Predicting Academic Performance in an Agricultural Statistics Course using Learning Analytics

Shuai Ma, *Zhihong Xu, Zilong Pan, Wenting Weng, Theresa Pesl Murphrey

TEXAS A&M UNIVERSITY

Introduction

Learning analytics has proven to be beneficial as it enables prediction of students' achievements, provides explanations for student outcomes, and identifies at-risk learners. Extensive research has highlighted the positive impact of learning analytics on enhancing students' learning outcomes and contributing to their success (Arnold & Pistilli, 2012). With the widespread use of Learning Management Systems (LMS) in higher education, log data generated by an LMS provides valuable information on student learning and engagement patterns. Building upon these widely recognized classifications in the online learning environment (Moore, 1989, Hillman et al., 1994), our study focuses on analyzing online interaction types using Canvas log data.

This study introduces specific instructional design approaches for teaching statistics and data visualization techniques. It focuses on implementing a problem-based learning (PBL) strategy in an agricultural statistics course over a semester with ten graduate students participating. Additional instructional design strategies applied include simulation enhancement, project-based learning, and a flipped classroom.

Research Questions

The study aims to answer three research questions:
1) What were the interaction types observed within the course?
2) What were the student interaction patterns throughout the semester?
3) What were the significant predictors of students' academic performance, including the potential role of online interactions?

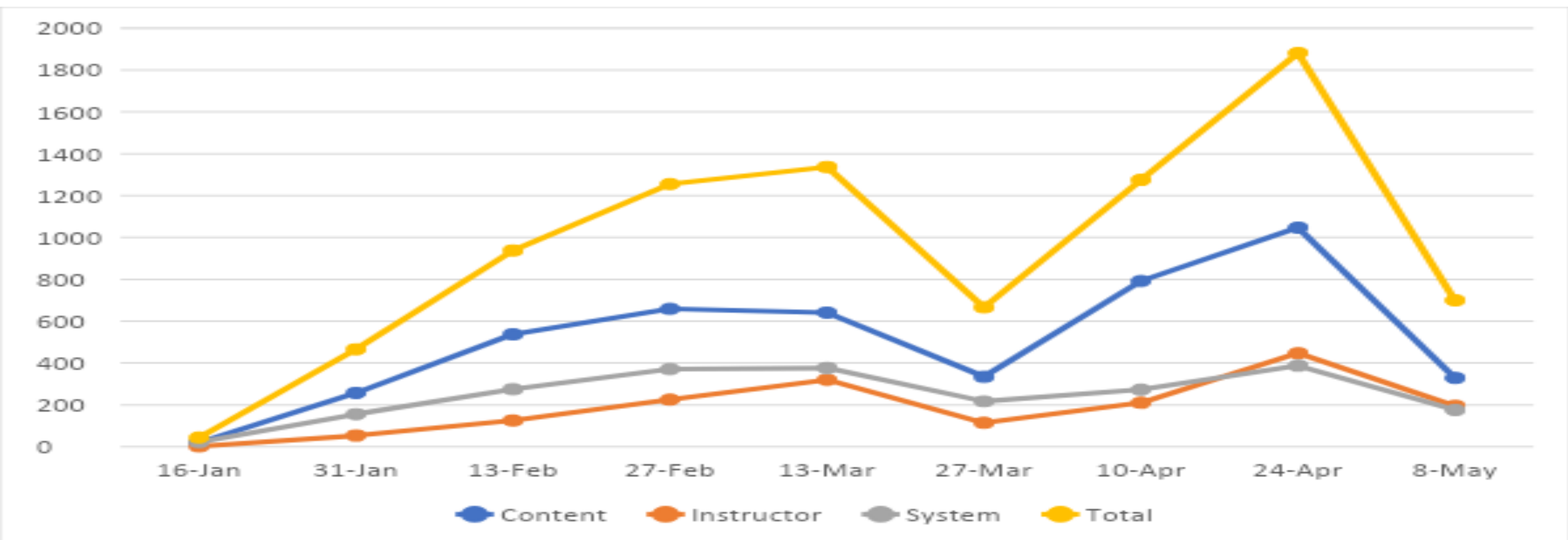
Methods

- Pre- and Post- surveys: self-efficacy, motivation, cognitive, social and teaching presence.
- Academic performance: final scores.
- Log data : Canvas. Log data were categorized as three interaction types: interaction with the instructor, interaction with the content and interaction with the system system (Moore, 1989; Hillman et al., 1994).
- Analytic tool: software R version 4.3.0 (R Core Team, 2022).

Results

First, log data analysis from the LMS revealed patterns of three interaction types (see Figure 1). The total number of interactions with the content (53.92%) was 4617, followed by interaction with the instructor (26.33%) with 2254, and finally interaction with the system (19.75%) with 1691. These findings reveal interaction with content as the most frequently utilized interaction type.

Figure 1. Number of interactions of each type.



Second, data showed that student engagement was highest around April 24th, nearing finals, followed by March 13th, around midterm exams. This behavior pattern demonstrated that the assessments in the course could positively influence student learning behavior and engagement.

Third, we analyzed survey data, log data, and students' final scores to investigate predictors of students' academic performance. Using the forward method, a best-fit model was established, including teaching presence, cognitive presence, social presence, impact of PBL, self-efficacy(post), self-efficacy(pre), and interaction with content. The model was found to be significant ($p < .05$) with an adjusted R-squared of 0.97, indicating that 97% of the variances were explained by the model. Among predictors, impact of PBL, self-efficacy(post), self-efficacy(pre), cognitive presence, and interaction with content were significant predictors of students' academic performance.

Table 1. Results of multiple regression to investigate predictors of academic performance.

Predictor	Estimate	Std.Error	t value	Pr(>|t|)
(Intercept)	0.59	0.04	13.33	< 0.01 **
Teaching presence	0.01	< 0.01	4.27	0.05
Cognitive presence	< 0.01	< 0.01	4.55	< 0.05 *
Social presence	< 0.01	< 0.01	1.53	0.27
Impact of PBL	0.02	< 0.01	5.28	0.03 *
Self-efficacy post	-0.01	< 0.01	-11.67	< 0.01 **
Self-efficacy pre	0.01	< 0.01	10.69	< 0.01 **
Interaction with the Content	-0.001	< 0.01	-9.60	0.01 *

*p<.05, **p<.01, ***p<.001.

Implications and Recommendations

Online interaction data has demonstrated its potential to predict students' learning outcomes and identify at-risk learners in the early stages of a course. To further explore this area, future studies could benefit from larger sample sizes and conduct experiments to assess the impact of using learning analytics on improving instruction and improving students' academic performance. Additionally, investigating the accuracy of identifying at-risk learners through learning analytics would be valuable for informing targeted interventions and support systems

References

Arnold, K. & Pistilli, M. (2012). Course signals at Purdue: Using learning analytics to increase student success. Proceedings of the second international learning . analytics and knowledge conference (pp. 267–270). ACM.

Hillman, D. C., Willis, D. J., & Gunawardena, C. N. (1994). Learner–interface interaction in distance education: An extension of contemporary models and strategies for practitioners. *American Journal of Distance Education, 8*(2), 30–42.

Moore, M. G. (1989). Three types of interaction. American *Journal of Distance Education, 3*(2), 1–6.

R Core Team (2022). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.