

DOI: https://doi.org/10.48009/3_iis_2026_116

From reactive grading to proactive intervention: building a behavioral early warning system for online learners

Hannah Roberts, *Georgia College & State University*, hannah.roberts1@bobcats.gcsu.edu
Daniel D. Wu, *Georgia College & State University*, daniel.wu@gcsu.edu

Abstract

In online and distance education, the ability to identify at-risk students early is essential for improving retention and academic success. This study develops a Behavioral Early Warning System (EWS) using clickstream data from the Open University Learning Analytics Dataset (OULAD), which includes 32,593 students and over 10.6 million interactions across 22 module presentations. The research introduces a multi-dimensional behavioral framework focusing on four key dimensions: Consistency (Gini Study Distribution), Intensity (average clicks per active day), Total Effort, and Spacing (time between sessions), supplemented by academic context variables. Using binary logistic regression, the full-dataset baseline model achieved 74.92% accuracy with a strong recall of 0.81 for failing students. When tested with limited temporal windows to simulate real-world deployment, the model reached 62.6% accuracy using only the first 30 days of data and improved to 68.7% by day 150. The findings demonstrate that temporal patterns of student engagement are detectable early enough to support proactive intervention. Interestingly, the Gini Study Distribution variable exhibited a positive relationship with student success, suggesting that moderate and strategically concentrated engagement patterns may reflect adaptive self-regulated learning behaviors rather than purely ineffective cramming. In contrast, excessively high-intensity study sessions were associated with elevated academic risk. These results highlight the importance of not only how much students engage, but also how they distribute and structure their effort over time. Overall, the study provides a practical, interpretable, and scalable framework for institutions seeking to shift from reactive grading to timely and data-driven student support in online learning environments.

Keywords: learning analytics, early warning system, clickstream analysis, student success prediction, Gini coefficient, online Learning, logistic regression

Introduction

The rapid expansion of online and distance education has transformed how institutions monitor and support student success. With millions of learners now engaging in virtual learning environments (VLE), educators can access rich behavioral data that extends far beyond traditional performance indicators such as midterm grades or final examinations. Learning Analytics has emerged as a powerful discipline that leverages this data to develop early warning systems (EWS), enabling timely and proactive interventions before academic difficulties become entrenched (Kuzilek et al., 2017; Tempelaar et al., 2015).

A critical limitation of many existing predictive approaches is their reliance on aggregate activity metrics, such as total login time or overall click volume. These measures often fail to capture the underlying rhythms

of student engagement. For example, two students may generate similar total activity levels, one through consistent daily study sessions and the other through infrequent, high-intensity cramming, yet resulting in markedly different learning outcomes. Psychological research on the spacing effect demonstrates that distributed practice leads to superior long-term retention compared to massed practice (Cepeda et al., 2006; Miyamoto et al., 2015). Consequently, temporal patterns of engagement, including consistency, intensity, and spacing of study sessions, represent promising yet underutilized predictors of persistence and achievement.

This study addresses this gap by developing a predictive model that prioritizes the temporal distribution and behavioral rhythm of student effort within the Open University Learning Analytics Dataset (OULAD). OULAD is one of the largest and most comprehensive publicly available datasets in distance education, containing anonymized data on 32,593 students across 22 module presentations from seven courses. The dataset integrates demographic information, assessment results, and more than 10.6 million timestamped interactions with the VLE, recorded as daily click summaries (Kuzilek et al., 2017).

We propose a behavioral framework organized around four key dimensions: Consistency (measured via the Gini Study Distribution), Intensity (average clicks per active session), Total Effort (cumulative activity), and Spacing (temporal gaps between study sessions). These dimensions are combined with selected academic background variables (credits studied and number of prior attempts) in a binary logistic regression model to predict the probability of passing a module.

The primary objectives of this study are:

1. To examine whether temporal behavioral features provide predictive value beyond traditional volume-based metrics.
2. To evaluate model performance using progressively limited temporal windows (30 to 150 days) to simulate real-world early warning deployment.
3. To assess the feasibility of identifying at-risk students early enough in the term for meaningful pedagogical intervention.

