White and Asian students are classified as non-URM. Taken as a whole, the course population had sufficient participation in each of the demographic levels for statistical analysis and there was a slightly greater number of students known to be at-risk of not succeeding in the course than the overall campus population due to their socio-economic status.

Student age exhibited low variation in the study population, with 350 (88%) of the students within the traditional college-attending age range of 18-21 years old. This population is typical for CSU Chico, but is not sufficient to conduct meaningful statistical analysis. Therefore, age was excluded from the student characteristic variables in the statistical analysis portion of the study.

Table 10

<table>
<thead>
<tr>
<th>Distribution of Participant Gender Race, Socio-Economic Status, and Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Race</td>
</tr>
<tr>
<td>-----------------------</td>
</tr>
<tr>
<td>White</td>
</tr>
<tr>
<td>Black/African American</td>
</tr>
<tr>
<td>American Indian/Alaskan Native</td>
</tr>
<tr>
<td>Asian</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
</tr>
<tr>
<td>Two or More Races/Ethnicities</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
</tr>
<tr>
<td>Decline to State</td>
</tr>
<tr>
<td>URM</td>
</tr>
<tr>
<td>Not URM</td>
</tr>
<tr>
<td>URM Student</td>
</tr>
<tr>
<td>Pell-Eligible</td>
</tr>
<tr>
<td>Not Pell-Eligible</td>
</tr>
<tr>
<td>Pell-Eligible</td>
</tr>
<tr>
<td>Age</td>
</tr>
<tr>
<td>17</td>
</tr>
<tr>
<td>18-21</td>
</tr>
<tr>
<td>21+</td>
</tr>
</tbody>
</table>

*Mean of students with <70 units

**Additional demographic data on family history and educational preparation.**

Additional participant demographic data in Table 11 shows a diverse course enrollment. These figures are consistent with the previous findings of a substantial number of students at-risk of low achievement in the course. There were 105 students (28%) who were the first in their families to attend college. Surprisingly for a lower-division course, only 139 (37%) of the students were first-time enrollees, and 217 (58%) were continuing students who had previously been enrolled at Chico State.
Student majors were coded into variables that indicated to which of the eight colleges the major belonged. The results indicate that students in the course majored in a broad variety of disciplines, with the largest number (98 students, 27%) belonging to the College of Behavioral and Social Sciences. There were 8% more students with majors from this college than the general university population, which was balanced by lower enrollment from students in the College of Communication and Education as well as Business. There were 49 students (13%) who had not declared a major.

Table 11

Course Participant Demographics: Additional Family and Educational Background

<table>
<thead>
<tr>
<th></th>
<th>Frequency</th>
<th>Percent</th>
<th>University Mean*</th>
<th>% Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>First in Family to Attend College</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>No</td>
<td>268</td>
<td>72</td>
<td>83</td>
<td>-12</td>
</tr>
<tr>
<td>Yes</td>
<td>105</td>
<td>28</td>
<td>17</td>
<td></td>
</tr>
<tr>
<td>Enrollment Status</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Continuing Student</td>
<td>217</td>
<td>58</td>
<td>62</td>
<td>-4</td>
</tr>
<tr>
<td>Transfer</td>
<td>17</td>
<td>5</td>
<td>10</td>
<td>-5</td>
</tr>
<tr>
<td>First-Time Student</td>
<td>139</td>
<td>37</td>
<td>28</td>
<td>10</td>
</tr>
<tr>
<td>Student Major by College</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agriculture</td>
<td>8</td>
<td>2</td>
<td>3</td>
<td>-1</td>
</tr>
<tr>
<td>Behavioral and Social Sciences</td>
<td>98</td>
<td>26</td>
<td>18</td>
<td>8</td>
</tr>
<tr>
<td>Business</td>
<td>50</td>
<td>13</td>
<td>16</td>
<td>-3</td>
</tr>
<tr>
<td>Communication and Education</td>
<td>56</td>
<td>15</td>
<td>18</td>
<td>-3</td>
</tr>
<tr>
<td>Engineering, Computer Science and Construction Management</td>
<td>17</td>
<td>5</td>
<td>12</td>
<td>-7</td>
</tr>
<tr>
<td>Humanities and Fine Arts</td>
<td>38</td>
<td>10</td>
<td>10</td>
<td>0</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>57</td>
<td>15</td>
<td>13</td>
<td>2</td>
</tr>
</tbody>
</table>
Student high school grade point average (HS GPA) data exhibited a distribution that conformed to a normal curve, as illustrated in Figure 4. There were few students at either end of the distribution, with the majority between 3 and approximately 3.3 grade points (a letter grade of B). A few students had GPA averages above 4.0, the maximum, due to participation in advanced placement courses during high school. HS GPA was the only variable with a continuous and ordinal data type of the student characteristic variables.

<table>
<thead>
<tr>
<th>University</th>
<th>Frequency</th>
<th>Percent</th>
<th>University Mean*</th>
<th>% Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undeclared</td>
<td>49</td>
<td>13</td>
<td>9</td>
<td>4</td>
</tr>
</tbody>
</table>

*Figure 4. High School GPA Histogram.*

**Student characteristic variable interactions and cross-tabulations.**

Two variables were generated to represent interaction between variables. The values for these variables are presented in Table 12. A systematic relationship between URM status and Pell-eligibility was found; 80 of the 109 (73%) URM students were Pell-eligible, while only 83
of the 264 (31%) non-URM students were Pell-eligible. Due to adequate student numbers in each cell of the cross-tabulations, there was substantial amount of statistical power to analyze the interaction of URM-Pell with course grade. A clear correlation between gender and URM status was not found, although there were enough observations in each cell to provide the power needed for statistical analysis.

Table 12

Cross Tabulations of Demographic Variables

<table>
<thead>
<tr>
<th>URM and Pell-Eligibility Interaction</th>
<th>Pell Eligible</th>
<th>Not Pell Eligible</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not URM</td>
<td>83</td>
<td>181</td>
<td>264</td>
</tr>
<tr>
<td>URM</td>
<td>80</td>
<td>29</td>
<td>109</td>
</tr>
<tr>
<td>Total</td>
<td>163</td>
<td>210</td>
<td>373</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URM and Gender Interaction</th>
<th>Male</th>
<th>Female</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not URM</td>
<td>105</td>
<td>159</td>
<td>264</td>
</tr>
<tr>
<td>URM</td>
<td>37</td>
<td>72</td>
<td>109</td>
</tr>
<tr>
<td>Total</td>
<td>142</td>
<td>231</td>
<td>373</td>
</tr>
</tbody>
</table>

Additional cross tabulations of student characteristic variables.

Descriptive data and cross-tabulations not included in the statistical analysis were also calculated to provide additional information about the characteristics of the study population. The number of observations in some cross-tabulations was so low that statistical analysis could not be performed. Table 13 provides cross-tabulations of several characteristics that the literature review indicated may be of relevance to student achievement. Race and Pell-eligibility had predictable outcomes, with students from each URM race/ethnicity more likely to be classified as Pell-eligible than White students. However, the both the Pell-eligibility / race and first in family to attend college / race cross-tabulations challenged stereotypes about the relationships between race, poverty, and family education levels. Almost one-third of White students were Pell-eligible. Many URM students had the benefit of family expectations and
support that are often provided by parents with college experience, with 52 of 109 (48%) of the URM students from families in which a parent had attended college. This finding suggests that racial background alone may not be a conclusive indicator of the presence or absence of factors contributing to persistence.

Table 13

Additional Cross Tabulations of Student Variables: Pell, Race, URM

<table>
<thead>
<tr>
<th>Race</th>
<th>Not Pell-Eligible</th>
<th>Pell Eligible</th>
<th>Total</th>
<th>% Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>White</td>
<td>146</td>
<td>42</td>
<td>188</td>
<td>50%</td>
</tr>
<tr>
<td>Black/African American</td>
<td>2</td>
<td>10</td>
<td>12</td>
<td>3%</td>
</tr>
<tr>
<td>American Indian/Alaskan Native</td>
<td>1</td>
<td>3</td>
<td>4</td>
<td>1%</td>
</tr>
<tr>
<td>Asian</td>
<td>7</td>
<td>28</td>
<td>35</td>
<td>9%</td>
</tr>
<tr>
<td>Native Hawaiian/Pacific Islander</td>
<td>1</td>
<td>0</td>
<td>1</td>
<td>0%</td>
</tr>
<tr>
<td>Two or More Races/Ethnicities</td>
<td>8</td>
<td>4</td>
<td>12</td>
<td>3%</td>
</tr>
<tr>
<td>Hispanic/Latino</td>
<td>25</td>
<td>67</td>
<td>92</td>
<td>25%</td>
</tr>
<tr>
<td>Decline to State</td>
<td>20</td>
<td>9</td>
<td>29</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td>210</td>
<td>163</td>
<td>373</td>
<td>99%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>URM and First in Family to Attend College</th>
<th>Not First in Family</th>
<th>First in Family</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Not URM</td>
<td>216</td>
<td>48</td>
<td>264</td>
</tr>
<tr>
<td>URM</td>
<td>52</td>
<td>57</td>
<td>109</td>
</tr>
<tr>
<td>Total</td>
<td>268</td>
<td>105</td>
<td>373</td>
</tr>
</tbody>
</table>

Student major was found to exhibit systematic variance by URM status as indicated in Table 14. Of the students enrolled in the course, there were a higher proportion of URM students than non-URM belonging to the colleges of Behavior and Social Science and Humanities and Fine Arts. Majors from these two colleges typically have lower attrition rates than majors from engineering or science-related disciplines, which had lower proportions of
URM students. This finding suggests that compared to non-URM students, URM students had less experience with demanding courses. On the other hand, URM students may have had more interest in the subject matter of the course, as it is closer to their major area. This factor may have made these students more motivated to participate in the course than non-URM students.

