

# Mining and Assessing Anomalies in Students' Online Learning Activities with Self-supervised Machine Learning

Lan Jiang  
University of Illinois Urbana–Champaign  
Champaign, IL, USA  
lanj3@illinois.edu

Nigel Bosch  
University of Illinois Urbana–Champaign  
Champaign, IL, USA  
pnb@illinois.edu

## ABSTRACT

Two students in the same course working toward the same learning objectives may have very different strategies. However, on average, there are likely to be some patterns of student actions that are more common than others, especially when students are implementing typical self-regulated learning strategies. In this paper, we focus on distinguishing between students' typical actions and unusual, anomalous sequences of actions. We define anomalous activities as unexpected activities given a student's preceding activities. We distinguish these anomalies by training a self-supervised neural network to determine how predictable activities happen (the complement of which are anomalies). A random forest model trained to predict course grades from anomaly-based features showed that anomalous actions were significant predictors of course grade (mean Pearson's  $r = .399$  across 7 courses). We also explore whether humans regard the anomalous activities labeled by the model as anomalies by asking people to label 20 example sequences. We further discuss the implications of our method and how detecting and understanding anomalies could potentially help improve students' learning experiences.

## Keywords

Anomalies, Human understanding of anomalies, Log activities

## 1. INTRODUCTION

Online education systems can provide personalized learning experiences by understanding students' learning behavior automatically, given rich data that can be collected through such systems [18, 6, 12, 33]. Most research in this area focuses on investigating specific, theory-driven phenomena via data analytics and employs data-driven approaches to understand typical learning behaviors (the most frequent actions, most commonly studied resources, etc.) [33, 44, 23] and prediction tasks (grade/dropout prediction, test recommendation, etc.) [6, 25, 32, 41, 35]. These approaches, while

valuable, rarely consider the role of *anomalous* behaviors, which are also important to understand. For example, existing work has shown that some anomalous behaviors positively correlate with high course grade [17]. Much remains to be discovered regarding anomalous actions, how to determine which kinds of activities are anomalous activities, and how humans perceive anomalies. In this paper, we define students' activities in terms of typical (i.e., predictable) activities and anomalous activities (i.e., unpredictable activities), and describe a method for uncovering these anomalies. In addition, we investigate whether human experts' perceptions of anomalies align with anomalies identified by the proposed method, and explore how humans distinguish anomalous versus typical student activities.

We focus on data from log files [5], which accumulate a great deal of interaction information to understand anomalies in student learning behaviors. Due to the large amount of data in log files, it is difficult to glean insights from these manually. Thus, researchers have devised methods like behavior mining to extract insights computationally [20, 31]. However, behavior-based inferences mainly rely on handcrafted features [40, 43, 36] (e.g., number of occurrences of specific activities), which usually capture frequent or expected activities. Conversely, there may be anomalous activities that relate to learning as well, which are—by definition—unexpected and thus difficult to discover. In this paper, we propose and evaluate a generalizable approach to reveal anomalous activities by examining the prediction errors of neural networks.

Analyzing anomalies requires defining them, which may be difficult. Manual examination and inference based on expert knowledge is one possible approach for discovering specific constructs in data [30]. However, defining and determining anomalies is time-consuming and constrained to the limits of expert knowledge. In statistics and data mining, anomalous activities refer to data deviating from patterns exhibited by the majority of data [13, 29]. In this perspective, anomalous activities are those where, given a sequence of activities, the activities that followed are unexpected—similar to definitions for time series data [26, 42]. Anomalies in this definition indicate deviations from predictable learning strategies [14], and thus may be the result of deviations from common learning strategies or from the ways in which instructors expect students to go through course materials. Consequently, a method to discover students' anomalous learning behaviors might inspire changes to our understanding

of e-learning strategies and could eventually help refine the design of learning experiences.

We approach this problem by training a self-supervised neural network to learn typical activity sequences, then detect anomalies based on the prediction errors. We demonstrate one aspect of the usefulness of our method by exploring the correlation between students' anomalous actions and students' learning outcomes. We further contribute to this problem by understanding how humans perceive anomalous activities and whether those perceptions are aligned with proposed approach.