By transforming granular clickstream data into actionable behavioral signatures, this study offers a scalable and interpretable framework for moving from reactive grading to proactive student support in online learning environments.

Literature Review

The growth of online and distance education has generated unprecedented volumes of student behavioral data, shifting the focus of educational research from reactive assessment to proactive prediction and intervention. Learning Analytics has become central to this transformation, with EWS representing one of its most practical applications (Akçapınar et al., 2019; Marbouti et al., 2016; Tempelaar et al., 2015).

OULAD has emerged as a foundational resource for such research. Released in 2017, it contains detailed demographic, assessment, and clickstream data for 32,593 students across 22 module presentations, totaling over 10.6 million interactions with the VLE (Kuzilek et al., 2017). Multiple studies using OULAD have demonstrated that behavioral features derived from clickstream data consistently outperform demographic variables in predicting student success (Chen & Cui, 2020; Gardner & Brooks, 2018; Hlosta et al., 2017).

A key insight from this body of work is that the rhythm of student engagement often matters more than raw volume. Psychological research on the spacing effect shows that distributing study sessions over time leads

to stronger long-term retention than massed practice (cramming) (Cepeda et al., 2006; Carvalho & Goldstone, 2020; Miyamoto et al., 2015). In MOOC and online course contexts, students who engage in more frequent, regularly spaced sessions demonstrate significantly higher completion rates, even after controlling for total time on task (Miyamoto et al., 2015).

To capture this temporal regularity, researchers have increasingly turned to inequality metrics such as the Gini coefficient. Originally developed in economics, the Gini index has been successfully adapted in educational research to quantify the uneven distribution of student effort or performance (Nguyen et al., 2020; Wagner et al., 2025). A high Gini value indicates “bursty” behavior, i.e., long periods of inactivity followed by intense cramming, while lower values reflect consistent and rhythmic engagement.

Beyond consistency, the intensity and timeliness of interactions provide additional predictive power (Dvorak & Jia, 2016). EWS implementations have moved toward multi-dimensional models that incorporate not only volume but also patterns of engagement and disengagement “drift” (Villano et al., 2018).

Despite these advances, several gaps remain. Many studies still rely heavily on aggregate metrics or late-stage predictors. Few have systematically examined how well temporal behavioral signatures perform when restricted to early- or mid-course data windows. This study addresses these gaps by developing a unified behavioral framework built on four dimensions, i.e. Consistency (Gini Study Distribution), Intensity, Total Effort, and Spacing, and evaluating performance across progressive temporal windows.

Methodology

The dataset utilized in this study comprises 29,230 student records derived from the OULAD, filtered to ensure data completeness for our behavioral analysis. We employed a standard 80/20 split, designating 80% of the data for model training and the remaining 20% for testing.

Target Variable

A binary outcome variable was created for classification. Students who achieved a “Pass” or “Distinction” were labeled as 1 (“Pass”). Students who “Failed” or “Withdrawn” were labeled as 0 (“Fail”) and defined as “at-risk”. This consolidation produced a balanced yet realistic representation of module outcomes.

Feature Engineering

Four behavioral dimensions were engineered to capture the rhythm and quality of student engagement:

- Consistency: Gini Study Distribution calculated on daily click counts.
- Intensity: Average clicks per active study day (each calendar day counted once).
- Total Effort: Total number of clicks across the observation window.
- Spacing: Mean and median days between consecutive active study days.

Academic context variables included Credits Studied and Number of Previous Attempts. Features were computed cumulatively for temporal windows (30, 60, 90, 120, and 150 days).

Modeling Approach and Temporal Analysis

Binary Logistic Regression was used:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_k x_k)'}}$$

where x_i represent the behavioral and academic features, and β_i are the estimated coefficients.

