2013 IEEE 13th International Conference on Advanced Learning Technologies

Correlation between Course Tracking Variables and Academic Performance in Blended Online Courses

Che-Cheng Lin
Graduate Institute of Information and Computer Education
NTNU
Taipei, Taiwan
80108001E@ntnu.edu.tw

Chiung-Hui Chiu
Graduate Institute of Information and Computer Education
NTNU
Taipei, Taiwan
cchui@ntnu.edu.tw

Abstract—**The purpose of this research was to identify which course tracking variables correlate significantly with academic performance in blended asynchronous online courses through an empirical analysis of Learning Management System (LMS) data. In this study, course tracking variables refers to number of online sessions, number of original posts created, number of follow-up posts created, number of content pages viewed and number of posts read. Academic performance defined as how well a student's final grad is. These five variables were collected from 15 undergraduate courses in the first semester of academic year 2012 at one national university in Taiwan. A total of 528 related final scores were transformed to z score and analyzed to investigate the correlation between course tracking variables and academic performance. A multiple regression analysis was used to evaluate how well course tracking variables measure predicted academic performance. Results indicated that approximately 16.4% of the variability in academic performance was accounted for by student's course tracking variables measure, and three of the five variables were statistically significant.**

Keywords-Academic performance; academic analytics; data mining; ICT

I. INTRODUCTION

The vast majority of Taiwan universities have adopted learning management systems (LMS) to support students learning for years. At one national university in Taiwan for instance, approximately 10,000 students and 3,000 staff use the commercial LMS Wisdom Master 3 (WM3) as part of their daily course experience. Increasingly, these systems provide an essential infrastructure which mediates students' access to learning resources and facilitates student-student and student-lecturer interaction. Additionally, these systems can provide sophisticated levels of institutional-wide data on areas of student demographics, academic performance, learning pathways, user engagement, online behavior, and development and participation within social networks. These data can also be used to promote practitioner reflection for professional development as well as identifying students who may require additional scaffolding and/or early learning support.

According to the Sloan Consortium, over 1.9 million students took at least one online course during Fall 2003, and 81% of all U.S. institutions of higher education offered at least one fully online or blended course [1]. Considering the rapid growth in online education, a key concern of educators and the public is the quality and effectiveness of online learning [2], [3]; consequently, recent studies have emphasized the importance of evaluating student behavior online to better understand student learning and achievement in the virtual environment [2], [4]. In addition, previous studies have evaluated students' engagement in online courses through an empirical analysis of student learning behavior and its relationship to persistence and achievement. However, there is rare research focusing on blended online course through an empirical analysis of LMS data.

The campus-based classroom, where students and instructors meet face-to-face (F2F), is physically and psychologically different to the virtual classroom. LMS appears more suited to giving and storing information. F2F is for responding to or discussing information. As these tasks are complementary, the optimum solution would appear to use both LMS and F2F. This pedagogic combination of the virtual and the real classroom is called blended learning and is advocated by an increasing number of authors [5], [6]. The goal of the present study was to examine blended course tracking variables through the analysis of LMS data. Specifically, this research addressed:

- Which course tracking variables of LMS data correlate to academic performance in blended online courses?
- How accurately can measures of significant course tracking variables of LMS data predict academic performance in blended courses?

II. LITERATURE REVIEW

A. Higher education institutions (HEIs) *Integrated Internet and communication technology (ICT) into teaching and learning*

Higher education institutions (HEIs) integrated ICT into teaching and learning rapidly in the past decade. The broad number of ICT tools now readily available within HEIs presents numerous pedagogical advantages to both educators and learners. However, a simple combination of hardware and software will not make integration naturally follow [7]. In addition, Integration of ICTs into teaching and learning has been driven by both pedagogical goals and the need for enhanced flexibility of content delivery and engagement with course materials. Learning management system (LMS) seems to meet the criterion. A LMS is an information system

978-0-7695-5009-1/13 $26.00 © 2013 IEEE
DOI 10.1109/ICALT.2013.57

Authorized licensed use limited to: University of Eastern Finland. Downloaded on September 05,2024 at 07:18:13 UTC from IEEE Xplore. Restrictions apply.

that facilitates e-Learning. HEIs are increasingly adopting LMS and consequently the analytics of LMS data are coming more to the fore.