## 2. RELATED WORK

In this section, we first discuss the concept of anomalies and provide an overview of applications of anomalies to highlight the potential for work in this area with educational data (section 2.1). We then investigate existing research on log data from learning management systems (section 2.2) to show the importance of behavioral data and how our approach contributes to related work in this area.

### 2.1 Anomalies

Anomalies are generally understood as rare data that do not conform to preconceptions or expectations derived from the majority of data [7]. Anomalies can be identified with statistical and machine learning techniques, and in various types of data, such as images and time series data [34, 3, 37, 26, 42]. However, in the context of students' behavioral sequences, anomalies are relatively poorly understood.

In image and video data (outside of educational contexts), anomalies refer to a set of features that are not expected, which provides context for how anomalies are defined and detected in general. For time series data, the data are linearly ordered and the definition of anomalies may differ as a result. A particular data value could be an anomaly in a specific context, and might be considered typical (not anomalous) in other contexts. Malhotra et al. [26] and Zhang et al. [42] leveraged prediction errors as an indication of anomalies.

We are aware of only one study that focused on anomalies in education-related sequential data [37]. They considered response time as an indicator of anomalous learning. After plotting the sequence of response times, the authors derived a posterior predictive distribution and regarded the learners as anomalous when they had an unusually high or low response time. However, time spent is not the only way in which actions might be anomalous; moreover, unusually high or low response times might actually be expected for some students when considered in the context of their previous behaviors.

An alternative way to distinguish anomaly versus normal actions is to analyze them in the context of a student's sequence of behaviors, which is the approach we take in this paper.

### 2.2 Data Mining in Log Activities

In recent years, there has been an increasing interest in analyzing log activities from e-learning environments. Re-

searchers have done a large number of tasks that try to understand students' behaviors, academic performance, and learning processes [10, 1, 28, 38, 39].

Much of the work [9, 44] on data-driven discovery in education focuses on extracting frequent sequential patterns that are common and thus may characterize the behaviors of students from a specific group or across an entire dataset. In contrast, our focus is on behaviors that distinguish students from their peers. Some existing works [30, 15, 19] build behavior models that incorporate relationships between past activities and current activities, or past state and current state, as we do in this study. Other works that rely on log files have explored connections between students' actions and high-level learning information. One research direction is to detect students' learning behaviors or learning preferences [25, 32, 41] as evident in logged activity data, with the goal of enabling personalization of learning experiences after identifying students' needs and preferences.

The methods discussed above do not directly examine anomalies; they rely on extracting features from log files to discover connections between those features and student outcomes or states, or to explore properties of the learning domain and task itself. However, these studies do show that the behavioral data reflects various states of students, and is thus a promising area for further exploration, such as with respect to anomalous behaviors.

## 3. METHOD

In this section, we describe the data used in our study and present the methods used to answer each of the research questions stated in the Introduction.

### 3.1 Dataset

We analyzed the Open University Learning Analytics Dataset (OULAD) [24] in our experiments. The data included in OULAD were collected from 2013 to 2014. The dataset contains information about 22 sections of 7 different courses (labeled *A* through *G*), including 32,593 students and their aggregated interactions with an LMS in terms of per-day counts of different types of actions.

We combined multiple sections of the same courses assuming that different sections of each course should be relatively similar. Based on the frequency of each activity, we observed that some activities rarely happened and appeared to be less meaningful for understanding student behaviors. In this work, we aim to detect anomalous activities in a given context instead of mining activities that happen rarely. We decided to group extremely uncommon interaction activities to simplify analysis and interpretation, though exploring the extremely rare events is one possible area for future work. Thus, we chose a threshold (i.e., less than twice per student on average) and grouped all interaction activities less frequent than the threshold into an "other" category. More information regarding specific actions can be found in existing work with this dataset (e.g., Figures 2 and 4 in [24], Table 2 in [21]).

### 3.2 Anomaly Detection

As discussed above, an "anomaly", generally speaking, refers to an unusual event. In our task, we formalized "typical"

events as predictable activities given a previous action sequence. In contrast, anomalies are students’ activities that do not conform with predictions. Anomaly detection included two steps: (1) use machine learning to model typical activities and (2) measure the model’s prediction errors. In particular, we trained a self-supervised sequential neural network to model activity sequences, leaving poorly-predicted activities as anomalies.