Table 14

*Additional Cross Tabulations of Student Variables: URM and Major College*

<table>
<thead>
<tr>
<th>Major College and URM Status</th>
<th>Total</th>
<th>Not URM</th>
<th>URM</th>
<th>Not URM (%)</th>
<th>URM (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agriculture</td>
<td>8</td>
<td>7</td>
<td>1</td>
<td>3%</td>
<td>1%</td>
</tr>
<tr>
<td>Behavioral and Social Sciences</td>
<td>98</td>
<td>57</td>
<td>41</td>
<td>22%</td>
<td>38%</td>
</tr>
<tr>
<td>Business</td>
<td>50</td>
<td>40</td>
<td>10</td>
<td>15%</td>
<td>9%</td>
</tr>
<tr>
<td>Communication and Education Engineering, Computer Science and Construction Management</td>
<td>56</td>
<td>43</td>
<td>13</td>
<td>16%</td>
<td>12%</td>
</tr>
<tr>
<td>Humanities and Fine Arts</td>
<td>38</td>
<td>24</td>
<td>14</td>
<td>9%</td>
<td>13%</td>
</tr>
<tr>
<td>Natural Sciences</td>
<td>57</td>
<td>42</td>
<td>15</td>
<td>16%</td>
<td>14%</td>
</tr>
<tr>
<td>Undeclared</td>
<td>49</td>
<td>37</td>
<td>12</td>
<td>14%</td>
<td>11%</td>
</tr>
<tr>
<td>Total</td>
<td>373</td>
<td>264</td>
<td>109</td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>

A correlation matrix of student characteristic variables was calculated in order to ensure that the variables had sufficient independent variation to permit their inclusion in the regression equations. The matrix provides both the level of statistical significance (whether the relationship is due to chance effect or not) and the magnitude of the relationship between the two variables. The correlation matrix in Table 15 indicates several statistically significant relationships, some of which were suggested in the previous cross tabulations.

Out of a total of 36 potential correlations, 21 pairs achieved statistical significance. Out of the six pairs with magnitudes greater than .40, four pairs were invalid as they included an interaction variable with the same term in both variables (e.g., they had high correlations by
First in family to attend college exhibited a high correlation with Pell-eligible (.48) as well as the Pell-URM interaction variable (.41). This finding indicates that first in family had low variation in these pairs, but it did have significant independence from six other variables and was, therefore, retained for statistical analysis. All other correlations were below .34, with most below .20. All variables were retained as a result of this procedure.
Table 15

*Intercorrelation Analysis of Student Characteristic Independent Variables*

<table>
<thead>
<tr>
<th></th>
<th>Gender</th>
<th>URM</th>
<th>Pell Eligible</th>
<th>HS GPA</th>
<th>First in Family</th>
<th>Major-College</th>
<th>Enroll. Status</th>
<th>URM-Pell</th>
<th>URM-Gender</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>URM</td>
<td>.05</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pell Eligible</td>
<td>.08</td>
<td>.39</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HS GPA</td>
<td>.15</td>
<td>-.16</td>
<td>.41</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First in Family</td>
<td>.09</td>
<td>.35</td>
<td>.48</td>
<td>-.02</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major-College</td>
<td>-.02</td>
<td>-.07</td>
<td>-.05</td>
<td>.06</td>
<td>.00</td>
<td>.14</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>URM-Pell</td>
<td>.05</td>
<td>.81</td>
<td>.59</td>
<td>-.09</td>
<td>.41</td>
<td>-.08</td>
<td>.19</td>
<td>1.00</td>
<td></td>
</tr>
<tr>
<td>URM-Gender</td>
<td>-.43</td>
<td>.52</td>
<td>.20</td>
<td>-.20</td>
<td>.15</td>
<td>-.02</td>
<td>.14</td>
<td>.42</td>
<td>1.00</td>
</tr>
</tbody>
</table>

*Bold* indicates statistical significance at the p = .05 level
CSU Chico maintains many additional variables about students that previous research indicates are related to student achievement. While these additional variables were considered for inclusion in the study, many were excluded due to a large number of observations with missing records. Variables that were excluded with the highest amount of potential relevance to this study are listed in Table 16 in descending order by the quantity of missing records. Several variables describing student academic preparation for college were excluded, including information about remediation and ACT/SAT scores.

Some of this missing data was due by student application and enrollment status processes. SAT and ACT scores are not required for students who transfer from another institution and are, thus, infrequently reported for these students. First time students do not have a current college GPA because they have not previously taken a college course. Other variables, such as status regarding remediation, were recently added to institutional databases and have been recorded for a low number of students.

These omitted variables are important due to the loss of the content they represent and the quality of the data for statistical analysis. Four of the omitted variables were continuous data types that were well suited to correlation and regression methods. Only one variable retained for analysis (HS GPA) included either background educational experience or continuous data. This issue is likely to have influenced the results for the analysis of student characteristic variables.

Table 16

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count of Records Missing Data</th>
<th>Percent of Population Missing Data</th>
<th>Data type</th>
</tr>
</thead>
<tbody>
<tr>
<td>Remediation currently</td>
<td>278</td>
<td>75</td>
<td>Dichotomous</td>
</tr>
</tbody>
</table>
### Needed Remediation Needed at Entry

<table>
<thead>
<tr>
<th></th>
<th>ACT Scores (English)</th>
<th>ACT Scores (Math)</th>
<th>SAT Scores (English)</th>
<th>SAT Scores (Math)</th>
<th>College Prep Courses Taken (English)</th>
<th>College Prep Courses Taken (Math)</th>
<th>Current College GPA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>241</td>
<td>240</td>
<td>59</td>
<td>59</td>
<td>21</td>
<td>21</td>
<td>126</td>
</tr>
<tr>
<td>Dichotomous</td>
<td>38</td>
<td>64</td>
<td>16</td>
<td>16</td>
<td>6</td>
<td>6</td>
<td>34</td>
</tr>
</tbody>
</table>

|                         | Continuous           | Continuous        | Continuous           | Continuous        | Dichotomous                         | Continuous                      |                    |

### LMS Use Descriptive Data Results

Log files for LMS use were available for every student in the course and all variables were continuous data types. However, log file data required substantial filtering and analysis in order to create accurate representations of student LMS activity. LMS website hits were found to have better value distribution than dwell time, even though hits do not distinguish a student writing a thoughtful discussion post from a student quickly reading someone else's discussion reply. The website hits were divided into categories based on the tool used in the log file entry. There was a low amount of variation between students in their use per each variable. This reduced the potential value of the distinctions between types of LMS use for the analysis as differences in grade could not be attributed to changes in the tools used.

#### LMS log file filtering results.

The initial LMS log file had 249,490 records. This file was initially filtered to factor out faculty usage, which removed 5,035 (2%) of the records from analysis. Discussion with Chico State staff familiar with the log file and review of the records led to further filtering that eliminated records that duplicated entries for the same student action (e.g., reading a discussion message requires clicking on a discussion topic). The most efficient initial approach to eliminate
these duplicate items was to filter the records by time. Consequently, records were filtered for inclusion in the data set that were greater than five seconds and less than one hour. Additional filters were applied by LMS action as described in the Methods section. The result was a reduction of the number of log file entries to 68,337 (-73%). Strikingly, almost three-quarters of the log file were eliminated by these criteria as not indicative of student educational activity. The dramatic reduction in the number of entries is an important finding when considering the potential use of LMS data for understanding and predicting student achievement.