With current LMSs ranking in the top 10 technologies for higher education [8], one of the most significant developments in the use of information technology (IT) in universities in the last decade has been the adoption of LMSs to support the teaching and learning process [9]. LMSs process, store and disseminate educational material and support administration and communication associated with teaching and learning. Most LMSs are the primary e-learning resources that bring together tools and materials to support learning. However, despite this ubiquity of LMS use, there has not been widespread change in pedagogic practice to take advantage of the functionality afforded by LMSs [10], [11]. Consistent with this, there has been very little analysis of the impact of LMSs on teaching or learning [9]. In order to take advantage of the potential associated with LMSs, research that addresses the role of LMSs in learning performance is needed. Through the analysis of LMS data, this study addressed the question of which tracking variables correlate significantly with learning performance.

B. *Mining LMS data*

LMSs capture and store large amounts of sophisticated user activity and interaction data. Course tracking variables include measures such as the number and duration of online sessions (student visits to the online course site), messages read or posted, and content pages visited. Perhaps the most promising source of automatically gathered online learning data is the learning software itself, particularly the LMS. Since students typically log in to such systems, keeping track of users and sessions–a major hurdle in examining server logs [12]–is done automatically. In addition, many such systems gather a range of relatively high-level student data such as quiz grades and forum posts [13]. Importantly, this data may represent aspects of learner behavior that are difficult or impossible to apprehend by other means: study patterns, engagement, and degree and mode of participation in learning networks [14].

To date, some investigators have been able to access, analyze, visualize and interpret this data, and current LMSs offer some data reporting options. Recent studies have developed or identified the potential for LMS tracking data variables. Reference [15] examined student engagement in totally asynchronous online courses through an empirical analysis of student behavior online and its relationship to persistence and achievement. They found approximately 31% of the variability in achievement was accounted for by student participation measures, and three of the eight variables were statistically significant. Reference [16] revealed that cumulative course data logs are predictive of both a student's sense of connectedness and student community in online courses. Reference [17] affirms that pedagogically meaningful information can be extracted from LMS-generated student tracking data, and discusses how these findings are informing the development of a customizable dashboard-like reporting tool for educators that will extract and visualize real-time data on student engagement and likelihood of success.

However, Most of the existing studies track LMS data in fully online courses. A recent survey of United States (US) higher education institutions indicated a greater than 70% adoption rate of campus-wide LMS (Campus Computing Project, 2008). 2007 figures for the US suggest that greater than 3.9 million students (>20% of all US higher education students) enrolled in at least one online course during the Fall 2007 term [1]. Even more significant is the uptake of LMS to support ‘traditional’ classroom-based courses, in formats better described as ‘blended’ or ‘web-supported’. LMS is a heavily utilized tool and a critical resource for supporting teaching and learning across the institution. Moreover, very little research exists, and no guidance is available for educators, to indicate which (if any) of the captured tracking variables may be pedagogically meaningful – that is to say, which of the many available data points are indicative of student participation in educationally purposeful activity that may contribute to their learning and performance in a blended online course.

C. *Academic analytics*

Initially developed in the corporate sector as ‘business intelligence’, [18] have called the application of these analytical tools and processes to educational systems academic analytics. In essence, academic analytics involves the extraction of large volumes of data from institutional databases and the application of various statistical techniques in order to identify patterns and correlations [19]. The concept of ‘academic analytics’ is now gaining increasing momentum as a process for providing HEIs with the data necessary to respond to the reportage and decision making challenges facing contemporary universities [15]. For example, using analyses on student demographics, student learning styles, course communication and external factors, [20] suggested that Web-based courses are more attractive to busy students who are also more likely to fail or drop the course. In addition, [19] have proposed that the analysis of data captured from various IT systems could be used to inform decision making process for university management and administration. The new millennium has, however, seen the emergence of a new approach to system-wide data harvesting and analysis that has the potential to unlock the value of the vast data sets captured by institutional LMS. Reference [21] outlined how academic analytics can be used to better inform institutions about their students learning support needs. They also provided examples of IT automation that may allow for student user-information to be translated into a personalized and semi-automated support system for students.