Our prediction model consists of three layers (though in principle the model could be expanded for datasets with more complex inputs): (1) the encoding layer, which is used for representation generation; (2) a sequential layer (e.g., convolution, recurrent), which is used for feature extraction; and (3) a fully connected layer with sigmoid activation to predict the next step in the sequence. In our experiments, we split the dataset into train and test sets with a ratio of 9:1. We ran experiments on two models: one with a convolutional layer for feature extraction, as described above, and an alternative model based on long short-term memory (LSTM) instead. The model takes three sequential actions, predicts the following action, and convolves over time. We set the kernel width of the convolutional layer to 3 and the number of filters to 20. For the LSTM model, we used a similar configuration (i.e., 20 LSTM cells). We trained models for 50 epochs with batch size 32. We used Adam as the optimizer [22] with a .004 learning rate for all seven courses after tuning the rate from .001 to .01 on course *A*.

To compute prediction error, we calculated the difference between actual actions and predicted actions for each timestamp. For each student, we computed the L2 loss (mean squared difference) between predicted actions and actual actions of each timestamp in the test set as an indicator of how well at each point a student conformed to expected behaviors. Thus, for a student activity sequence of length  $l$ , the error between actual and predicted action sequences can be represented by an  $l$ -length sequence, which has  $d$  dimensions (one for each action type).

### 3.3 Correlation with Grade

A common approach for predicting students’ outcomes is to engineer features from students’ activities that are believed to have some relationship to outcomes [4]. Analogously, we expected that a student’s activity typicality (anomalous versus typical actions) directly relates to their learning outcomes if anomalous behaviors are evidence of adapting behaviors (e.g., via self-evaluation) or the opposite. We tested this hypothesis by calculating correlations between anomaly loss features and students’ outcomes.

To model the relationship between activity typicality and student’s outcomes, we represented each student with a  $d$ -dimensional vector where each element in the vector is the aggregated error for a specific action. That is, we defined *anomaly loss* as the aggregated error for each possible action for each student. We computed anomaly loss by aggregating error from anomaly detection at the student level by calculating the mean for each possible action. We further determined which actions were most important by training random forest regression with 25 trees to predict students’ outcomes from the anomaly loss. During whole process, we conducted experiments only on the test set in the dataset

we used to model students’ activities. We further split the test set into train and test subsets randomly in a 2:1 ratio, ensuring that data from each student appeared in only one subset or the other. The usefulness of anomaly loss for each action can then be calculated from the feature importance values of the random forest regressor.

### 3.4 Human Perceptions of Anomalies

Anomaly detection is mostly rooted in the statistics and machine learning communities [2, 7, 14]. Whether or not the anomalies detected by the proposed approach align with human intuition is still unclear. Thus, we conducted a small survey in which we asked four people with data mining experience to make their own assessments of anomalous student behaviors. We selected 10 sequences that were representative from the top 5% of most anomalous sequences, and 10 sequences from the 5% most typical. We presented 20 sequences (half typical activities, half anomalous activities), descriptions of the types of activities, and asked the participants (i.e., “coders”) to determine whether the activities that happened on day 4 were typical or anomalous given three previous days of activities. After they finished coding the activities, we asked participants to provide insight into their coding strategy and their perceptions of what an anomaly is. Specifically, we asked them to describe how they perceive anomalies versus typical activities.

## 4. RESULTS AND DISCUSSION

### 4.1 Behavioral Prediction Models

The losses of CNN and LSTM models indicated that they worked approximately equally (mean of loss of LSTM was 1.3% higher; details can be found in the Appendix) well for feature extraction and the behavioral prediction task. We focused on the CNN model alone in the remainder of analyses for the sake of simplicity, since the two models were similarly accurate.

### 4.2 Correlation with Grade

In this section, we present the results of the analysis comparing anomalies and grade.