As illustrated in Figure 5, filtering was unevenly distributed between LMS tools and use categories. For the purposes of identifying the source of filtered items, “engagement” was analyzed by the discussion and mail tools that pertain to the category. The majority of entries filtered out of the data set used the discussion tool. Entries using this tool were reduced from a mean value of 382 hits per student to a mean value of 54 hits per student. This large decrease resulted from how discussion items were recorded in the log file. If a student selected the "compile discussion" action on a discussion thread, each message in that thread created a log file entry with time = 0. Filtering by time removed these duplicate entries. Although Chico State staff had identified the compile discussion issue prior to this study, the magnitude of the issue was not known.

The second-highest area from which entries were removed was the content tool, which was used to view index pages and duplicated the action of viewing the content item. This finding demonstrates the importance of analyzing log file entries prior to analysis. Without filtering, there is a potential for a high degree of inaccuracy in summary statistics and subsequent inferential analysis.
To confirm that LMS use data met the assumption of normal distribution required for regression analysis, a histogram was created. Histograms also test for the presence of outlier data values. These distributions were created for two measures of LMS use: the count of log file entries and the total dwell time for each student.

Dwell time was considered as a variable that might provide a better indication of the educational relevance of a log file entry, as it not only indicated the action that had been performed, but also the length of time a student spent on that activity. The histograms in Figure 6 indicate a relatively normal distribution of LMS website hits with few outliers. The outliers do not deviate sufficiently from the distribution to skew the results. However, the dwell time variable significantly skews toward low values. It also has more outlying values than website hits. Dwell time exhibited a small range that was skewed and was therefore ill-suited for inferential analysis.
These distribution characteristics are stronger when applied to LMS use categories, as indicated in Figure 7, which calculates histograms for actions using administrative tools (announcement, calendar). The findings are consistent with how these tools are used: calendar items are quickly read and announcements are meant to provide updates. Thus, these distributions indicated that the frequency of access provided by hits was the best measure to use for inferential statistics.
LMS Use frequencies.

The data in Table 17 indicates substantial student use of the LMS with high variation in that use between tools for the entire population. There was a mean of 182 hits per student with a range of 8-674 and a standard deviation of 105. Total dwell time had a mean value of 5 hours with a range from 0-25 hours. As indicated previously, there was a high standard deviation in dwell time at 4 hours. The total hits by tool demonstrate how the course was used by students for learning activities, with the largest amount of activity in the discussion tool (53), followed by content pages (39) and mail (35). There were no other tools with a mean value above 20 hits per student. There were also large ranges in student frequency of use.

One of the tools with the lowest frequency was “assessment,” with a mean value of 2. There was a high standard deviation for this tool at 3 hits per student. This tool is used to submit quizzes and, therefore, does not have a high frequency of use. Whereas content pages record a log file entry for each page opened, assessments only record a log file entry for opening or submitting an assessment. The other tool in the assessment category, assignments, had a higher frequency of use with a mean value of 14. In several tools (e.g., calendar, my grades, web links) the standard deviation was nearly as large as the mean value, suggesting that there was a high dispersion of use at the tool level. Given these results, it is not advised to conduct analysis at the tool level because of the weak dispersion of data values for some tools.

Table 17

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
</table>

Descriptive Data on LMS Use: Summary and by Tool
Instead of conducting analysis with observed variables, or dropping selected variables due to the distribution of use, this study aggregated LMS use by tools into categories. This method led to improved distributions of the data without eliminating log file entries from any tool variables. The descriptive data on the use by category of tools in Table 18 is consistent with the data of the individual tools. For example, engagement activities (discussion, mail) had the highest frequency of use (91 mean hits per student) and a range from no use to 429 maximum hits. The next highest use was content tools, with almost a 50% reduction in the mean value (51 mean hits), which was followed by assessment tools (23 mean hits) and administrative tools (16 mean hits). The distribution of the data was improved by grouping the observations into categories with standard deviations far below the mean value for all categories. The grouping into categories appeared to improve the data quality and provide a sufficient range of values for statistical analysis.
Table 18

**Descriptive Data on LMS Use: Categories Used for Inferential Analysis**

<table>
<thead>
<tr>
<th>Category</th>
<th>Tools in Category</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Administrative</td>
<td>Announcement</td>
<td>16</td>
<td>9</td>
<td>0</td>
<td>57</td>
</tr>
<tr>
<td></td>
<td>Calendar</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assessment</td>
<td>Assessment</td>
<td>23</td>
<td>17</td>
<td>0</td>
<td>97</td>
</tr>
<tr>
<td></td>
<td>Assignments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>My-grades</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Content</td>
<td>Content-page</td>
<td>51</td>
<td>39</td>
<td>0</td>
<td>265</td>
</tr>
<tr>
<td></td>
<td>Web-links</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Engagement</td>
<td>Discussion</td>
<td>91</td>
<td>60</td>
<td>0</td>
<td>429</td>
</tr>
<tr>
<td></td>
<td>Mail</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In addition to the distribution of LMS use across the entire population, the variation by student between LMS use category variables were calculated using factor analysis. The analysis found that there was a low degree of variation. The factor analysis results are presented in Table 19. The eigenvalue for the first variable (administrative hits) explained 66% of the variance by student for all four LMS use categories (administrative, assessment, content, engagement). This variable had the lowest frequency of use and the smallest range, which would make it less likely to explain the total variation than another value with a higher amount of use and variation between students. The next variable (assessment) explained an additional 16% of the variation, resulting in 82% of the total variance in LMS use explained by these two variables alone. Content and engagement hits, which had the highest frequency of use, only explained 18% of the remaining variance in LMS use by student.

These high eigenvalues indicate a low amount of variation by student between the LMS use categories: in other words, some students used the LMS frequently, and others used it infrequently (or not at all), and they did so consistently across all categories of use. Despite this high degree of shared variance, the LMS use categories were retained in consideration of the
problems framing this study and the findings in previous literature, in particular research on how specific uses of educational technology support learning (Bernard et al., 2004; Tamim et al., 2011).

Table 19

*Factor Analysis Results of LMS Use Categories*

<table>
<thead>
<tr>
<th>Component</th>
<th>Initial Eigenvalues</th>
<th>Extraction Sums of Squared Loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Total</td>
<td>% of Variance</td>
</tr>
<tr>
<td>Administrative Activity Hits</td>
<td>2.64</td>
<td>66</td>
</tr>
<tr>
<td>Assessment Activity Hits</td>
<td>.62</td>
<td>16</td>
</tr>
<tr>
<td>LMS Content Activity Hits</td>
<td>.47</td>
<td>12</td>
</tr>
<tr>
<td>Engagement Activity Hits</td>
<td>.26</td>
<td>6</td>
</tr>
</tbody>
</table>

**Distribution of course grade dependent variable.**

The course grade distribution is illustrated in Figure 8. The course had a high occurrence of low grades: 23% of the students received a D or F grade in the course. The grade distribution is not notable in other respects for a lower-division undergraduate course because 45% of the students received grades of B- or higher. However, the number of D and F students were 9 and 11% higher than two other sections of this course taught concurrently by the same instructors using traditional methods with a smaller number of students (Vela, 2011).
Research Question 1 results: correlation results of LMS use with course grade.

The first research question investigated the relationship between LMS use and the final course grade. In addition to the LMS use categories, the total hits per student was correlated as a point of comparison. As detailed in Table 20, each of the five variables was positively correlated with final grade. For all variables, statistical significance was achieved at the $p < .0000$ level, with coefficients of determination ranging from .35 (administrative activities) to .48 (total hits). This finding rejects the null hypothesis for the research question and indicates that there was systematic variance between LMS use variable and final grade.

Squaring the coefficient calculates the amount of variance in the dependent variable explained by variance in the independent variable. Using this calculation, LMS use accounts for
between 12% and 23% of the variation in course grade. Overall, the amount of variation between the coefficients is surprisingly high when the results of the factor analysis are considered. Given that there is a low degree of variance in LMS use by category by student, a low degree of variance in coefficients would seem to be a likely result. The coefficients have a range of .12, over one-third of the value in the smallest coefficient. This range in coefficients suggests that there is a substantial amount of variance in final grade due to the difference in LMS use in student practices.

Ordering the variables by magnitude of the coefficient shows that total hits had a higher coefficient than any of the variables for specific uses of the LMS. This result is difficult to explain on the basis of educational relationships; the total use would seem more likely to be a mean value, with some coefficients greater than the total hits coefficient and others lower than it. One explanation for this result could be related to data distribution and analysis. There is a larger amount of variation between students in total hits than in the other LMS use variables. This variation could lead to a greater differentiation between students, which could lead to a higher coefficient. This result did not test for the whether the combined LMS use category variables had a lower predictive relationship with final course grade than the total use. The regression analysis tests for that question.