The widespread adoption of ICT across the HEIs has provided institutions with additional expansive data sets that capture student learning behaviors through user online interactions. However, there has been limited interest in analytics within the academy [18]. The few projects undertaken within the higher education sector have primarily focused on the analysis of institutionally-collected data on

Authorized licensed use limited to: University of Eastern Finland. Downloaded on September 05,2024 at 07:18:13 UTC from IEEE Xplore. Restrictions apply.

student demographics and overall academic performance, with the goal of understanding and improving institutional recruitment and retention [17]. In addition, most of the existing studies analyzed only in fully online courses, regardless of the majority fraction of blended online courses. Through an empirical analysis of LMS data, this study aimed to identify which course tracking variables correlate significantly with academic performance in blended asynchronous online courses.

III. Methodology

Initial exploratory research was undertaken to identify the course tracking variables that would inform the development of an academic analytics tool for institutions. This involved the mining of five LMS course tracking variables of the commercial LMS Wisdom Master 3 (WM3) at one national university in Taiwan.

A. Study population and context

This paper reports on the analysis of the commercial LMS tracking data from one semester of fifteen blended courses offered by one national university in Taiwan from Feb., 2012 to July, 2012. A 'blended' course is defined as one in which partial content delivery, communication and assessment is carried out via the institutional LMS. In this study, a blended course is specifically defined as at least half of course hours were conducted with online. This comprehensive blended course makes extensive use of available tools to provide access to course content (content pages, learning modules, web links, self-assessment quizzes), promote student communication and engagement (discussion forums, chat), assess student learning (quizzes, assignments), and allow instructors and students to self-manage their learning (mail, calendar, grades, progress, announcements). In summary, there are 15 blended courses, including 528 related final scores. A total of 528 students participated in this study. They were primarily freshmen aged 18-20 years old, with a gender balance of males to females of 15:19. The 15 blended courses were equally spread around different colleges.

B. Data collection and procedures

The data analyzed in this exploratory research was extracted from the course-based tracking variables. Data in course tracking variables were stored in the selected commercial LMS database. The selected commercial LMS database had ten related tables which stored course tracking data for this study. The interface of database query tool such as phpMyAdmin provides SQL (Structured Query Language) to insert, query, update and delete course tracking variables data which stored in different tables. Using the phpMyAdmin, data were extracted a greater set of course tracking data and exported to CSV files. The procedures of finding out specific data were using SQL to query with different tables step by step under some conditions.

Tables in a database were related to each other with keys. Through the JOIN function were used in an SQL statement to query data from two or more tables, based on a relationship between certain columns in these tables. Data collected on each student included 'whole semester' counts for frequency of usage of course materials, engagement and discussion. In some cases, variables were combined to give a more accurate and complete measure of post read. For instance, counts per student for 'discussing board post read' and 'chat room post read' were aggregated to provide a complete score for student read of post. Table 1 shows the initial set of course tracking variables examined in this study for relationship with student performance in the course. Data was exported into an Excel spreadsheet and merged with final course grade data received from institution administration database. The function of Excel was used to standardize each final score data as z scores. This complete student data set was later imported into SPSS for multiple regression analysis. Descriptive data for the initial set of course tracking variables are presented in Table 2.

TABLE1. Course tracking variables for further analysis.