For all courses, predicted grade—from a random forest model with anomaly-based features—had a substantial correlation (mean  $r = .399$ ,  $SD = .073$ ) with actual grade. Correlations ranged from  $r = .295$  (for course *D*) to  $.494$  (for course *G*) with all  $p$ -values less than .05, indicating that correlations were positive across courses; even the lowest correlation indicated a moderate relationship between course grades and predictions made based on anomalies. Thus, we conclude that anomalies are important to investigate since they relate to students’ academic outcomes.

To further explore which types of anomalous actions were most related to students’ course grades, we analyzed feature importance of the random forest model. We used course *A* as an example. Feature importances in Table 1 show that the top five most important actions were *exam*, *other*, *oucontent*, *resource*, and *gap*. These actions were not necessarily the most common, yet still important because of their role in learning. Anomalous exam-related actions, in particular, explained over half of the model’s feature importance (which sums to 1 in a random forest). Other important anomalous

actions may relate to self-regulated learning behavior; for example, *gap*-related anomalies may indicate irregular course participation.

**Table 1: Importance for each feature in random forest grade prediction. Top five important actions shown in bold. Detailed information of actions can be found in existing work with this dataset (e.g., Figure 2 and 4 in [24], Table 2 in [21])**

Action	Feature Importance	Action	Feature Importance
Exam	<b>.596</b>	Quiz	.001
Forumng	.020	Register	.006
Gap	<b>.064</b>	Resource	<b>.054</b>
Homepage	.018	Subpage	.008
Other	<b>.085</b>	Transfer	.012
Oucontent	<b>.073</b>	Unregister	.022
Ouwiki	.000	Url	.040

### 4.3 Results of Anomaly Detection

Self-supervised neural networks, such as the model we trained in this study, learn the conditional probability distribution of possible elements in a sequence [16, 27, 11]. In our anomaly detection framework in particular, the model learns the probability of each activity occurring given activities in the preceding three days. We provided example sequences of activities labeled by the proposed method in Appendix Table 4 in appendix. The criteria the model learns for predicting activities may be complex, but Appendix Table 4 does illustrate some reasonable high-level patterns learned by the model. For example, if the activities that occurred on day 4 were not consistent with preceding activities (either students performed many more or fewer than the previous activity pattern), they were tagged as anomalous. In contrast, if the activities on day 4 appeared in the previous days once or more, then they were usually labeled as typical. The conditional probability of an activity is just one way of defining anomalies, however, and may not align with human perceptions of what anomalies are. Thus, we also explored human perceptions of anomalies, as described next.

### 4.4 Results of Human Perception

For each pair of four human coders, we calculated Cohen’s kappa coefficient to measure their inter-rater agreement [8], as shown in Table 2.

**Table 2: Kappa coefficients among machine learning and human coders of anomalous vs. typical sequences.**

	Model	Coder 1	Coder 2	Coder 3
<b>Coder 1</b>	.00			
<b>Coder 2</b>	.60	.00		
<b>Coder 3</b>	.00	-.20	-.20	
<b>Coder 4</b>	.40	.20	.42	-.20

Inter-rater agreement results show that coder 2 and coder 4 agreed somewhat with each other and with the model: the kappa coefficient between coder 2 and the model was .60, between coder 4 and model it was .40, and between coder 2 and coder 4 kappa was .42. Conversely, both coder 1 and coder 3 had close to zero agreement with others: the mean kappa coefficients between coder 1, coder 3 and others

were .00 and -.13 respectively. Similarly, kappa coefficients among coder 1, coder 3, and the model were -.07, on average, indicating that coder 1 and coder 3 also did not agree with each other or with the model.

