Majumdar, R. et al. (Eds.) (2025). Proceedings of the 1st International Conference on Learning Evidence and Analytics. Asia-Pacific Society for Computers in Education

Integrating Learning Analytics and Learning Theories: A Clustering Approach to Student Engagement and Performance

Mu-Sheng CHEN & Ting-Chia HSU*
Technology Application and Human Resource Development, National Taiwan Normal University, Taiwan
*ckhsu@ntnu.edu.tw

Abstract: This study employed learning analytics to explore how students' login behavior and practice attempts relate to their academic performance. Using Pearson correlation and K-means clustering, we identified two distinct engagement profiles. Results showed that total time spent and practice attempts positively correlated with performance, whereas login frequency alone did not. These findings align with Constructivist and Self-Regulated Learning theories, highlighting the value of sustained engagement over superficial activity. This study contributes to the field by integrating theoretical frameworks with data-driven approaches to support personalized interventions.

Keywords: Learning analytics; Learning theories; K-means; Constructivist Learning; Self-Regulated Learning

1. Introduction

Learning analytics (LA) provides data-driven insights that can enhance educational practices (Siemens, 2013). However, to meaningfully interpret behavioral data, it is essential to align LA with learning theories. Constructivist Learning Theory emphasizes active knowledge construction through participation and iterative practice (Piaget, 2005), while Self-Regulated Learning (SRL) Theory underscores self-monitoring and strategic engagement (Zimmerman, 2002). Integrating these theories with LA can offer deeper understanding of student behaviors (Wise & Shaffer, 2015; Gašević et al., 2015).

This study examines how engagement behaviors—login frequency, time spent, and practice attempts—relate to academic performance. We define "login behaviors" as both the weekly frequency of platform visits (every week visit, EWV) and the average duration per login session (every time visit, ETV), and "practice attempts" as the total number of completed exercises or activities (count score, CS). The total amount of time a student spent using the platform is referred to as total visit time (TVT), and average performance is measured by the average score (AS). Given potential limitations in interpreting raw login counts, we argue that regularity and quality of logins, rather than quantity alone, provide more meaningful indicators of engagement (Clow, 2013).

We further apply clustering to categorize learners and assess performance differences. Although clustering is widely used in learning analytics (Viberg et al., 2018), our study contributes by grounding the clustering process in Constructivist and SRL theories, and by using engagement metrics derived from actual platform usage to produce practically interpretable student profiles for early identification and pedagogical intervention. The research questions are:

- How are students' login behaviors and practice attempts related to their learning performance?
- Can clustering method effectively categorize students based on their login behaviors and practice attempts?
- Do students in different levels of engagement clusters exhibit significant differences in learning performance?

2. Method

Data were collected from 471 students using the Taipei CooC Cloud platform between August 2020 and January 2021. Five variables were derived: EWV, TVT, ETV, AS, and CS.

We applied Pearson correlation analysis to examine relationships among engagement variables and academic performance. To classify students by engagement profiles, K-means clustering was selected for its efficiency and interpretability. Optimal cluster number (k = 2) was determined using the elbow method and silhouette coefficient.

3. Result and Discussion

3.1 Correlation analysis

To examine the relationship between students' engagement behaviors and academic performance, Pearson correlation analyses were conducted. The results in Table 1 show that there are significant correlations between several key variables.

EWV negatively correlated with ETV and CS, and showed no significant relation to AS, suggesting frequent logins did not indicate effective engagement. Conversely, TVT, ETV, and CS were all significantly positively correlated with AS, indicating that time investment and practice are better predictors of performance.

Table 1. *Results of Pearson correlation analyses.*

	Mean	SD	EWV	TVT	ETV	AS	CS
EWV	1.45	0.98					
TVT	3661.30	2807.61	0.194**				
ETV	129.11	110.08	-0.433**	0.547**			
AS	67.73	16.26	0.01	0.241**	0.212**		
CS	18.88	8.43	-0.197**	0.307**	0.276**	0.221**	

**p<.01

3.2 K-means

K-means produced two clusters. The Cluster 1 had significantly higher scores on TVT, ETV, AS, and CS, but lower EWV. This implies that students who engaged less frequently but more deeply achieved better outcomes, reinforcing that depth of interaction matters more than frequency.

To statistically compare the variables between the two clusters, the independent sample t test was used. The results showed that, except for EWV, each variable in the first cluster was significantly higher than that in the second cluster, as shown in Table 2.