Total number of online sessions
Total number of original posts created
Total number of follow-up posts created
Total number of posts read
Total number of content pages viewed

TABLE2. Population descriptive indicators for the courses under study

Variable	N	Min.	Max.	Mean	SD
Final z score	528	-4.05	2.35	0.00	1.000
# of online session	528	10	506	90.57	72.125
# of original posts created	528	0	61	6.25	9.925
# of follow-up posts created	528	0	207	17.71	26.292
# of posts read	528	0	224	32.05	45.475
# of content viewed	528	0	657	108.48	96.875

IV. Results

A. Simple (bivariate) correlations of LMS tracking variables with final grade

To further investigate the significance of selected variables as indicators of student performance in this study, a simple correlation analysis of each variable with student final grade was undertaken. Of the selected five course tracking variables examined, all of them demonstrate a positive and statistically significant correlation with student final grade ($p < 0.01$) (Table 3). Within the significant subset of the LMS variables, one demonstrate a medium-large effect size ($r = 0.386$), with explaining 15% of the variance in student final grade. The remaining 4 variables have a small-medium effect size ($r = 0.146–0.278$), with each explaining from 2% to 8% of variance in student final grade.

TABLE3. Simple correlation of relevant LMS course tracking variables with student final grade

Variable	r_s	r^2	p
# of online session	0.386*	0.149	0.000
# of original posts created	0.152*	0.023	0.001
# of follow-up posts created	0.199*	0.040	0.000
# of posts read	0.146*	0.021	0.001
# of content viewed	0.278*	0.077	0.000

Authorized licensed use limited to: University of Eastern Finland. Downloaded on September 05,2024 at 07:18:13 UTC from IEEE Xplore. Restrictions apply.

* Correlation is significant at the 0.01 level (2-tailed).

B. *Multiple regression*

Although five LMS variables for this study appeared to show significant correlation with student final grade, it would be erroneous to rely too heavily on the predictive power of simple correlations. Students do not show simple univariate patterns of blended online behavior within course websites, but instead undertake complex composite behaviors in which they make decisions to give more or less time to different activities. Some combinations of online activities are likely to translate into effective learning strategies but 'more time spent on online activities' does not simply predict higher performance. In other words, some students are making more effective strategic decisions about time use within the virtual classroom that is not adequately represented by simple correlations with time online.

From the set of significantly correlated course tracking variables (Table 3), five potentially significant indicator variables were identified for inclusion in a multiple regression analysis. The previously identified predictors are entered into the model in order of their importance in determining the outcome variable [22]. However, in the absence of such information, a stepwise approach for entering potentially significant variables into a model is a robust and valid approach. In these instances a forward or backward stepwise regression analysis is undertaken. However, a forward regression does present a higher risk in terms of excluding potential predictors that are involved in suppressor effects and thus may generate Type II errors (that is, accidentally excluding significant predictors) [22]. Consequently, this study elected to adopt a backwards stepwise method, in which variables that are not statistically significant in relation to the predictive power of the model are removed.

A linear multiple regression analysis was therefore conducted, in order to develop a predictive model in which 'Student final grade' was the continuous dependent variable. As shown in Table 4, this process generated a 'best predictive model' of student final grade (F (32.900), p = 0.00) as a linear combination of the LMS tracking data variables measuring only three online activities: total number of online sessions, total number of follow-up posts created, and total number of posts read. All three variables are statistically significant contributors ($p < 0.05$). The multiple squared correlation coefficient for this model is 0.164, indicating that some 16.4% of the variability in student performance in this study can be explained by this combination of student online activities within the course site.

TABLE4. MULTIPLE REGRESSION ANALYSIS SUMMARY FOR LMS COURSE TRACKING VARIABLES (N = 528)

Variable	Unstandardized coefficients		Standardized coefficient
	β	*Std. Error*	β
(Constant)	-0.414	0.063	
# of online sessions	0.006	0.001	0.414
# of follow-up posts created	0.005	0.002	0.132
# of posts read	-0.003	0.001	-0.152

$r = 0.405$, $r^2 = 0.164$.
$p < 0.05$.