Coders’ perceptions of anomalies largely aligned with whether they agreed with each other and with the model: coder 2 and coder 4 determined anomalies from a more statistical perspective, while coder 1 and coder 3 mostly determined anomalies subjectively (they imagined whether or not the sequences were consistent with their own behaviors). In their descriptions of anomalies, they mentioned that if students had different activities (either fewer or much more activities) on day 4 than what happened on day 1 to day 3, they considered the activities on day 4 to be anomalous. For example, as coder 2 said:

*Coder 2: The first [criterion I used] is if the activities included many more than what I expected to be there from the past three days. For example, if on days 1-3 the user only went to the Homepage plus one other place, and on day 4 the user went to many different places, I considered that anomalous.*

Conversely, coder 1 suggested that if a student did not use the discussion forum on day 1 to day 3 but used it on day 4, then the activities on day 4 are anomalies. In addition, coder 1 thought if a student only watches content to prepare for an exam, then the exam happening on day 4 is an anomaly. Coder 3 decided the typicality of activities by linking them to his/her own experience. Coder 3 thought that classes rarely require a consistent effort on the same activities throughout any given week:

*Coder 3: If activity types appear to be too consistent in Days 1-3, I became doubtful that any activity in Day 4 would be a typical activity. From my experience taking college courses, classes rarely require a consistent effort on the same types of activities throughout any given week, so too much consistency made me more likely to believe that the activities on Day 4 were anomalous.*

## 5. CONCLUSIONS

Our goal in this study was to more deeply understand students’ behavior in web-based learning systems, specifically in terms of anomaly detection. We formally defined anomalies as unexpected activities given preceding activities and demonstrated that these anomalies are significantly related to student outcomes. We only tested our method with a selection of the previous three days’ activities as context for the next day’s activities prediction. However, larger fixed sequence intervals could be of interest. We further investigated if anomalies detected by our method aligned with human perceptions of anomalies, finding that the method indeed aligns with some conceptions of what anomalies are, though further research is needed to explore alternative conceptions. Ultimately, anomaly detection may lead to improvements in student modeling, activity recommendations, and even modifications of course materials and learning environments as researchers and teachers rely on methods like these to identify and address critical moments in learning processes.