To evaluate the model's utility as an EWS, separate models were trained and tested using only data available up to each cutoff date (30, 60, 90, 120, and 150 days). A standard 70/30 train-test split with stratified sampling was applied within each window. Model performance was assessed using overall accuracy, precision, recall, F1-score, and confusion matrices. All analysis was conducted in Python using pandas, NumPy, scikit-learn, and matplotlib/seaborn libraries.

Results

The binary logistic regression model was first evaluated using the full course dataset to establish a performance baseline. The model achieved an overall accuracy of 74.92%. Detailed performance metrics are presented in Table 1.

Table 1. Classification Report (Full Dataset).

Class	Precision	Recall	F-1 Score	Support
Fail	0.71	0.81	0.75	2,781
Pass	0.80	0.70	0.74	3,065

The model demonstrated particularly strong recall (0.81) for the Fail class, which is highly desirable for an Early Warning System as it successfully identifies the majority of at-risk students. The confusion matrix (Figure 1) shows 2,246 true positives for Fail and 2,134 true positives for Pass, with 535 false positives and 931 false negatives.

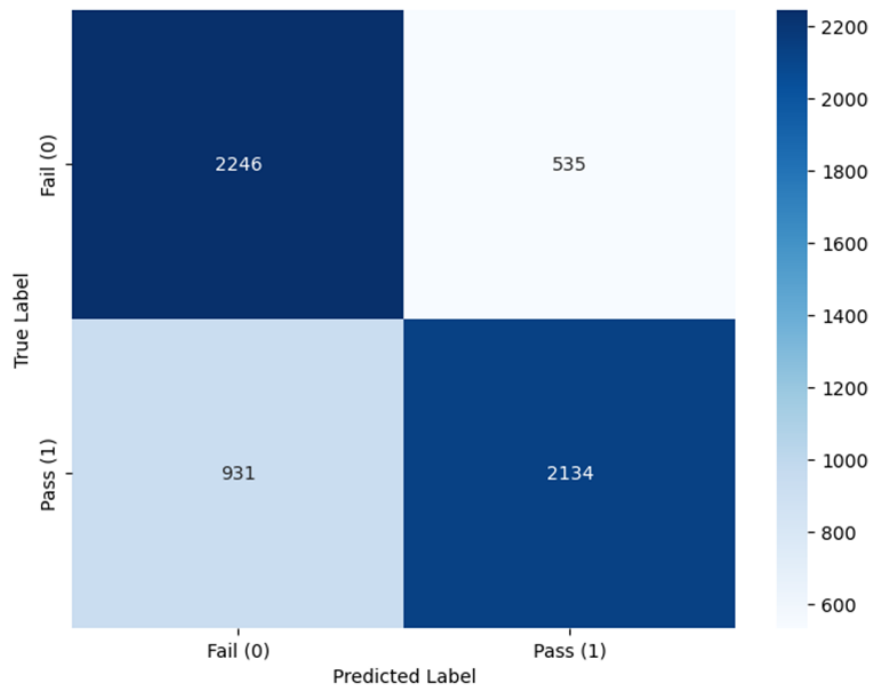


Figure 1. Confusion Matrix (Full Term Model).

Interestingly, the Gini Study Distribution variable exhibited a positive coefficient in the logistic regression model. While high Gini values are often associated with irregular or concentrated engagement patterns, this finding suggests a more nuanced relationship between temporal study behavior and academic success. In online learning environments, some degree of concentrated engagement may reflect adaptive and strategically timed study behavior rather than purely ineffective cramming. Successful students may intensify effort around assessments, challenging content, or key deadlines while still maintaining sufficient overall engagement throughout the course. When considered alongside the negative coefficient for high-intensity session behavior, the results suggest that moderate strategic concentration of effort may be beneficial, whereas excessively bursty or binge-like engagement patterns are associated with elevated academic risk. This finding highlights the complexity of temporal learning behavior and suggests that engagement quality and timing may matter as much as consistency alone.

