# Student Engagement in a Gamified Online Learning Environment: A Data Mining Approach

Laura McNeill, Jewoong Moon, & Christopher Edmonson

DOI:10.59668/1269.15624

THE JOURNAL OF APPLIED INSTRUCTIONAL DESIGN

*In education, gamification applies game design elements to non-game contexts to engage learners. This preliminary study explored the feasibility of using data mining to analyze student engagement patterns in an asynchronous online course that utilized gamification. We conducted a series of exploratory data analysis approaches, including keyword extraction, clustering, concordance analysis, and sequential pattern mining. Results showed peaks in engagement aligned with gamified challenges and rewards. Students discussed both academic and non-academic topics, indicating community-building. Sequence pattern mining revealed more consistent academic engagement for highly engaged students. Findings demonstrate the viability of using data mining techniques to assess cognitive engagement based on discussion patterns in a gamified online course.*

## Introduction

Online learning has become ubiquitous in higher education, with most courses offering a mix of fully online and hybrid courses (Allen & Seaman, 2017). However, this transformation brings the challenge of student engagement to the forefront, yet limited interactions in virtual/remote environments with a lack of physical presence remain (Dixson, 2015).

In response to learning interaction and subsequent engagement challenges, gamification has surfaced as a promising strategy. By including online courses with game design elements—points, badges, leaderboards, and more—educators aim to promote student motivation and engagement (Nacke & Deterding, 2017). While gamification's merits are clear, its ability to gauge students' engagement during a gamified online course is questionable. Traditional approaches, particularly self-reported measures, may fall short of capturing the full spectrum of student engagement, often missing subtle but valuable behavioral cues. The research echoes a more unobtrusive and alternative measurement approach to track students' learning engagement patterns in online learning.

This gap highlights the need for sophisticated, insightful approaches, where the potential of Educational Data Mining (EDM) becomes particularly significant. As a rapidly emerging field within learning analytics and computational social science, EDM offers a unique advantage in providing educators with deep insights into student learning. This goes beyond superficial analysis, delving into the complex dynamics of student interactions. Such depth is particularly critical in large-scale online learning environments, where subtle student behaviors might otherwise remain undetected. It is evident that EDM's ability to track detailed micro-interactions, as noted by Daghestani et al. (2020), allows educators and researchers to discern underlying behavioral patterns, including those that might indicate students at risk (Bošnjaković & Đurđević Babić, 2023). Given this context, our study aims to bridge these gaps by applying EDM to investigate learning engagement within gamified online courses. Specifically, we seek to answer two pivotal research questions:

- RQ1. How do students engage in a gamified online, asynchronous course?
- RQ2. What learning engagement patterns emerge from using a gamification approach in an online, asynchronous course?