In the categories of LMS use, coefficients for engagement activities (.40) and content activities (.41) were nearly equivalent. Considering that engagement activities had the highest frequency of student activity of any LMS use category, and were anticipated to fill an instructional gap created by the large format course, this finding is surprising. Engagement activities were expected to have a higher value than content activities. Instead of engagement activities having a lower value than anticipated, this finding could be interpreted to demonstrate
that content activities were more effective than anticipated. This is possible since they were the only mechanism by which some course content was available to students. Administrative activities had the lowest coefficient value (.35), and assessment activities had the highest coefficient (.47). These are predictable findings; for example, reading a calendar item would reasonably have a lower impact on student achievement than submitting a graded quiz.

Table 20

*Correlation results of LMS use with course grade.*

<table>
<thead>
<tr>
<th>LMS Use Variables (ordered by r values)</th>
<th>r</th>
<th>% Variance</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Hits</td>
<td>.48</td>
<td>23%</td>
<td>.000</td>
</tr>
<tr>
<td>Assessment Activity Hits</td>
<td>.47</td>
<td>21%</td>
<td>.000</td>
</tr>
<tr>
<td>LMS Content Activity Hits</td>
<td>.41</td>
<td>17%</td>
<td>.000</td>
</tr>
<tr>
<td>Engagement Activity Hits</td>
<td>.40</td>
<td>16%</td>
<td>.000</td>
</tr>
<tr>
<td>Administrative Activity Hits</td>
<td>.35</td>
<td>12%</td>
<td>.000</td>
</tr>
<tr>
<td>Mean Value All Significant Variables</td>
<td></td>
<td>18%</td>
<td></td>
</tr>
</tbody>
</table>

Although correlation coefficients are a powerful statistic for making comparisons between variables and communicating results, they reduce the complexity of individual observations contributing to these trends. Correlation coefficients smooth out the variation and provide a single measure. In Figure 9 below, a scatterplot was plotted to demonstrate the complete relationship between assessment activity LMS hits and the course grade. Validating the results of the correlation, the scatterplot displays a higher density of low assessment hits at lower grades and increased hits at higher grades. The trend based on these plotted values is graphed in the line, with a 95% confidence interval based in standard error calculations surrounding that line.

The scatterplot also reveals some interesting differences from this general trend. There are a large number of low LMS use values at every grade level. At the "D" and "F" grade levels
almost every student had less than 20 assessment hits, but there are a considerable number of low values in every grade category. It appears that there may be a small positive correlation in the entire population and a small number of students in the high grade categories who use the LMS much more frequently than other students. These students do not have extreme values that should be excluded from the analysis, but their activities affect the results by increasing the correlation coefficient. This difference in use by a small number of students merits further consideration.

Figure 9. Scatterplot of Assessment Activity Hits and Course Grade

Research Question 2 results: correlation of student characteristics with final course grade.

The student characteristic variables had a weaker correlation with course grade than LMS use. As indicated in Table 21, seven out of the nine correlated variables were statistically
significant. Gender and major-college did not achieve significance at the p < .05 level. Out of the remaining variables, significance was achieved but was lower than LMS use in six of the seven variables. The effect size had a range from \( r = -.11 \) to \( r = .31 \). When these values were squared, student characteristic variables accounted for between 1% - 9% of the variation in final course grade.

The difference in effect size of these variables from the effect size of the LMS use variables was pronounced. Considered as a group for comparison purposes, the mean amount of variance explained by student characteristic variables was 4% and the mean value for LMS use was 18%. LMS use variables explained over four times the amount of variance on course grade as the student characteristic variables.

Ordering the variables by magnitude provides reveals that HS GPA had the highest coefficient (.31) and was significant at the p < .0000 level. This finding could be the result of the impact of pre-college educational experience and academic preparation. It could also be the result of the data characteristics of HS GPA. The remaining variables are all dichotomous or categorical, which makes them poorly-suited for correlation analysis.

Table 21

*Correlation results of Student Characteristics with Final Course Grade*

<table>
<thead>
<tr>
<th>Statistically Significant Student Characteristics (ordered by sign.)</th>
<th>r</th>
<th>% Variance</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS GPA</td>
<td>.31</td>
<td>9%</td>
<td>.000</td>
</tr>
<tr>
<td>URM and Pell-Eligibility Interaction</td>
<td>-.26</td>
<td>7%</td>
<td>.000</td>
</tr>
<tr>
<td>URM</td>
<td>-.21</td>
<td>4%</td>
<td>.000</td>
</tr>
<tr>
<td>Enrollment Status</td>
<td>.19</td>
<td>3%</td>
<td>.000</td>
</tr>
<tr>
<td>URM and Gender Interaction</td>
<td>-.15</td>
<td>2%</td>
<td>.003</td>
</tr>
<tr>
<td>Pell Eligible</td>
<td>-.15</td>
<td>2%</td>
<td>.005</td>
</tr>
<tr>
<td>First in Family to Attend College</td>
<td>-.11</td>
<td>1%</td>
<td>.033</td>
</tr>
</tbody>
</table>

*Mean Value All Significant Variables* 4%
A scatterplot was also plotted between HS GPA, the student characteristic coefficient with the highest magnitude, and course grade to evaluate the relationship between the variables at a more detailed level (Fig 10). The scatterplot has a moderate slope to the trend line. Unlike the LMS assessment hits, there is less of a variation in the central tendency in this data. Instead, the range of values for HS GPA increased with increased course grade. However, there are some exceptions. Students with "F" course grades had HS GPA scores ranging from 2.3-4.0. Low achieving students as a sub-population would not have been predicted by this variable. Students with "A" course grades had higher HS GPA scores than other grades. Overall, the scatterplot reveals a large variation in HS GPA by student by course grade.

Figure 10. Scatterplot of HS GPA and Course Grade.
Research question 3a results: regression of multivariate regression analysis of final grade on LMS use (restricted model 1).

The next set of analyses used multivariate regression to examine the relation between final course grade and multiple variables. The first regression was calculated with the combined LMS use variables by instructional category. The variable “total LMS hits” was excluded due to co-linearity with the other LMS variables. The results for this regression are provided in Table 22. The $R^2$ value was .25; in other words, the regression found that 25% of the variation in the final course grade could be explained by these variables. The calculations for error terms indicate that the predicted values are relatively precise. The adjusted $R^2$ value is less than .01 lower than the initial $R^2$ value, and the mean square of errors (MSE) is within a reasonable level (1.11).

The relative contribution of each variable to the regression was also calculated. The $\beta$ statistic indicates the percentage of variance in the dependent variable that is predicted for a one percent increase in the independent variable. Only two of the four LMS variables (content activities and administrative activities) were significant at the $p = .05$ level, and the $\beta$ values for each significant variable were small (.02 and .00). However, $p$ levels for variables that did not achieve significance were near the threshold level (.058 and .082).

The small value obtained for the LMS variables is most likely due to the high range of the LMS variables compared to the range of the values for course grade. For example, while content activities ranged from 0 to 57 units, final grade ranged from 0 to 4. Even if there was a strong relationship between the variables, the higher range of LMS content activity hits would lead to a relatively low $\beta$ value.
An unanticipated finding was that assessment activities and engagement activities were not significant in the regression results. These variables had the highest correlation coefficients when analyzed as isolated variables. Lack of significance indicates that these two variables did not account for additional variance to that explained by administration activities and content activities. The two significant variables (content and administration) are not necessarily more effective than the other variables, but instead may have a larger variation between students.

Table 22

*Multivariate Regression Results of Restricted Model 1: Final Grade on LMS Use*

<table>
<thead>
<tr>
<th>Combined Results</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>.25</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.24</td>
</tr>
<tr>
<td>Root MSE</td>
<td>1.11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables (Ordered by Sign.)</th>
<th>$\beta$</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Activity Hits</td>
<td>.02</td>
<td>.001</td>
</tr>
<tr>
<td>Administrative Activity Hits</td>
<td>.00</td>
<td>.019</td>
</tr>
<tr>
<td>Assessment Activity Hits</td>
<td>.01</td>
<td>.058</td>
</tr>
<tr>
<td>Engagement Activity Hits</td>
<td>.00</td>
<td>.082</td>
</tr>
<tr>
<td>Constant</td>
<td>1.23</td>
<td>.000</td>
</tr>
</tbody>
</table>

**Research question 3b results: regression of multivariate regression analysis of final grade on LMS use and student characteristics (restricted model 2).**

Student characteristic variables were added to the previous regression equation to answer research question 3b. An increase of approximately 9% in predictive relationship between the regressed variables and course grade was found compared to the regression on LMS use variables in restricted model 1. As indicated in Table 23, the adjusted $R^2$ and MSE indicate that the results are similar to the previous equation and are within acceptable ranges.