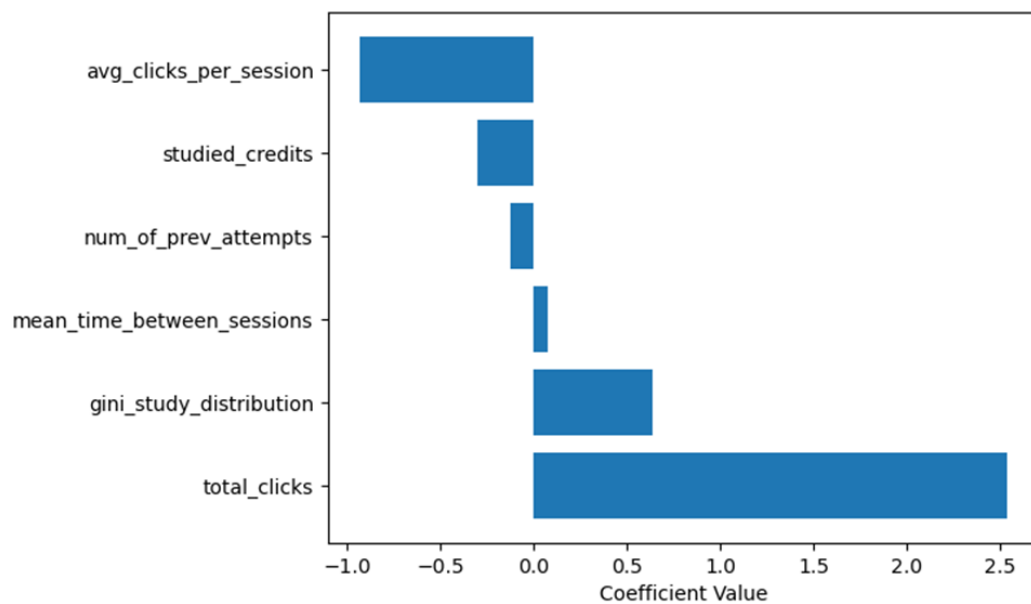


Figure 2. Feature Importance (Logistic Regression Coefficients).

Table 2. Selected Logistic Regression Coefficients (Full Model).

Feature	Coefficient	Interpretation
Total Clicks	Positive (strongest)	Higher Volume → Better Outcomes
Gini Study Distribution	Positive	Less Consistent Patterns → Better Outcomes
Average Clicks per Session	Negative (strongest)	Bursty/High-Intensity Sessions → Higher Risk
Credits Studied	Negative	Higher Load → Higher Risk
Number of Previous Attempts	Negative	Prior Difficulty → Higher Risk
Mean Time Between Sessions	Marginal	Smaller Gaps → Better Outcomes

The predictive capability of the full-term model is also visually summarized through the ROC curve, which illustrates the trade-off between the True Positive Rate (Sensitivity) and the False Positive Rate across various classification thresholds. As shown in Figure 3, the model achieves an AUC of 0.84, a value that indicates strong discriminative power and signifies that the model can correctly distinguish between a student who will pass and one at risk of failure approximately 84% of the time. This confirms that engineered features such as Consistency, Intensity, Total Effort, and Spacing provide a reliable "behavioral signature" for success. Furthermore, the proximity of the curve to the upper-left quadrant graphically

reinforces the model's ability to maintain the high Recall for the "Fail" class (0.81), which is essential for a proactive Early Warning System (EWS).