V. DISCUSSION AND CONCLUSIONS

A. *LMS data variables are significant indicators of student performance*

This study was conducted to analyze the five course tracking variables mined from LMS data in fifteen blended online courses over one semester. There were two research questions. The first was which course tracking variables of LMS data correlate to academic performance in blended online courses? A simple correlation analysis of variables indicated that the five variables were all correlated with student final grade which defined as student performance. The number of online session variable explained 15% of the variation in student final grade. Although the other four variables resulting the correlation were significant, but the variation was lower than 8%, from 2% to 8%. It meant that the major factor of influencing student performance in blended online course was the student total number of online session.

The second research question was how accurately can measures of significant course tracking variables of LMS data predict academic performance in blended courses? Our findings indicated that a regression model of student performance, developed using course tracking variables relevant to the measures of time on task (variables indicating number of online session, number of follow-up posts created and number of posts read) explained 16.4% of the variation in student final grade. This regression model suggested that three measures of student online activity (number of online session, number of follow-up posts created and number of posts read) function as significant predictive variables of student final grade. However, while past studies have focused on 'time online' as a measure of student engagement and effort, our current investigation in blended online course indicated that number of post created and number of content viewed were not significant predictor variables of student final grade in regression model. The single most significant predictive variable reported here, with regression coefficient (β) of 0.414 is the total number of student online session. This result support and confirm that the degree of student engagement in an online course is an important indicator of academic performance. The other interesting findings were the regression coefficient of the other two significant predictive variables. The first is the total number of follow-up posts created, with only regression coefficient (β) of 0.132. One explanation is that the student's interaction in blended online course not only happened in virtual classroom but also in F2F classroom. Another alternative is that neither the total number of original post created nor the total number of follow-up post created had any impact. One possible reason for the lack of the effect could be cultural. In Taiwan cultural context, students in blended course are used to interact with

Authorized licensed use limited to: University of Eastern Finland. Downloaded on September 05,2024 at 07:18:13 UTC from IEEE Xplore. Restrictions apply.

classmate in F2F classroom rather than in virtual classroom. The second interesting finding is the total number of post read indicated negative predictive variable, with regression coefficient (β) of -0.152. There are several possibilities. The first of these is possibly the nature of student behavior in blended online course. It seems that students read more posts were less aggressive create posts themselves. A second possibility is probably that student behave less confident when read posts more than create posts in Taiwan e-Learning context. A third possibility is that the contribution of statistic significant was just too small.

To sum up, simple bivariate regression indicated that all five selected course tracking variables correlated with the student final score which refer to academic performance. Multiple regression analysis, however, indicated that only course tracking variable refer to total number of online session can explain the adequate variation of academic performance. The rest four course tracking variables either show no impact on final score, or with very little variation of academic performance. We concluded that in blended online course, the major factor influencing the student academic performance was the course tracking variable refers to total number of online session.

B. *Limitations of this study*

There are two limitations that impact the overall generalizability and interpretation of the findings of this study. First, the implications of the study are limited by its focus on data derived from blended online courses within one institution and one LMS. Second, in blended online courses, it is reasonable to expect that the proportion of F2F interaction will disturb the impact of interaction in virtual classroom. This suggests that indicators of student online activity maybe be over-estimated. Future studies should be directed towards the investigation and analysis of institutional-wide course tracking indicators, including virtual and F2F classroom, measure in relation to student academic performance for alternate pedagogical designs and course delivery modalities.

REFERENCES

[1] Allen, I., & Seaman, J. (2008). Entering the mainstream: The quality and extent of online education in the United States, 2003–2008. The Sloan Consortium. Available online at http://www.sloan-c.org/resources/entering_mainstream.pdf

[2] Nachmias, R. (2002). Research framework for the study of a campus-wide web-based academic instruction project. Internet and Higher Education, 5, 213–229

[3] Peled, A., & Rashty, D. (1999). Logging for success: Advancing the use of WWW logs to improve computer mediated distance learning. Journal of Educational Computing Research, 21, 413–431.