Since the independent variables used for the regression are on different scales, the $\beta$ weights cannot be directly compared. A one percent change in the HS GPA variable is not
equivalent to a one percent change in the LMS engagement activity variable; therefore, comparing the results of the $\beta$ coefficients is not valid. The significance levels and order of variables by significance indicate which variables were relevant to the regression results and provide an indication of the systematic relationship between variables.

Six of nine variables were significant in this model. HS GPA had the highest level of significance ($p < .0000$) and a $\beta$ weight of .57. The two LMS activity variables (engagement and assessment) that followed HS GPA in significance were not significant in restricted model 1, revealing salient variances change when different combinations of variables are used. Differences are not due to dissimilarity in the underlying relationship between the use of the LMS and student achievement, but they point toward a divergence in data calculation.

Table 23

Multivariate Regression Results of Restricted Model 2: Final Grade on LMS Use and Student Characteristics

<table>
<thead>
<tr>
<th>Combined Results</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>.34</td>
<td>.33</td>
<td>1.05</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables (Ordered by Sign.)</th>
<th>$\beta$</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS GPA</td>
<td>.57</td>
<td>.000</td>
</tr>
<tr>
<td>LMS Engagement Activity Hits</td>
<td>.00</td>
<td>.001</td>
</tr>
<tr>
<td>LMS Assessment Activity Hits</td>
<td>.01</td>
<td>.005</td>
</tr>
<tr>
<td>Pell-Eligible</td>
<td>-.36</td>
<td>.006</td>
</tr>
<tr>
<td>LMS Administrative Activity Hits</td>
<td>.02</td>
<td>.018</td>
</tr>
<tr>
<td>First in Family to Attend College</td>
<td>-.33</td>
<td>.025</td>
</tr>
<tr>
<td>LMS Content Activity Hits</td>
<td>.00</td>
<td>.184</td>
</tr>
<tr>
<td>URM</td>
<td>-.15</td>
<td>.291</td>
</tr>
<tr>
<td>Enrollment Status</td>
<td>-.02</td>
<td>.545</td>
</tr>
<tr>
<td>Constant</td>
<td>-.27</td>
<td>.579</td>
</tr>
</tbody>
</table>
**Research question 3c results: multivariate regression analysis of final grade on LMS use, student background characteristics and student background characteristic interactions (complete model).**

The complete model added the interaction variables of gender and URM status and Pell-eligibility and URM status to the previous restricted regression model of LMS activity and student characteristic variables. As indicated in Table 14, the complete model resulted in a very small increase in the $R^2$ value of .01 compared to restricted model 2. Both the adjusted $R^2$ value and MSE are within acceptable bounds. Accordingly, the null hypothesis is rejected, but the result demonstrates a negligible impact in practical relevance.

Only one of the two interaction variables (URM and Pell-eligibility) was significant in the regression results. Since URM and Pell-eligibility was the fifth in order of significance, it only made a small contribution to the regression model. The order of significance of the other variables in the model was identical as the previous model for the first three variables. This result shows that adding these variables had little impact on the final regression analysis.

Table 24

*Multivariate Regression results of Complete Model: Final Grade on LMS Use, Student Background Characteristics, and Student Background Characteristic Interactions*

<table>
<thead>
<tr>
<th>Combined Results</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>.35</td>
</tr>
<tr>
<td>Adjusted $R^2$</td>
<td>.33</td>
</tr>
<tr>
<td>Root MSE</td>
<td>1.04</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Variables (Ordered by Sign.)</th>
<th>$\beta$</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>HS GPA</td>
<td>.58</td>
<td>.000</td>
</tr>
<tr>
<td>Engagement Activity Hits</td>
<td>.00</td>
<td>.001</td>
</tr>
<tr>
<td>Assessment Activity Hits</td>
<td>.01</td>
<td>.011</td>
</tr>
<tr>
<td>Administrative Activity Hits</td>
<td>.02</td>
<td>.017</td>
</tr>
<tr>
<td>URM and Pell-Eligibility Interaction</td>
<td>-.60</td>
<td>.026</td>
</tr>
<tr>
<td>First in Family to Attend College</td>
<td>-.31</td>
<td>.037</td>
</tr>
<tr>
<td>LMS Content Activity Hits</td>
<td>.00</td>
<td>.130</td>
</tr>
</tbody>
</table>
Comparing the results of the regression models in Table 25 reveals a modest increase in the $R^2$ results with the addition of student characteristic variables and student characteristic variable interactions to the LMS use variables. The first restricted model accounted for 25% of the variance in the final course grade, which was increased by 10% with the addition of all student characteristic variables. Although this result is a substantial increase in the $R^2$ value, it is modest when considering that student characteristic variables are the typical predictors used of student achievement. A higher magnitude might have been predicted with the addition of these variables. The addition of interaction variables resulted in a negligible increase in the $R^2$ value, while the adjusted $R^2$ and MSE were similar between models. Overall, the regression results indicate a small effect size between the LMS use, student characteristics, and course grade.

Table 25

*Comparison of Partial and Complete Regression Model Results*

<table>
<thead>
<tr>
<th>Regression Results Comparison</th>
<th>$R^2$</th>
<th>Adjusted $R^2$</th>
<th>Root MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS Use Only (Restricted Model 1)</td>
<td>.25</td>
<td>.24</td>
<td>1.11</td>
</tr>
<tr>
<td>LMS Use and Student Characteristics (Restricted Model 2)</td>
<td>.34</td>
<td>.33</td>
<td>1.05</td>
</tr>
<tr>
<td>LMS Use, Student Characteristics and Interactions (Complete Model)</td>
<td>.35</td>
<td>.33</td>
<td>1.04</td>
</tr>
</tbody>
</table>
Research question 4 results: comparison of multivariate regression analysis of final grade on LMS use and restricted student background characteristics by population sub-samples (restricted models by sub-sample).

The complete model indicated that a negligible amount of additional variance could be explained by adding interaction variables to the regression equations. However, this finding did not indicate if the results from restricted models 1 and 2 applied equally to populations at-risk of not succeeding, as determined by the student URM status and Pell-eligibility, as they did to populations coming from backgrounds providing a greater likelihood of success in the course. This question was a driving issue behind this study, and the final research question was added to the study design to investigate this issue after the results of the complete model were calculated.

Regressions were run again by population sub-samples (URM and Pell-eligible, not URM and Pell-eligible) to determine if these models applied equally. Table 26 presents the differences between results based on these sub-samples. The restricted model 1 (LMS use variables) had an \( R^2 \) difference of .07 between at-risk and not at-risk populations. With an \( R^2 \) value of .28 for the not at-risk population, this is a 25% reduction in effect size for the at-risk students. This finding leads to potential concerns about the effectiveness of LMS use in the academic achievement of at-risk students. The smaller magnitude of the relationship indicates that at-risk students using LMS have a less systematic relationship between their use of the LMS and achievement. The use of the LMS does not lead to the same increases in grade for at-risk students as it does for the other groups.

In addition to a lower effect size, the complete regression models on combined LMS have different constant value for these populations. The at-risk students have a higher constant value (1.39) than the not at-risk student (.68). This statistic shows that the at-risk students also have a higher beginning value of LMS use that is related to achievement. At-risk students must use the
LMS more to achieve the same result and have a lower impact on achievement from using this technology.

A third restricted regression model was also run that added the five student characteristic variables that were not related to the sampling criteria. This model resulted in a small (.03) difference in $R^2$ value between at-risk and not at-risk populations. This model had a higher result with the overall study population, which was unexpected. However, this additional model was not as powerful a predictor of student success as the previous models that included the entire set of student characteristic variables.

Table 26

Comparison of Regression Models Results by Population Sub-Sample and Selected Completed Regression Results

<table>
<thead>
<tr>
<th>Regression Results Comparison by Population Subsample (R² values)</th>
<th>URM &amp; Pell-Eligible</th>
<th>Not URM &amp; Pell-Eligible</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>LMS Use Only (Restricted Model 1)</td>
<td>.21</td>
<td>.28</td>
<td>.25</td>
</tr>
<tr>
<td>LMS Use and Limited Student Characteristics (not URM or Pell-Eligible)</td>
<td>.29</td>
<td>.31</td>
<td>.33</td>
</tr>
<tr>
<td>LMS Use and Student Characteristics (Restricted Model 2)</td>
<td>N/A</td>
<td>N/A</td>
<td>.34</td>
</tr>
<tr>
<td>LMS Use, Student Characteristics and Interactions (Completed Model) (Complete Model)</td>
<td>N/A</td>
<td>N/A</td>
<td>.35</td>
</tr>
</tbody>
</table>

Regression Results: Restricted Model 1: LMS Use Only - URM and Pell-Eligible (At-Risk) Combined Results

R² .28
Adjusted R² .27
Root MSE 1.04

Variables (ordered by sign.)