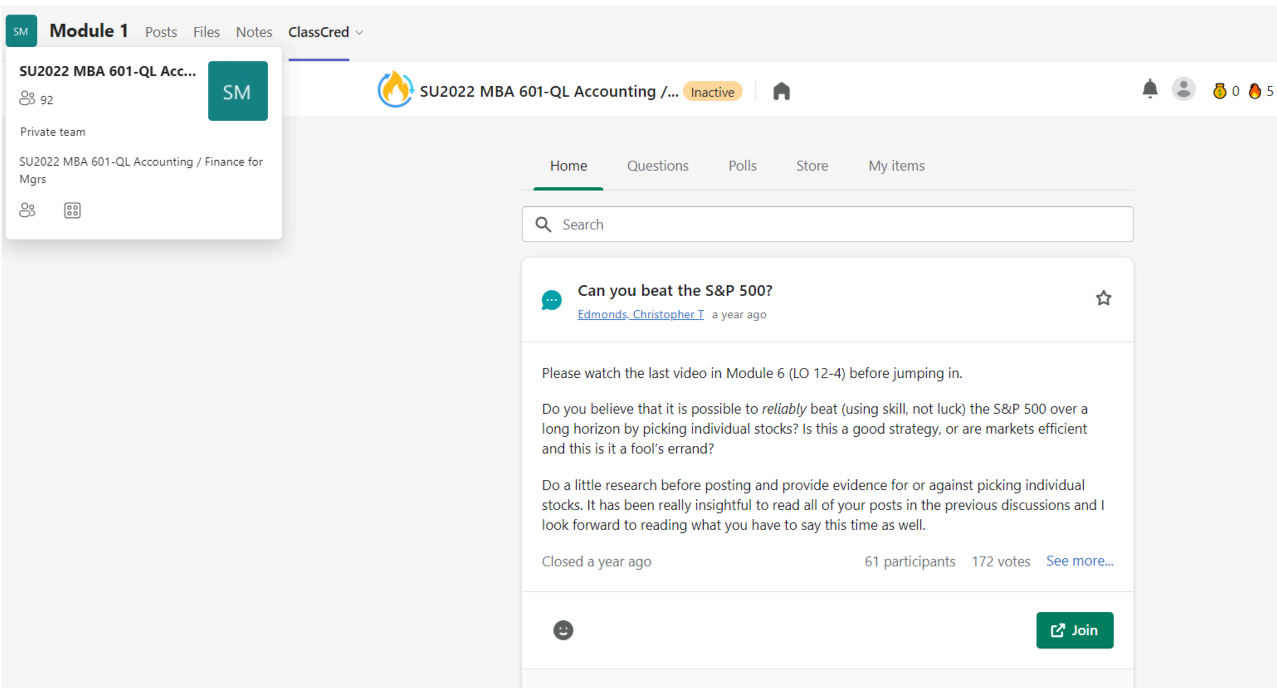
## Methods

This case study analyzed data from 50 students in one online financial literacy course section at a large public university. The course utilized ClassCred, a gamification plug-in app (Figure 1) embedded within the Microsoft Teams learning management system (LMS). This plug-in allows instructors to use or customize a collection of gamification functions, including points, badges, emojis, rewards, and social discussion features. We gathered a full semester of discussion posts (N=4,760), and data preprocessing was done to ensure accuracy, ethical compliance, and standardization of the textual data for subsequent analysis. This process included converting text to lowercase, removing punctuation and numbers, tokenizing the text, eliminating stop words, and applying stemming to reduce words to their root form. Analysis was completed using the following educational data mining techniques:

- Keyword extraction to identify academic topic-related posts.
- Clustering students into high and low-engagement groups
- Concordance analysis to examine the context of keyword usage (Adolphs, 2006).
- Sequence pattern mining to categorize discussion types and identify patterns (Mabroukeh & Ezeife, 2010).

**Figure 1**

*ClassCred Interface*




# Preliminary Findings

The data analysis is incomplete as this study is currently in progress. However, preliminary results suggest that data mining techniques are feasible and useful in exploring students' online engagement patterns in a gamified learning system.

RQ1. How do students engage in a gamified online, asynchronous course?