[4] Zaiane, O. R., & Luo, J. (2001). Towards evaluating learners' behavior in a web-based distance learning environment. Available online at http://www.cs.ualberta.ca/~zaiane/postscript/icalt.pdf

[5] Hakkarainen, K., & Palonen, T. (2003). Patterns of female and male students' participation in peer interaction in computer-supported learning. Computers & Education, 40(4), 327–342. http://dx.doi.org/10.1016/S0360-1315(02)00134-3.

[6] Lee, D. (2003). New technologies in the politics classroom: using internet classrooms to support teaching and learning. Politics, 23(1), 66–73. http://dx.doi.org/10.1111/1467-9256.00181.

[7] Earle, R.S. (2002). The integration of instructional technology into public education: Promises and challenges. Educational Technology, 42(1), 5–13.

[8] Yanosky, R., Harris, M., & Zastrocky, M. (2004). Top technology priorities for higher-education e-learning. Stamford, CT: Gartner Research Group. <http://www.mtholyoke.edu/lits/ris/CourseMgmt/Gartner/top_tech_priority-g.pdf>.

[9] Coates, H., James, R., & Baldwin, G. (2005). A critical examination of the effects of learning management systems on university teaching and learning. Tertiary Education and Management, 11(19–36).

[10] Becker, R., & Jokivirta, L., (2007). Online learning in universities: Selected data from the 2006 Observatory survey - November 2007. The observatory on borderless higher education (OBHE). [Electronic Version] <http://www.obhe.ac.uk> Retrieved 21.09.07.

[11] Collis, B., & van der Wende, M., (2002). Models of technology and change in higher education: An international comparative survey on the current and future use of ICT in higher education. [Electronic Version] <http://www.utwente.nl/cheps/documenten/ictrapport.pdf> Retrieved 11.02.08.

[12] Zorrilla, M. E., Menasalvas, E., Marin, D., Mora, E., & Segovia, J. (2005). Web usage mining project for improving web-based learning sites. Computer Aided Systems Theory Eurocast 2005: 10th International Conference on Computer Aided Systems Theory, Las Palmas de Gran Canaria, Spain, Revised Selected Papers.

[13] Mazza, R., & Milani, C. (2005). Exploring usage analysis in learning systems: Gaining insights from visualisations. 12th International Conference on Artificial Intelligence in Education (AIED 2005). The Netherlands: Amsterdam.

[14] Dawson, S., McWilliam, E., & Tan, J. P.-L. (2008). Teaching Smarter: How mining ICT data can inform and improve learning and teaching practice. Paper presented at the ASCILITE 2008, Melbourne. Australia. <http://www.ascilite.org.au/conferences/melbourne08/procs/dawson.pdf>.

[15] Morris, V. L., Finnegan, C., Wu, S.-S. (2005). Tracking student behavior, persistence, and achievement in online courses, Internet and Higher Education, 8, 221–231.

[16] Black, W. E., Dawson, K., Priem, J. (2008). Data for free: Using LMS activity logs to measure community in online courses, Internet and Higher Education, 11, 65–70.

[17] Macfadyen, P. L., & Dawson, S. (2010). Mining LMS data to develop an ''early warning system" for educators: A proof of concept. Computer & Education, 54(2), 588–599.

[18] Goldstein, P. J., & Katz, R. N. (2005). Academic analytics: The uses of management information and technology in higher education. Washington, DC: EDUCAUSE Center for Applied Research.

[19] Campbell, J., De Blois, P. B., & Oblinger, D. (2007). Academic analytics: A new tool for a new era. EDUCAUSE Review, 42(4), 42-57.

[20] Doherty, W. (2006). An analysis of multiple factors affecting retention in web-based community college courses. Internet and Higher Education, 9(4), 245-255.

[21] Dawson, S., Heathcote, L., Poole, G. (2010). Harnessing ICT potential: The adoption and analysis of ICT systems for enhancing the student learning experience. The International Journal of Educational Management, 24(2), 116-128.

[22] Field, A. (2005). Discovering statistics using SPSS. London: Sag

Authorized licensed use limited to: University of Eastern Finland. Downloaded on September 05,2024 at 07:18:13 UTC from IEEE Xplore. Restrictions apply.