<table>
<thead>
<tr>
<th>Administrative Activity Hits</th>
<th>β</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content Activity Hits</td>
<td>.01</td>
<td>.024</td>
</tr>
<tr>
<td>Assessment Activity Hits</td>
<td>.02</td>
<td>.037</td>
</tr>
<tr>
<td>Engagement Activity Hits</td>
<td>.00</td>
<td>.068</td>
</tr>
<tr>
<td>Constant</td>
<td>1.39</td>
<td>.000</td>
</tr>
</tbody>
</table>
Regression Results: Restricted Model 1: LMS Use Only - Not URM and Pell-Eligible (Not At-Risk)

Combined Results

<table>
<thead>
<tr>
<th></th>
<th>$\beta$</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R^2$</td>
<td>.21</td>
<td></td>
</tr>
<tr>
<td>Adj $R^2$</td>
<td>.17</td>
<td></td>
</tr>
<tr>
<td>Root MSE</td>
<td>1.18</td>
<td></td>
</tr>
</tbody>
</table>

Variables (Ordered by sign.)

<table>
<thead>
<tr>
<th>Variable</th>
<th>$\beta$</th>
<th>Sign.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Engagement Activity Hits</td>
<td>.01</td>
<td>.108</td>
</tr>
<tr>
<td>Administrative Activity Hits</td>
<td>.00</td>
<td>.326</td>
</tr>
<tr>
<td>Assessment Activity Hits</td>
<td>.02</td>
<td>.327</td>
</tr>
<tr>
<td>Content Activity Hits</td>
<td>.00</td>
<td>.978</td>
</tr>
<tr>
<td>Constant</td>
<td>.68</td>
<td>.030</td>
</tr>
</tbody>
</table>
CHAPTER 5
DISCUSSION

Overview

This chapter presents implications of the research findings and broader problems identified by this study, which is ultimately concerned with how to improve student achievement. The majority of the discussion focuses on issues relevant to instructional practice as well as potential use of LMS data for Learner Analytics. The findings are applicable to both instructional methods and data analysis.

Overall, the results of this study indicate that academic technologists, staff working on student persistence, and researchers can confidently use LMS log files as indicators of student educational effort. Because the LMS is used by nearly every university in the United States, the significance of this finding should not be understated. Currently, these log files are rarely analyzed. These data are an untapped resource for universities in their efforts to improve student persistence.

Further, this study found that at the course level the LMS data is better than the typical student demographic characteristics that are used as predictors for achievement. This is intuitively reasonable; how much effort a student puts into a course seems like it should have a stronger relationship with achievement than a student's background upon entering the course. However, this research empirically demonstrating this intuition and quantifies the magnitude of the difference. LMS variables were over four times as strongly related to achievement as demographic variables. LMS data has the potential to be an important variable for early-alert systems to improve student achievement and, ultimately, persistence in high enrollment courses.
The correlation results indicate that different uses of the LMS have different impacts on student achievement. Graded assessment activities had the strongest relationship with achievement, followed by content and engagement-related activities. The research also found that students used all LMS features to a similar degree. This finding indicates these categories as potentially useful for comparisons between courses.

In addition to these outcomes, the research demonstrated methods that can be used in future research with LMS data. In the original format, LMS logfiles contained a large number of entries related to server processes that were unrelated to student learning. Single student actions (such as reading a discussion post) resulted in multiple logfile entries. In the future, LMS data should be pre-processed carefully before it is used as an indicator of teaching and learning.

**Conclusions for Instructional Practices**

**Diverse student populations.**

The study population was diverse by multiple demographic data measures. Only half of the students in the course were from White ethnic backgrounds, and Latino students constituted 25% of the course population. A large number of students were from low socio-economic status backgrounds as directly indicated by 44% of the students qualifying for Pell grants and indirectly indicated by 28% of the students who were the first in their family to attend college. Considering the barriers that these students are predicted to have in higher education achievement, and especially for skills needed to succeed in online courses (Boyette, 2008; Chen, 2006; Vanderpool, 2009; Z. Zhang, 2007), these population demographics may explain the above average number of students with failing grades in this course compared to the same course that was not redesigned (taught in the traditional method).
While students with low socio-economic status may appear to be at risk for low achievement, cross-tabulations between variables demonstrated complex student identities and background experiences that were not evidenced by a single variable or interaction between variables. For example, almost 50% of the URM students came from families with a parent that had attended college, although 73% of the URM students were in Pell-eligible families. Most of the URM students were female (67%), and the literature shows that female URM students are likely to outperform their male URM counterparts. The data suggests that previous literature may overstate the challenges of acculturation and family support that these students encountered (Gloria et al., 1999; Rendon et al., 2000).

**LMS use is a systematic indicator of academic achievement.**

The correlations between student use of the LMS and academic achievement revealed a highly systematic relationship ($p < .0000$) for each of the variables. The variables explained between 12% and 23% of the variation in final course grade (research question one). This finding indicates that students who used the LMS more frequently received higher grades. The correlation methods analyzed the variables independently to determine their relationship with the final course grade.

Compared to the two traditional sections, this redesigned course had a larger number of students who failed. These particular students had a low amount of LMS activity. Therefore, the results indicate that the learning materials and activities provided through the LMS were effective for the students who used them, and the range of values in the correlation coefficients largely confirms what would be anticipated about the strength of the relationship between student participation in LMS activities and the outcome. Graded activities in the assessment category had the strongest impact on achievement, explaining 22% of the variance in grade.
Administrative activities had the lowest impact on achievement, explaining only 12% of the variance in grade.

A surprising finding was that the impact of content activities on grade was nearly identical to the impact of engagement activities on grade (r = .40 and r = .41, respectively). Students participated in engagement activities more frequently than any other category of activity. Given the large number of students in the course, and the importance of student interactions on persistence, it was anticipated that engagement activities would have a large impact on student success. Indeed, engagement was identified as an important consideration when designing the LMS to support large-enrollment course sections.

In this hybrid course, some concepts were only available through LMS materials. As a result, if a student did not use the LMS, they were at a disadvantage compared to students that did. Engagement activities may have helped to deepen student understanding of the materials and apply the course to their personal life, but this understanding may not have been revealed in course assessments, regardless of the importance in terms of the larger impact of the course. It is possible that alternative outcome variables for the course, such as future student interest in course content or ability to apply course concepts to their own life, would reveal a greater impact on learning provided by engagement activities.

The LMS total hits variable explained more variance (23%) in the correlations than any of the other variables describing use within a category. The assessment category had the highest effect size, explaining 22% of the total variance. At face value, this finding seems counterfactual. Total hits should have a value within the range of the use category coefficients, since it includes both variables with higher and lower effect sizes. Characteristics of the data in the analysis and not the underlying conceptual issues may have caused this result. The total hits
variable had the largest range of values between students than any of the other LMS category variables, which may have led to greater differentiation between students and, thus, provided a basis for stronger correlations. Limited variation in use also affected regression analysis.

Overall, the findings here on using the LMS as a systematic indicator of academic achievement are consistent with previous research in the No Significant Difference literature (Clark, 1983; Kozma, 1994a; Russell, 1997). It is not technology in isolation, but the specific instructional practices for which technology is used that make a difference in student achievement. The tools used are relevant in understanding the impact of technology, but it is the deployment of those tools within specific activities and for specific goals that determines whether they are effective or not.

**Students consistently use the LMS across categories.**

Factor analysis indicated that students use the LMS consistently across all categories of use. The first variable (administrative activities) explained 66% of the variation between students in the use of the combined LMS variables. Adding the final variable (engagement) increased the variation between students in combined LMS variable use by only 6%. When considered independently in correlation analysis, the small differences between use resulted in different findings. However, when the variables are analyzed together, these differences are greatly reduced. As a result, there were small effects on findings from the addition of LMS variables in the regression analysis compared to the correlation results of separate variables (research question 3).

Regressing the student grade by the combined LMS use category variables resulted in a nearly equivalent explanation of variance as the correlation by total hits (25% and 23%, respectively). The contributions of the variables to the regression findings were different from
the correlation results: only content and administrative activities were statistically significant (at the p < .05 level). These same variables had the lowest coefficient of determination in the correlations. It is likely that these two variables, although each having a relatively small relationship with student achievement, had the largest amount of variation in use between students and, therefore, the largest impact on the regression equation results.