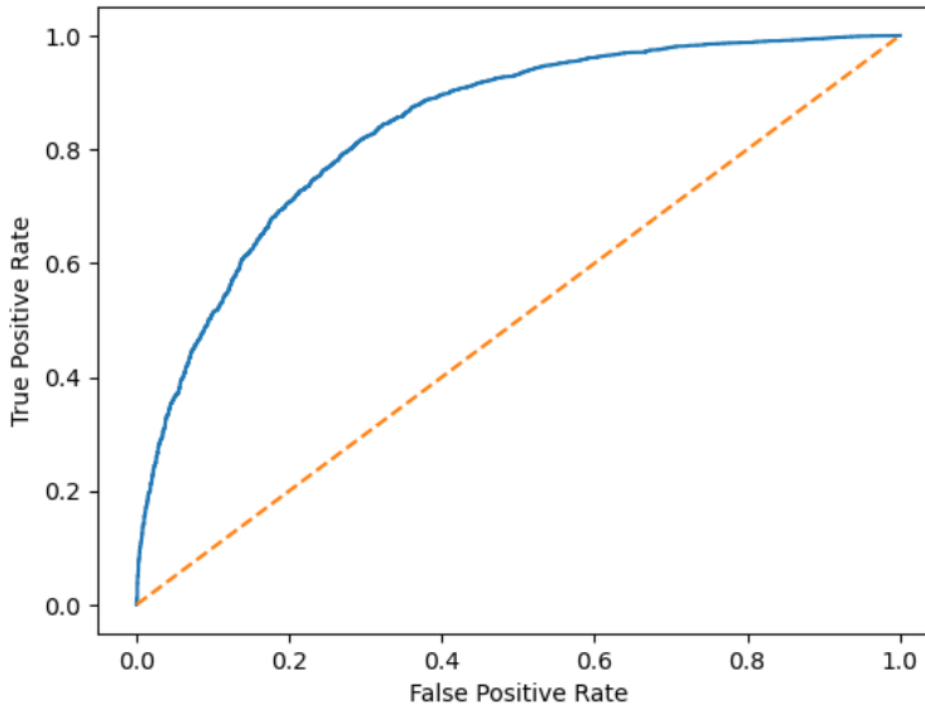


Figure 3. ROC (AUC = 0.84).

To simulate real-world deployment of the EWS, the model was retrained and evaluated using only data available within progressively larger temporal windows (30, 60, 90, 120, and 150 days). Results are summarized in Table 3.

Table 3. EWS Performance Across Temporal Windows.

Time Window	Precision	Recall	F-1 Score	Accuracy
30 Days	0.62	0.74	0.67	0.626
60 Days	0.63	0.75	0.68	0.637
90 Days	0.65	0.76	0.70	0.658
120 Days	0.65	0.75	0.70	0.656
150 Days	0.68	0.78	0.73	0.687

Performance improved steadily as more behavioral data accumulated, reaching a peak of 68.7% accuracy by day 150 (Figure 3). Notably, even with only the first 30 days of data, the model achieved 62.6% accuracy with strong recall for at-risk students (approximately 0.74). These results demonstrate that meaningful early warning signals can be detected well before the end of the course.