Table 1. *Results of independent sample t test.*

	Cluster	N	Mean	SD	*t*	*d*
EWV	1	207	1.13	0.70	-6.96***	0.62
	2	264	1.70	1.09		
TVT	1	207	5249.54	3078.68	15.21***	1.13
	2	264	2415.97	1766.56		
ETV	1	207	203.15	110.08	11.81***	1.45
	2	264	71.06	66.83		
AS	1	207	77.68	10.28	14.68***	1.33
	2	264	59.93	15.84		
CS	1	207	24.73	9.10	15.53***	1.50
	2	264	14.28	3.73		

***p <.001

3.3 Theoretical Implications

These findings align with Constructivist and SRL theories, which emphasize meaningful, self-directed learning (Piaget, 2005; Zimmerman, 2002). Our clustering approach demonstrates how log data can be used to generate actionable insights for both teachers and students. Teachers may use these cluster profiles to identify at-risk students early, while students may be empowered by real-time feedback on their engagement patterns to self-regulate their learning (Sailer et al., 2024). As Selwyn (2014) suggests, however, interpreting learning analytics requires awareness of sociotechnical contexts, including issues such as fragmented login behavior due to unstable internet or inconsistent device use.

This integration of theory and analytics not only enhances the interpretability of behavioral data but also facilitates the development of pedagogical strategies grounded in evidence. The use of cluster-informed profiles, for instance, enables differentiated support for learners who may otherwise be overlooked by aggregate measures such as average scores alone (Khalil et al, 2023). Moreover, by aligning behavioral metrics with theoretical constructs, such as strategic regulation and constructive engagement, this study provides a replicable framework for future research that aims to close the loop between learning behaviors, system design, and instructional feedback (Kitto et al, 2023; Wise & Shaffer, 2015).

4. Conclusion

This study demonstrates the importance of engagement quality over quantity in predicting academic success. By integrating learning theories with LA methods like clustering, we can identify meaningful behavior patterns and tailor interventions. These insights inform the design of adaptive learning environments that prioritize sustained, self-regulated engagement (Sailer et al., 2024). Future work should investigate finer-grained login regularity and contextual factors to improve predictive accuracy and student empowerment.

References

Clow, D. (2013). An overview of learning analytics. *Teaching in Higher Education, 18*(6), 683-695. https://doi.org/10.1080/13562517.2013.827653

Gašević, D., Dawson, S., & Siemens, G. (2015). Let's not forget: Learning analytics are about learning. *Techtrends Tech Trends, 59*, 64-71. https://doi.org/10.1007/s11528-014-0822-x

Khalil, M., Prinsloo, P., & Slade, S. (2023). The use and application of learning theory in learning analytics: A scoping review. *Journal of Computing in Higher Education, 35*(3), 573-594. https://doi.org/10.1007/s12528-022-09340-3

Kitto, K., Hicks, B., & Buckingham Shum, S. (2023). Using causal models to bridge the divide between big data and educational theory. *British Journal of Educational Technology, 54*(5), 1095-1124. https://doi.org/10.1111/bjet.13321

Piaget, J. (2005). *The psychology of intelligence.* Routledge.

Sailer, M., Ninaus, M., Huber, S. E., Bauer, E., & Greiff, S. (2024). The End is the Beginning is the End: The closed-loop learning analytics framework. *Computers in Human Behavior, 158*, 108305. https://doi.org/10.1016/j.chb.2024.108305

Selwyn, N. (2014). *Distrusting educational technology. Critical questions for changing times*. Routledge.

Siemens, G. (2013). Learning analytics: The emergence of a discipline. *American Behavioral Scientist, 57*(10), 1380-1400. https://doi.org/10.1177/0002764213498851

Viberg, O., Hatakka, M., Bälter, O., & Mavroudi, A. (2018). The current landscape of learning analytics in higher education. *Computers in Human Behavior, 89*, 98-110. https://doi.org/10.1016/j.chb.2018.07.027

Wise, A. F., & Shaffer, D. W. (2015). Why theory matters more than ever in the age of big data. *Journal of Learning Analytics, 2*(2), 5-13.

Zimmerman, B. J. (2002). Becoming a self-regulated learner: An overview. *Theory into practice, 41*(2), 64-70. https://doi.org/10.1207/s15430421tip4102_2