**Student characteristics are low predictors of achievement.**

The correlation of student characteristics with achievement (research question 2) resulted in significant findings for seven of nine variables. The range of correlation coefficients was r = .31 to r = -.11, with two variables (gender and major-college) not achieving significance. High School GPA had the strongest magnitude and explained 9% of the total variance in grade, but this variable is of low value to help understand the causal factors behind student achievement or indicate a situation that can be affected by changes in institutional policies or classroom practices. Students who earned better grades in high school are likely to earn higher grades in college. The reasons why this population of students received high grades in this course were not revealed nor was their level of academic preparation known. Although high school GPA has the highest effect size of all student characteristic variables, it is not helpful for efforts to improve student achievement, and emphasizes the value of the findings from LMS use data.

The interaction variable of URM and Pell-eligibility had the second-largest coefficient, explaining 7% of the variance in student grade. This result confirms the findings in previous research that a large number of barriers that confront URM students also confront students from low socio-economic status families. However, an unexpected finding is the difference of this coefficient from Pell-eligibility alone. While significant, Pell-eligibility had a small negative effect size (p = -.15). As indicated in the cross-tabulation results, many Asian students were
from Pell-eligible families, which may have reduced the magnitude of this variable. The URM and Pell-eligible combination removes these students from the correlation results. URM did have a larger effect size ($r = -0.21$) that was increased by .05 through the interaction of these two variables.

First in family to attend college had a small magnitude, explaining only 1% of the variance. Gender did not yield a statistically significant result, although the level of significance was very close to the $p < .05$ threshold level for significance (.0557). This finding is consistent with previous literature and the complex relationship between gender and other variables. The interaction of URM and gender did achieve significance, although the magnitude was relatively small (explaining 2% of the variance).

Results with the major-college variable indicate that differences in major had a highly unsystematic variation with final course grade. This may have been due to the underlying relationship between major and grade. Alternatively, the analysis of majors by the college to which they belong may have distorted the relationship between major and grade.

**Combined LMS and student characteristic data makes small advance.**

Regression analyses were also conducted on combined LMS use data and student characteristics (research question 3b), as well as on combined LMS use data, student characteristics, and interactions of student characteristics (research question 3c). These regressions resulted in a small increase (+10%) in the effect size of the final regression model compared to the initial restricted model on LMS use only (research question 3a). This is a substantial increase to the effect size that advances the model to predict one-third of the student achievement in the course. These statistics reveal that student characteristics still have a
meaningful impact on student grade. A student’s background and experiences make a difference in how they approach the barriers and opportunities encountered in a course.

High school GPA made the highest contribution to the regression equations that included student characteristics. As noted previously, this contribution could be caused more by data characteristics than the underlying phenomenon. In any case, this research seeks to discover barriers to student achievement.

**LMS use data is proxy for student effort.**

Perhaps the most important conclusion from this research is that student characteristics are relatively low contributors to achievement compared to LMS use. This finding has important ramifications for the potential value of the LMS and practitioners and researchers predicting factors in student achievement.

Considered as a group, the LMS use variables explained over four times the amount of variation in student achievement as the variation explained by the student characteristics. Although the regression equations were affected by the order of entry of the variable groups (and thereby were biased toward increased results for the LMS use variables), the student characteristics provided only a modest increase in effect size upon the LMS use variables. This result implies that what students do within a course is more important to their achievement than their background upon entering higher education or their current enrollment status. The barriers (and opportunities) that the students faced in learning based on their race/ethnicity, socio-economic status, high school achievement or other factors did not impact their achievement to the degree that was affected by the frequency of LMS use. This finding leads to consideration of what deeper learning issue LMS use frequency might represent.
The low variance found between LMS use categories in the factor analysis indicates that students used the LMS consistently across all functions. Some variation in the correlation findings suggest that the various categories of use had different degrees of impact. However, considered as a whole, total use was found to be the best predictor of student achievement with a higher effect size than any individual variable. Rather than indicating specific learning materials or activities, LMS use appears to be a proxy for an underlying issue. One interpretation of this variable is that it represents the amount of student effort put into the course through the LMS (Arnold, 2010) whether by posting more frequently to the discussion board, reading more online content, watching online videos, or submitting assessments.

However, effort is a simple one-dimensional measure: students can put in “more” or “less”. The magnitude of the relationship is also small, explaining 25% of the variation in final grade. This magnitude is similar to several previous studies that investigate the relationship between frequency of LMS use and final course grade (Morris et al., 2005; Peled & Rashty, 1999; Rafaeli & Ravid, 1997). In order to increase the magnitude of these findings, future research must use other methods of analysis that advance beyond the LMS as a measure of effort. These methods might include time series analysis that analyze the LMS as a measure of study habits, content analysis that investigates the relative importance of different activities, and social learning analytics that investigates how students are interacting with others. Methods beyond frequency of use can provide more precise insights into student behavior and achievement.

**LMS use and expanded instructional method effectiveness.**

The ultimate purpose of this study was to find methods of improving student persistence in large-enrollment courses, where enrollment size prevents dynamic pedagogies and interactive
learning activities. This study indicates that the LMS is an effective institutional strategy to increase the average course grade for a large-enrollment course.

Some differences in achievement based on the type of LMS use were found through correlation analysis, but these differences appear not to be from more dynamic or engaging activities. Instead, more conventional activities, such as taking quizzes or submitting homework, had the strongest relationship with grade. Because grades are based on completing these activities, this finding is not surprising. Content activities had a slightly stronger effect than engagement activities, even though there was a large frequency of use of engagement activities. This finding suggests that student achievement, as measured by final grade, is at least as strongly affected by traditional mastery of content as by engagement of students with one another or faculty through the LMS.

**LMS use is less effective for at-risk students.**

A concerning finding in this study was a 25% reduction in effect size of the LMS use regression for at-risk students (URM and Pell-eligible) compared to not at-risk students (research question 4). Regression analysis controlled for differences in overall population averages to find the systematic relationship between each individual and the course grade. The results suggest that even if at-risk students are motivated to use the LMS, they have lower gains in achievement by using the LMS than students who are not at-risk, which supports persistence literature concepts of race and socio-economic status as conditioning factors on student achievement (Pascarella, Seifert, & Whitt, 2008; Peltier et al., 1999; Reason, 2009; Terenzini & Pascarella, 1998).

Because 80 students in the course met the at-risk criteria, sample size cannot explain these findings. The cause for this difference is not indicated by the research questions in this
study and it calls for future research by scholars interested in social justice and educational equity. This finding advances previous literature on URM populations in online learning that report lower average effects across the entire student without considering the relationship between individual student use of online materials and achievement.

**Demographic data in the Postmodern age.**

The magnitude of difference between LMS activity and student characteristics contributing to course achievement points to the need to reconsider cultural identity and social processes in the 21st century. Not only are most of the demographic data points dichotomous, with only two potential values, but they are also sampled at a single point in time. This single point is assumed to apply throughout the life of a student and to do so in a homogeneous way for all students. In the case of race/ethnicity, this variable is attributed at birth and assumed to apply consistently for all students of the same race/ethnicity. No consideration is given to the importance of racial identity to the individual student or their family. While Pell-eligibility is based on student financial status at application to college, it does not indicate whether the student has lived in a low-income family for their entire life or just recently become low-income due to temporary circumstances. Gender, likewise, is represented by a single value, but it is a complex identity that develops over time.

Generalizations based on these variables may have been true at one point in history, but in the contemporary historical period, social identity is less cohesive and stable. Postmodern social theory, which emerged in the 1970's, emphasizes flux, change, and hybridity in cultural identities and social forms (Lyotard, 1979). Postmodernism proposes that once-static "master narratives" such as race/ethnicity, or any other theoretical construct, are no longer cohesive in the
information age. Taken to its logical conclusion, Postmodernism suggests that there is no objective knowledge or social structures that are meaningful in the present historical era.

More meaningful student characteristics having a stronger relationship with student achievement may require incorporating insights from Postmodern theory. Moving beyond interactions of static point-in-time variables, characteristics could be considered as dynamic processes that need to be understood through a student’s interpretation and are subject to change over time. Instead of gender, perhaps self-perceptions of masculinity or femininity could be surveyed. Questions of race could be supplemented with questions to students about the importance of race in their personal identity. If the goal of research is to predict, and ultimately affect the local context, then research should use measures that are accurate and relevant to students.

The study’s findings confirm previous research on student persistence and the recommendations of others to use multiple demographic criteria and conduct research focused on the complexity of the contemporary student population (Pascarella & Terenzini, 2005; Reason, 2003). The high numbers of URM students in the course validates the increasing diversity of students found in the national college enrollment statistics.