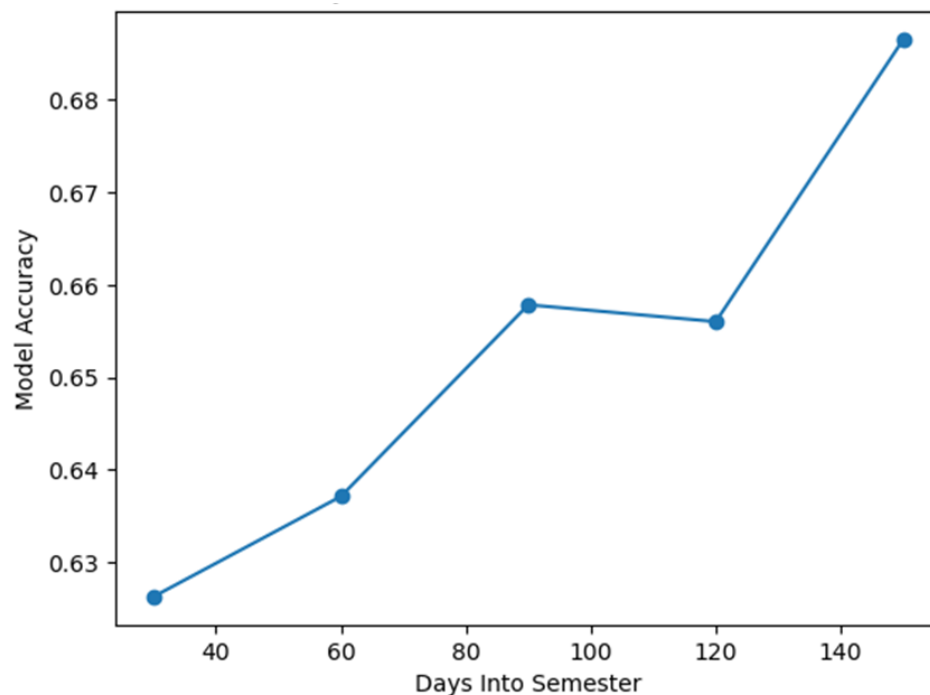


Figure 4. How early can we predict student outcomes?

Discussion and Conclusion

This study demonstrates that behavioral rhythms extracted from clickstream data can provide meaningful predictive insight into student success in online learning environments. Using OULAD, the study developed a behavioral EWS centered on four dimensions of engagement: Consistency, Intensity, Total Effort, and Spacing. Rather than relying solely on aggregate activity measures, the framework emphasizes how students distribute and sustain their effort throughout a course. The results suggest that temporal engagement patterns contain actionable signals that can support earlier and more proactive educational intervention.

The full-term logistic regression model achieved an overall accuracy of 74.92%, with particularly strong recall for at-risk students. This is an important outcome in the context of EWS because identifying struggling students is often more valuable than maximizing overall classification accuracy alone. More importantly, the temporal-window analysis demonstrated that predictive signals emerge relatively early in the course timeline. Using only the first 30 days of behavioral data, the model achieved over 62% accuracy, and performance improved steadily as additional engagement data accumulated. These findings indicate that institutions may be able to identify vulnerable students early enough to provide timely support before disengagement becomes difficult to reverse.

The feature analysis also provides important educational insights into how students organize and regulate their learning behavior. Total engagement volume remained a strong positive predictor of success, reinforcing the importance of sustained participation in online learning environments. Interestingly, the Gini Study Distribution variable also exhibited a positive coefficient in the logistic regression model. While traditionally an economic inequality measure, we justify its use here as a robust proxy for quantifying the "burstiness" and temporal dispersion of engagement, providing a more granular behavioral signature than

simple volume-based metrics. Although high Gini values are commonly associated with irregular or concentrated engagement patterns, this finding suggests a more nuanced relationship between temporal study behavior and academic performance. In online learning environments, some degree of concentrated engagement may reflect adaptive and strategically timed study behavior rather than purely ineffective cramming. Successful students may intensify their effort around assessments, challenging course content, or important deadlines while still maintaining sufficient engagement throughout the course.

This interpretation becomes more meaningful when considered alongside the strong negative coefficient for Average Clicks per Session. Together, these findings suggest that moderate and strategically concentrated engagement may be beneficial, whereas excessively intense, bursty, or binge-like study sessions are associated with elevated academic risk. In other words, productive self-regulated learning appears to involve intentional allocation of effort across the course timeline rather than either extreme uniformity or last-minute cramming. These results highlight the complexity of temporal learning behavior and suggest that the timing and structure of engagement may matter as much as overall activity volume.

Academic context variables also contributed meaningfully to the model. Students carrying heavier credit loads and those with multiple prior attempts demonstrated greater likelihood of unsuccessful outcomes, indicating that behavioral indicators should be interpreted within broader academic circumstances. This reinforces the importance of combining clickstream behavioral data with contextual academic variables when designing institutional early warning systems.

From a practical perspective, this study contributes a transparent and interpretable framework for behavioral EWS. Unlike more complex black-box models, logistic regression allows institutions and instructors to understand how specific behavioral variables influence predicted outcomes. This interpretability is particularly valuable in educational settings where actionable intervention and explainability are essential. The framework is also computationally lightweight and scalable, making it potentially adaptable across a variety of online and hybrid learning environments.