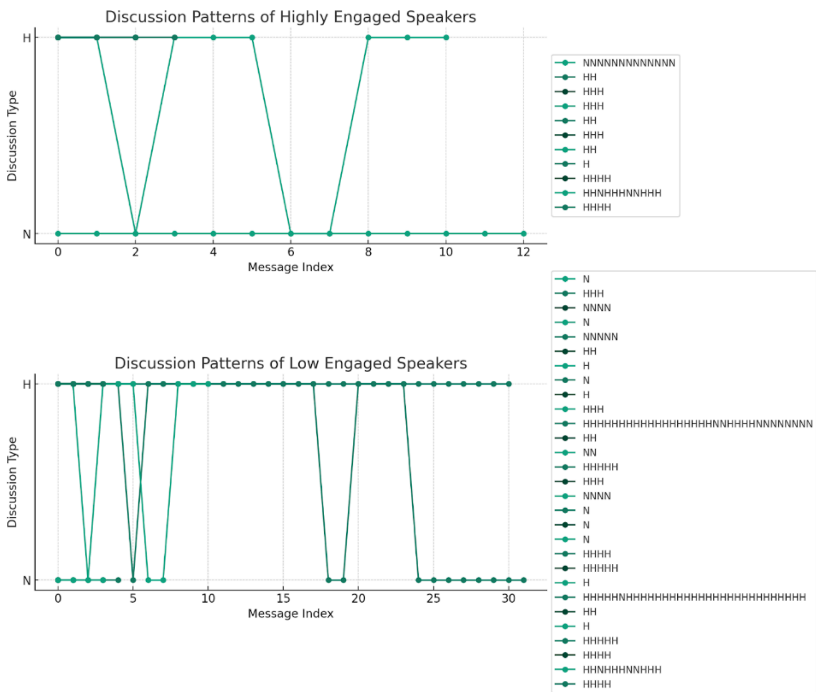
We clustered students into 'high' and 'low' engagement groups based on their cumulative engagement-related term frequency, identifying the top and bottom 20%, respectively, through keyword extraction techniques. In our study involving 50 students, this resulted in 10 students in the 'high' engagement group and 10 students in the 'low' engagement group. Our findings indicated that both groups, highly engaged and low engaged, raised questions about grading and exams. Highly engaged students, totaling 14, were more involved in academically engaged discussions, whereas the seven low-engaged students primarily raised general queries about exams and time limits and made comments about their performance expectations.

RQ2. What learning engagement patterns emerge from using a gamification approach in an online, asynchronous course?

The study's sequence pattern mining findings indicate that students with high engagement consistently engage in academic discussions, unlike their low-engagement counterparts, who display a wider variety of interaction patterns. Concordance analysis, particularly focusing on academic keywords like "exam" and "assignment," highlights the academic orientation of these discussions. Notably, student engagement fluctuated throughout the course, peaking alongside gamified rewards and milestones. Figure 2 exemplifies this by depicting two distinct learner groups, differentiated by their engagement levels (low vs. high). The high-engagement group predominantly participated in "Highly academically engaged" (H) discussions, interspersed with "Neutral" (N) discussions. The low-engagement group, while also engaging in some (H) discussions, showed a more diverse pattern, including a mix of (N) and "Irrelevant" (I) discussions. The categorization of discussion posts into (H), (N), and (I) was accomplished using an automated keyword extraction technique. This method identifies specific keywords and phrases that signify various levels of engagement. For example, frequent mentions of "exam" and "assignment" likely led to a post being classified as (H), reflecting a strong academic focus. In contrast, (N) posts, while course-related, lacked these specific academic keywords, indicating moderate engagement. Posts categorized as (I) were devoid of such keywords, suggesting they were off-topic. This approach allowed for efficient processing of a large data set, facilitating the analysis of engagement patterns in the online course. However, it primarily relies on keyword frequency and may not fully capture the depth and context of discussions, a limitation of automated text analysis. In sequence pattern mining, different sequences of symbols with more frequent H represent varying lengths or intensities of the same behavior or characteristic. In the context of your study, where "H" signifies "Highly academically engaged" discussions, these sequences would indicate different extents of sustained high engagement in academic discussions. Longer sequences (like "HHHHHHHHH") typically represent more prolonged and consistent behaviors or patterns, while shorter sequences (like "HH" or "HHH") may indicate more sporadic occurrences of the behavior. In educational contexts, longer sequences of high engagement could be particularly informative, possibly correlating with deeper or more effective learning experiences, or they might be influenced by course design elements like gamification, deadlines, or other motivating factors.