## 6. REFERENCES

- [1] M. A. Abdullah. Learning style classification based on student's behavior in moodle learning management system. *Transactions on Machine Learning and Artificial Intelligence*, 3(1):28–40, 2015.
- [2] C. C. Aggarwal. An introduction to outlier analysis. pages 1–34, 2017.
- [3] S. Akcay, A. Atapour-Abarghouei, and T. P. Breckon. Ganomaly: Semi-supervised anomaly detection via adversarial training. In *Asian Conference on Computer Vision*, pages 622–637. Springer, 2018.
- [4] M. M. Ashenafi, G. Riccardi, and M. Ronchetti. Predicting students' final exam scores from their course activities. In *Frontiers in Education Conference (FIE)*, pages 1–9, 2015.
- [5] R. S. Baker and P. S. Inventado. *Educational data mining and learning analytics*. Springer, 2014.
- [6] W.-L. Chan and D.-Y. Yeung. Clickstream knowledge tracing: Modeling how students answer interactive online questions. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, pages 99–109, 2021.
- [7] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM computing surveys (CSUR)*, 41(3):1–58, 2009.
- [8] J. Cohen. A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, 20:37–46, 1960.
- [9] O. Dermý and A. Brun. Can we take advantage of time-interval pattern mining to model students activity? In *Proceedings of the 13th International Conference on Educational Data Mining*, pages 69–80. ERIC, 2020.
- [10] A. Dutt and M. A. Ismail. Can we predict student learning performance from lms data? a classification approach. In *Proceedings of the 3rd International Conference on Current Issues in Education*, pages 24–29. Atlantis Press, 2019.
- [11] S. R. Eddy. Hidden markov models. *Current opinion in structural biology*, 6(3):361–365, 1996.
- [12] E. Er, C. Villa-Torranó, Y. Dimitriadis, D. Gasevic, M. L. Bote-Lorenzo, J. I. Asensio-Pérez, E. Gómez-Sánchez, and A. Martínez Monés. Theory-based learning analytics to explore student engagement patterns in a peer review activity. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, pages 196–206, 2021.
- [13] R. Foorthuis. A typology of data anomalies. In *International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems*, pages 26–38. Springer, 2018.
- [14] R. Foorthuis. On the nature and types of anomalies: A review of deviations in data. *International Journal of Data Science and Analytics*, 12(4):297–331, 2021.
- [15] C. Geigle and C. Zhai. Modeling mooc student behavior with two-layer hidden markov models. In *Proceedings of the 4th ACM Conference on Learning at Scale*, pages 205–208, 2017.
- [16] K. Greff, R. K. Srivastava, J. Koutník, B. R. Steunebrink, and J. Schmidhuber. Lstm: A search space odyssey. *IEEE transactions on neural networks and learning systems*, 28(10):2222–2232, 2016.
- [17] J. Huang, A. Dasgupta, A. Ghosh, J. Manning, and M. Sanders. Superposter behavior in mooc forums. In *Proceedings of the 1st ACM Conference on Learning at Scale*, pages 117–126, 2014.
- [18] Y. Huang, N. G. Lobjcowski, J. E. Richey, E. A. McLaughlin, M. W. Asher, J. M. Harackiewicz, V. Aleven, and K. R. Koedinger. A general multi-method approach to data-driven redesign of tutoring systems. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, pages 161–172, 2021.
- [19] P. Hur, N. Bosch, L. Paquette, and E. Mercier. Harbingers of collaboration? the role of early-class behaviors in predicting collaborative problem solving. *Proceedings of the 13th International Conference on Educational Data Mining*, pages 104–114, 2020.
- [20] A. Jalal and M. Mahmood. Students' behavior mining in e-learning environment using cognitive processes with information technologies. *Education and Information Technologies*, 24(5):2797–2821, 2019.
- [21] L. Jiang and N. Bosch. Predictive sequential pattern mining via interpretable convolutional neural networks. In *Proceedings of the 14th International Conference on Educational Data Mining*, pages 761–766. ERIC, 2021.
- [22] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *Proceeding of the 3rd International Conference on Learning Representations*, 2015.
- [23] J. S. Kinnebrew, K. M. Loretz, and G. Biswas. A contextualized, differential sequence mining method to derive students' learning behavior patterns. *Journal of Educational Data Mining*, 5(1):190–219, 2013.
- [24] J. Kuzilek, M. Hlosta, and Z. Zdrahal. Open university learning analytics dataset. *Scientific data*, 4(1):1–8, 2017.
- [25] M. P. P. Liyanage, L. G. KS, and M. Hirakawa. Detecting learning styles in learning management systems using data mining. *Journal of Information Processing*, 24(4):740–749, 2016.
- [26] P. Malhotra, L. Vig, G. Shroff, and P. Agarwal. Long short term memory networks for anomaly detection in time series. In *Proceeding of European Symposium on Artificial Neural Networks, Computational Intelligence, and Machine Learning*, volume 89, pages 89–94, 2015.
- [27] T. Mikolov, M. Karafiát, L. Burget, J. Cernocký, and S. Khudanpur. Recurrent neural network based language model. volume 2, pages 1045–1048. Makuhari, 2010.
- [28] B. Motz, J. Quick, N. Schroeder, J. Zook, and M. Gunkel. The validity and utility of activity logs as a measure of student engagement. In *Proceedings of the 9th International Learning Analytics and Knowledge Conference*, pages 300–309, 2019.
- [29] M. Nassir. Anomaly detection using principles of human perception. *arXiv preprint arXiv:2103.12323*, 2021.
- [30] L. Paquette and R. S. Baker. Comparing machine learning to knowledge engineering for student behavior modeling: a case study in gaming the system.

*Interactive Learning Environments*, 27(5-6):585–597, 2019.