At the same time, several limitations should be acknowledged. Clickstream data captures the quantity and timing of interactions but cannot fully measure cognitive engagement, motivation, or conceptual understanding. The study also relies on data from a single institutional dataset, which may limit generalizability to other learning management systems, course structures, or student populations. In addition, while the model demonstrates predictive capability, the study does not evaluate whether interventions based on these predictions actually improve student outcomes. Furthermore, we characterize the 75% predictive accuracy as moderately successful, positioning the model as a practical decision-support tool for educators. While the proposed model demonstrates reasonable efficacy within standard academic term lengths, we acknowledge that its deployment in accelerated 8-week or 10-week courses would necessitate model recalibration to account for the requisite re-scaling of temporal data windows.

Future research should expand the framework by incorporating richer behavioral and instructional features, such as discussion participation, assignment submission patterns, sentiment analysis, or content interaction types. More advanced temporal modeling approaches, including recurrent neural networks or survival analysis, may also improve predictive performance. Future studies should also examine potential nonlinear relationships between engagement consistency and student success, particularly whether moderate variability in study behavior may be more beneficial than either extreme uniformity or highly concentrated cramming patterns. Most importantly, future work should evaluate the effectiveness of intervention strategies triggered by early warning predictions in real educational settings.

Overall, this research demonstrates that temporal engagement behaviors provide valuable early indicators of student success and academic risk. By shifting the focus from reactive grading toward proactive support, behavioral early warning systems have the potential to help institutions create more responsive, personalized, and student-centered online learning environments.

References

- Akçapınar, G., Hasnine, M. N., & Ueda, H. (2019). Using learning analytics to develop early-warning system for at-risk students. *International Journal of Educational Technology in Higher Education*, 16(1), 40. <https://doi.org/10.1186/s41239-019-0172-z>
- Berens, J., Schneider, K., Görtz, S., Oster, S., & Burghoff, J. (2019). Early detection of students at risk: Predicting student dropouts using administrative student data from German universities and machine learning methods. *Journal of Educational Data Mining*, 11(3), 1–41. <https://doi.org/10.5281/zenodo.3594771>
- Carvalho, P. F., Sana, F., & Yan, V. X. (2020). Self-regulated spacing in a massive open online course is related to better learning. *NPJ Science of Learning*, 5, 2. <https://doi.org/10.1038/s41539-020-0061-1>
- Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, 132(3), 354–380. <https://doi.org/10.1037/0033-2909.132.3.354>
- Chen, F., & Cui, Y. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics*, 7(2), 1–17. <https://doi.org/10.18608/jla.2020.72.1>
- Crossley, S., Paquette, L., Dascalu, M., McNamara, D. S., & Baker, R. S. (2016). Combining click-stream data with NLP tools to better understand MOOC completion. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16)* (pp. 6–14). Association for Computing Machinery. <https://doi.org/10.1145/2883851.2883931>
- Dvorak, T., & Jia, M. (2016). Online Work Habits and Academic Performance. *Journal of Learning Analytics*, 3(3), 318–330. <https://doi.org/10.18608/jla.2016.33.15>
- Gardner, J., & Brooks, C. (2018). Student success prediction in MOOCs. *User Modeling and User-Adapted Interaction*, 28(2), 127–203. <https://doi.org/10.1007/s11257-018-9203-z>
- Hlosta, M., Zdrahal, Z., & Zendulka, J. (2017). Ouroboros: Early identification of at-risk students without models based on legacy data. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17)* (pp. 6–15). Association for Computing Machinery. <https://doi.org/10.1145/3027385.3027449>
- Jin, L., Wang, Y., Song, H., So, H. J. (2024). Predictive Modelling with the Open University Learning Analytics Dataset (OULAD): A Systematic Literature Review. In: Olney, A.M., Chounta, I.A., Liu, Z., Santos, O.C., Bittencourt, I.I. (eds) Artificial Intelligence in Education. Posters and Late Breaking Results, Workshops and Tutorials, Industry and Innovation Tracks, Practitioners,