Conclusions about Data for Learning Analytics

**Filtering LMS log file for educationally relevant activity.**

LMS use data is relatively easy to acquire from log files, but it requires processing in order to be used for educational research. Programmers created the LMS log file to record the operations of web servers to troubleshoot technical problems. The log file describes what the server does, not what students and faculty do. Transforming the log file into a meaningful
representation of learning activities requires data filtering and preliminary analysis before any conclusions should be drawn about the meaning of the log file.

This study found that the log file contained more entries for technical operations than entries that are meaningful to learning activities. Filtering for meaningful activities resulted in a 74% reduction in the amount of log file data. An effective data analysis approach was filtering log file entries by both the amount of time spent on activities and the descriptions of activities. This method, combined with aggregating the data into categories, provides an advance over previous research methods that hand-selected LMS actions relevant to a course (Macfadyen & Dawson, 2010; Morris et al., 2005). The method used for this study can be applied to any course using the LMS to provide a data set that more accurately describes student activity.

The implications of filtering apply to all campuses reporting LMS use data. Like many other campuses, prior to this research, Chico State aggregated unfiltered LMS log file data to produce reports on use of the LMS for internal department use and to represent the activities of the department to campus administration. The data was summarized into measures of central tendency (e.g., mean, max, min) with a focus on fundamental indicators of adoption (e.g., number of course sections using the LMS, number of faculty using the LMS). Without a preliminary filtering process, these reports were summarizing data about server actions that would result in an inaccurate representation of student LMS use. It is critically important that any LMS reporting conduct preliminary analysis to ensure that the data used is accurate.

Each LMS platform and version will ultimately have a unique filtering process. Although the log files have the same elements, they record server actions differently. Future research could apply the general processes from this study but would require re-analysis and
inspection of the results on the new log file. A different degree of data reduction may result, which would be interesting.

Data range and limits of dichotomous and categorical variables.

This research shows that the range of data values for independent variables highly influences the results. High school GPA was the only continuous student characteristic variable, and similar to the continuous LMS use variables, there was a large range of potential data values to analyze. The other student characteristic variables were either dichotomous (with only two potential values) or categorical (with a small number of potential values). Inferential statistical analysis methods calculate the relationship between changes in the independent variable and the dependent variable. If there is a larger range of data values to be analyzed, there is more likelihood of detecting subtle relationships that are not present in a highly restricted data value range.

Given the complexity of achievement at the course level, a larger range of values is more likely to discover significant relationships with final grade. Many student characteristics are dichotomous or categorical in nature and not appropriate to a broader array of data values. However, the range of data should be considered, especially in the course context, where there are likely to be subtle relationships between these characteristics and student achievement.

Contributions to Previous Learner Analytic Research Studies

As discussed in the literature review, four previous studies investigated the relationship between frequency of LMS use and course grade (J. P. Campbell, 2007; Macfadyen & Dawson, 2010; Morris et al., 2005; Rafaeli & Ravid, 1997). The results of the complete regression model with combined LMS use variables (R² = .25) was nearly the same as the other hybrid course (Rafaeli & Ravid, R² = .24), but lower than the two studies of fully online courses (Morris et. all,
Although this is a small number of studies from which limited conclusions can be drawn, the similarity of the hybrid and online course findings is striking. These results suggest that LMS use is more predictive of students success in the fully online environment than in a hybrid environment, in which some learning activities are conducted outside of the LMS. Thus, as common sense would dictate, the more that the LMS is required for the course, the greater degree to which student achievement is related to frequency of student LMS use.

The complete model with student characteristics and interaction variables ($R^2 = .35$) was slightly larger than any of the previous research with LMS use data alone. However, it was smaller than the results of Campbell's study ($R^2 = .51$), which combined LMS and student characteristic data (2007). The results may be due, at least in part, to the scale of Campbell's study context; the study population was the entire student population of Purdue university for one semester ($n = 26,014$), and (30) variables were included in the regression equation. Large sample sizes and numbers of variables, such as in the Campbell study, provide a much greater potential to uncover systematic relationships than is possible in a course with less than 400 students.

Campbell found that the LMS had a negligible impact on prediction of student achievement to that provided by student characteristic variables, with correlations of individual LMS use variables below $p = .1$ and an increase in the overall $R^2$ value of less than .01%. In contrast, the study described here has correlations as high as $p = .47$ and an $R^2$ value for LMS variables of .25, with a .10 increase with the addition of student characteristic variables to the regression model. The difference between this study and Campbell’s is most likely due to the number of course sections included in the study.
Methodological Limitations and Potential Revisions

The LMS independent variables used in this study describe the frequency of student use. While these variables describe what the students did on the LMS, they do not indicate the reason(s) behind their use. The conclusion that frequency is a proxy for effort could obscure a deeper learning phenomenon. Increased use of the LMS may actually reflect the effectiveness of online learning modality or materials in this course for students who used them frequently. It is also possible that students who did not use the materials found them an ineffective means to learn, were not appropriately prepared to use them, did not have access to the technology other students had, or some other factor besides their motivation to work on the course. Discovering the cause behind frequency of LMS use would require additional research that incorporates different data sources and analyses.

This additional research could be conducted through the additional of qualitative research methods. Individual interviews or focus groups could be conducted with students based on selected criteria, such as the at-risk student characteristic. These students could be asked about the factors that supported or hindered their use of the LMS and how effective they found the features to help them learn course content. Students could also be grouped by final grade. These methods would be helpful to uncover potential deeper factors that could help in the design of LMS course materials.

Implications for Practitioners and Suggestions for Future Research

The findings and conclusions drawn from this research have implications for practitioners seeking to improve student persistence and for scholars conducting research in learning analytics and other fields. The following recommendations summarize issues discussed throughout this monograph.
Implications for practitioners.

1. Total frequency of LMS use can be used as a meaningful predictor of student achievement at the course level. This variable provides an excellent starting point for early warning systems and predictive models to improve student achievement.

2. LMS use data is potentially more valuable than conventional student demographic variables. This data is also easier to access for academic technologists and faculty who may not have access to the complete student information system.

3. LMS use data requires filtering before accurate analysis can be conducted. The purpose of filtering is to separate meaningful educational actions from technical server processes. Dwell time and LMS action are useful criteria to use in filtering and must be customized for each LMS platform due to differences in log file recording.

Suggestions for future research.

1. Qualitative research could be added to a similar study to explore the reasons underlying the differences in use, especially by under-represented minority or at-risk students.

2. This study is limited to a single redesigned hybrid course section. Repeating this study with a different course, especially a hybrid course that was adopted without a thorough redesign, would provide an interesting comparison.

3. Categories of LMS use (administration, assessment, content, engagement) may be useful categories to compare LMS activity at the course level. However, students have low variation in frequency by type of LMS use, which reduces the value of these categories to understand student achievement.
4. New behavioral-based methods could be developed to increase the effect size. These methods should be based on known learning activities that have greater variation between students, such as activity patterns, text mining, discourse analysis, social network analysis, or others.

5. New content-based methods could be developed to increase the effect size and compare to behavioral methods. These new methods could be based on the importance of an activity (e.g., identifying core content for a course) or particular set of activities (e.g., submitting an important assessment, posting a message on an important topic).

6. The reasons why at-risk students had a lower effect size for LMS use than students who are not at-risk could be further explored. This is an important issue for educational equity and could be studied through quantitative analysis of LMS use or qualitative analysis.

7. Student characteristics could be aggregated into broader profiles with different pertinent variables, which may reduce problems of missing data encountered in this research and identify variables relevant for different populations. Profiles could be run as different regression equations or compared through other methods, such as those beyond single demographic variables. Interactions of variables are clearly called for to learn about the importance of student characteristics at the course level.
References

ACT. (2004). Crisis at the core: Preparing all students for college and work. Iowa City, IA: Author.


California State University Analytic Studies. (2008). Statistical Brief on Graduation Rates for CSU Full-Time, First-Time Freshmen By Gender and Ethnicity. Long Beach: California State University Office of the Chancellor


Done, K., Sr. (2009). *Factors in the implementation of online courses of study in selected community colleges in Mississippi*. (Ph.D. 3352272), Mississippi State University, United States -- Mississippi. Retrieved from http://0-


Russell, Thomas L. (2001). *The no significant difference phenomenon: a comparative research annotated bibliography on technology for distance education: as reported in 355 research reports, summaries and papers*. [S.l.]: IDECC.


Tobias, Sheila. (1990). *They're not dumb, they're different : stalking the second tier / by Sheila Tobias:* Research Corp.