- [31] Y. Psaromiligkos, M. Orfanidou, C. Kytagiias, and E. Zafiri. Mining log data for the analysis of learners’ behaviour in web-based learning management systems. *Operational Research*, 11(2):187–200, 2011.
- [32] S. Rajper, N. A. Shaikh, Z. A. Shaikh, and G. A. Mallah. Automatic detection of learning styles on learning management systems using data mining technique. *Indian Journal of Science and Technology*, 9(15):1–5, 2016.
- [33] F. Rodriguez, H. R. Lee, T. Rutherford, C. Fischer, E. Potma, and M. Warschauer. Using clickstream data mining techniques to understand and support first-generation college students in an online chemistry course. In *Proceedings of the 11th International Learning Analytics and Knowledge Conference*, pages 313–322, 2021.
- [34] W. Sultani, C. Chen, and M. Shah. Real-world anomaly detection in surveillance videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6479–6488, 2018.
- [35] O. M. Teodorescu. Building test recommender systems for e-learning systems. In *Proceedings of the 13th International Conference on Educational Data Mining*, pages 810–814, 2020.
- [36] N. Tomasevic, N. Gvozdenovic, and S. Vranes. An overview and comparison of supervised data mining techniques for student exam performance prediction. *Computers and Education*, 143:103676:1–18, 2020.
- [37] M. Ueno. Online outlier detection system for learning time data in e-learning and it’s evaluation. *Proc. of Computers and Advanced Technology in Education (CATE2004)*, pages 248–253, 2004.
- [38] S. Van Goidsenhoven, D. Bogdanova, G. Deeva, S. v. Broucke, J. De Weerd, and M. Snoeck. Predicting student success in a blended learning environment. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge*, pages 17–25, 2020.
- [39] H. Wei, H. Li, M. Xia, Y. Wang, and H. Qu. Predicting student performance in interactive online question pools using mouse interaction features. In *Proceedings of the 10th International Conference on Learning Analytics and Knowledge*, pages 645–654, 2020.
- [40] T.-Y. Yang, C. G. Brinton, C. Joe-Wong, and M. Chiang. Behavior-based grade prediction for moocs via time series neural networks. *IEEE Journal of Selected Topics in Signal Processing*, 11(5):716–728, 2017.
- [41] F. Zehner, S. Harrison, B. Eichmann, T. Deribo, D. Bengs, N. Anderson, and C. Hahnel. The naep edm competition: On the value of theory-driven psychometrics and machine learning for predictions based on log data. In *Proceedings of the 13th International Conference on Educational Data Mining*, pages 302–312, 2020.
- [42] C. Zhang, D. Song, Y. Chen, X. Feng, C. Lumezanu, W. Cheng, J. Ni, B. Zong, H. Chen, and N. V. Chawla. A deep neural network for unsupervised anomaly detection and diagnosis in multivariate time

series data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 1409–1416, 2019.

- [43] W. Zhang, X. Huang, S. Wang, J. Shu, H. Liu, and H. Chen. Student performance prediction via online learning behavior analytics. In *2017 International Symposium on Educational Technology (ISET)*, pages 153–157. IEEE, 2017.
- [44] Y. Zhang and L. Paquette. An effect-size-based temporal interestingness metric for sequential pattern mining. In *Proceedings of the 13th International Conference on Educational Data Mining*, pages 720–724, 2020.

## APPENDIX

In the appendix, we provide details of our experiments. We include the losses of our behavior model for all courses in Appendix Table 3. We also introduce samples of anomalous activities and typical activities labeled by our method in Appendix Table 4.

**Table 3: Loss (binary cross-entropy) of the behavioral prediction model based on different architectures.**

Loss	Course A	B	C	D	E	F	G
LSTM	0.0039	0.0025	0.0028	0.0037	0.0035	0.0041	0.0026
CNN	0.0037	0.0024	0.0029	0.0037	0.0034	0.0042	0.0025

**Table 4: Examples of activities labeled by our proposed approach The last column refers to the typicality of activities on day 4 given the activities from day 1 to day 3, which are labeled by the proposed approach.**

Activities for each day				
Day 1	Day 2	Day 3	Day 4	Day 4 Type
Homepage	Forum	Forum	Exam	Anomalous
	Homepage	Homepage	Forum	
	Content		Homepage	
			Content	
			Resource	
			Subpage	
			URL	
Forum	Forum	Forum	Gap	Anomalous
Homepage	Homepage	Homepage		
Content	Content	Content		
Subpage	Subpage	Subpage		
URL	URL	URL		
Gap	Homepage	Gap	Exam	Anomalous
	Content			
	Subpage			
Forum	Homepage	Homepage	Homepage	Typical
Homepage				
Content				
Exam	Gap	Forum	Gap	Typical
		Homepage		
		Content		
		Resource		
		Subpage		
Homepage	Homepage	Homepage	Homepage	Typical
			Content	
			Subpage	
			URL	