- Doctoral Consortium and Blue Sky. AIED 2024. *Communications in Computer and Information Science*, vol 2150. Springer, Cham. https://doi.org/10.1007/978-3-031-64315-6_46
- Johar, N. A., Kew, S. N., Tasir, Z., & Koh, E. (2023). Learning Analytics on Student Engagement to Enhance Students' Learning Performance: A Systematic Review. *Sustainability*, 15(10), 7849. <https://doi.org/10.3390/su15107849>
- Kuzilek, J., Hlosta, M., & Zdrahal, Z. (2017). Open University Learning Analytics dataset. *Scientific Data*, 4, 170171. <https://doi.org/10.1038/sdata.2017.171>
- Marbouti, F., Diefes-Dux, H. A., & Madhavan, K. (2016). Models for early prediction of at-risk students in a course using standards-based grading. *Computers & Education*, 103, 1–15. <https://doi.org/10.1016/j.compedu.2016.09.005>
- Merchant, A., Shenoy, N., Bharali, A., & Kumar, M. A. (2022). Predicting students' academic performance in virtual learning environment using machine learning. In *Proceedings of the 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T)* (pp. 1–6). <https://doi.org/10.1109/ICPC2T53885.2022.9777008>
- Miyamoto, Y. R., Coleman, C., Williams, J. J., Whitehill, J., Nesterko, S., & Reich, J. (2015). Beyond Time-on-Task: The Relationship Between Spaced Study and Certification in MOOCs. *Journal of Learning Analytics*, 2(2), 47-69. <https://doi.org/10.18608/jla.2015.22.5>
- Nguyen, Q., Rienties, B., & Richardson, J. T. E. (2020). Learning analytics to uncover inequality in behavioural engagement and academic attainment in a distance learning setting. *Assessment & Evaluation in Higher Education*, 45(4), 594–606. <https://doi.org/10.1080/02602938.2019.1679088>
- Roberts, L. D., Howell, J. A., Seaman, K., & Gibson, D. C. (2016). Student attitudes toward learning analytics in higher education: “The Fitbit version of the learning world”. *Frontiers in Psychology*, 7, 1959. <https://doi.org/10.3389/fpsyg.2016.01959>
- Sabnis, S., Yu, R., & Kizilcec, R. F. (2022). Large-scale student data reveal sociodemographic gaps in procrastination behavior. In *Proceedings of the Ninth ACM Conference on Learning @ Scale (L@S '22)* (pp. 133–141). Association for Computing Machinery. <https://doi.org/10.1145/3491140.3528285>
- Seo, E. Y., Yang, J., Lee, J. E., & So, G. (2024). Predictive modelling of student dropout risk: Practical insights from a South Korean distance university. *Heliyon*, 10(11), e30960. <https://doi.org/10.1016/j.heliyon.2024.e30960>
- Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: Learning analytics in a data-rich context. *Computers in Human Behavior*, 47, 157–167. <https://doi.org/10.1016/j.chb.2014.05.038>
- Villano, R., Harrison, S., Lynch, G., & Chen, G. (2018). Linking early alert systems and student retention: A survival analysis approach. *Higher Education*, 76(5), 903–920. <https://doi.org/10.1007/s10734-018-0249-y>

Wagner, D. A., Nair, T., Thapa, A., & Kumar, A. (2025). The Gini Learning Index: A new framework to measure learning inequality across contexts. *International Journal of Educational Development*, 119, 103433. <https://doi.org/10.1016/j.ijedudev.2025.103433>

Yu, C.-H., Wu, J., & Liu, A.-C. (2019). Predicting Learning Outcomes with MOOC Clickstreams. *Education Sciences*, 9(2), 104. <https://doi.org/10.3390/educsci9020104>