**Figure 2**

*Sequence Pattern Mining Result*




# Discussion

The results of this study offer empirical evidence that discussion analysis through a data mining approach is feasible to capture more objective clues on engagement assessments. This preliminary work suggests that gamification contributes to multifaceted engagement. In this study, the introduction of gamification elements like points, badges, and leaderboards led to a 15.09% increase in the quantity of posts and a 25.88% increase in the average length of threads. Specifically, threads featuring badge incentives saw a 22% increase in replies, suggesting a significant boost in engagement. Point-based leaderboards, particularly when enhanced with specific emojis, caused a 17% rise in daily active users, indicating the effectiveness of these strategies in tapping into intrinsic motivations for recognition and competition. The strategic use of gamification, particularly points (mentioned 134 times), played a notable role in enhancing engagement and participation. While the direct impact on replies and idea integration was not quantifiable from the data, the frequent mention of points suggests they were a motivating factor for many participants. As a mixed qualitative lens, the data suggests that the gamified interactions in this business course likely fostered a vibrant social learning environment where students quickly build consensus and engage in problem construction. The diverse sequence patterns among less engaged students identify an opportunity to adapt course design to better support at-risk students. This study also complements existing empirical works by exploring data mining techniques. These insights are crucial for understanding how to deploy evidence-based gamification strategically, creating an environment that adaptively fosters learning engagement.

# Conclusion

This exploratory study highlighted the potential of educational data mining as a powerful tool for identifying student learning engagement patterns in an asynchronous, gamified online course. We applied various data mining techniques, including keyword extraction, clustering, concordance analysis, and sequence pattern mining, to gain deep insights into student interactions within the gamified learning environment ClassCred. Building on these initial findings, future research should explore the long-term impact of gamification on student engagement and academic performance. Moreover, it would be beneficial to investigate the interaction between different gamification elements and personal factors such as learner profiles and individual motivation levels. To expand the scope of research, it is advisable to use a broader range of data analysis techniques beyond those employed in this study. For instance, methods that can examine the relationships between multiple variables could provide a more comprehensive understanding of the effects of gamification. This approach would allow for a more nuanced comparison to non-gamified courses and help assess the influence of specific game elements. Such research efforts will be crucial for enhancing instructional/learning strategies to ensure they are effectively tailored and adaptive, enhancing student learning engagement.

## References


Adolphs, S. (2006). *Introducing electronic text analysis: A practical guide for language and literary studies.* Routledge.

Allen, I.E. & Seaman, J. (2017). *Digital compass learning: Distance education enrollment report 2017*. Babson Survey Research Group. https://onlinelearningsurvey.com/reports/digtiallearningcompassenrollment2017.pdf

Bošnjaković, N., & Đurđević Babić, I. (2023). Systematic review on educational data mining in educational gamification. *Technology, Knowledge and Learning*, 1-18. https://doi.org/10.1007/s10758-023-09686-2

Daghestani, L. F., Ibrahim, L. F., Al-Towirgi, R. S., & Salman, H. A. (2020). Adapting gamified learning systems using educational data mining techniques. *Computer Applications in Engineering Education, 28*(3), 568-589. https://doi.org/10.1002/cae.22227

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

Mabroukeh, N. R., & Ezeife, C. I. (2010). A taxonomy of sequential pattern mining algorithms. *ACM Computing Surveys (CSUR), 43*(1), 1-41.

Nacke, L.E., & Deterding, S. (2017). The maturing of gamification research. *Computers in Human Behavior, 71*, 450-454. https://doi.org/10.1016/j.chb.2016.11.062

Rizvi, S., Rienties, B., Rogaten, J., & Kizilcec, R. F. (2020). Investigating variation in learning processes in a FutureLearn MOOC. *Journal of Computing in Higher Education, 32*, 162-181. https://doi.org/10.1007/s12528-019-09231-0

Sarra, A., Fontanella, L., & Di Zio, S. (2019). Identifying students at risk of academic failure within the educational data mining framework. *Social Indicators Research, 146*, 41-60. https://doi.org/10.1007/s11205-018-1901-8