The Feasibility of Utilizing ChatGPT in Learning Analytics for the Identification of At-Risk Students

Zhi-Qi Liu[1], Hsiao-Ting Tseng[2], Owen H.T. Lu[3]

[1] *International College of Innovation, National Chengchi University, Taiwan*
[2] *Department of Information Management, National Central University, Taiwan*

Abstract

The value-added applications of ChatGPT occur in many fields. Cooperation with ChatGPT has gradually become inevitable. This study aims to explore the potential of ChatGPT in the field of learning analytics, with a specific focus on predicting risk students while tackling prevalent challenges in learning analytics. Traditionally, learning analytics classification tasks have relied on machine learning models, leading to issues related to model interpretability and tailor learning suggestion generation. By utilizing the LBLS467 learning behavior dataset, experimental findings with ChatGPT-4 reveal its potential as a fundamental and accessible tool. While occasional performance variations are noted, ChatGPT holds promise as an alternative approach for basic at-risk student prediction within learning analytics. This study paves the way for further exploration of ChatGPT's potential in enhancing student support mechanisms and improving educational outcomes.

Keywords

Learning analytics, ChatGPT, Risk student prediction

1. Introduction

With the popularization of technology in recent decades, on-line learning platform such as Google classroom had become increasingly popular. During the Covid-19 pandemic, the shift to remote teaching had expanded the utilization of on-line learning environment in many aspects. An advantage of online learning environments is their ability to comprehensively record students' study habits, offering valuable data for learning analytics (LA). LA has recently become a necessity in the educational environment, for example, research has demonstrated how LA works on two Japanese universities to support education and learning [1]. LA is the interpretation and analysis on students' learning behavior data that aims to understand their learning progress, detect potential issues, and formulate interventions to improve education [2]. Predicting students' academic performance is a crucial task in LA because it enables teachers to offer tailored assistance to those who are unable to catch up with the class while conserve time and resources and make sure students are receiving helpful and appropriate support.

Typically, LA for predicting at-risk students is carried out through machine learning (ML) models or statistic methods. However, applying these approaches for risk student prediction may require certain level of advanced knowledge in the field. Nonetheless, technology advancements have expanded the range of options available to benefit needed users. Among various technologies, Artificial intelligence (AI) had become one of the latest tools that can efficiently and effectively help humans to deal with a variety of tasks. Among all the AI, ChatGPT is considered as one of the most popular and powerful, one known for its successful application to a wide range of domains, e.g., healthcare, translation, etc. [3]. Thus, it can be inferred that ChatGPT could likely be make LA and risk student prediction work better than previous. As an AI chatbot, ChatGPT's user-friendly natural language interface can lower the barriers to adopt LA techniques, reducing the proficiency required for their implementation. If ChatGPT is able to achieve excellent performance on making risk student prediction, it can possibly become a more convenient and easier way to assist educators in various discipline to do LA and provide necessary assistants to at-risk students.

To confirm the role and value of ChatGPT in LA is the core task of this study. In this study, experiments will be conducted using a set of educational datasets referred to as LBLS467

LAK-WS 2024: Joint Proceedings of LAK 2024 Workshops, March 18–19, Kyoto, Japan

(Learning Behavior Learning Strategy 467). This dataset will be processed by ChatGPT-4 to make prediction on who are the risk students. Then, their performance will be evaluated and compared by calculate the accuracy. The study aims to evaluate ChatGPT's capability in risk student prediction, addressing the following research questions:

- **RQ1:** How do ChatGPT make prediction?
- **RQ2:** How accurate is the prediction result from ChatGPT?

2. Literature Review

2.1. Previous studies and application of ChatGPT in the realm of education

In the realm of education, previous studies have mainly focused on the impact of ChatGPT on student's learning behavior, academic integrity concerns [4], and discussion about how course instructors can response to the rapidly developing of technology [5]. Since the introduction of ChatGPT in November 2022, it had significantly change various domains including education with its outstanding capability in handling a variety of text-based tasks. On one hand ChatGPT has opened up the possibility to integrate AI into education and enhance student learning, such as easily and quickly organized information for students or provide instructors course materials [6]. On the other hand, it also raises concerns regarding the misuse of AI generated content such as students using ChatGPT to write their homework, which can lead to unethical and unlearning [6].

Nevertheless, few studies have explored ChatGPT's application in LA, possibly due to its text-based chatbot nature, which is better suited for tasks like text generation and summarization than numerical data analysis. Additionally, instructors and researchers may prioritize addressing AI misuse which is a more immediate concern over LA. Still, LA serves as a valuable long-term resource for course strategy that is worth investing in. It's also crucial to acknowledge the potential of applying ChatGPT to various domains including LA.

2.2. Inclusion of ChatGPT in learning analytics and risk student predication

Despite the shortage of studies on the implementation of ChatGPT in LA, some existing research has involved ChatGPT in LA for various purposes. Research has pointed out the need of interpreting the internals of predictive analytics and provide tailored advice according to the analytics result to at-risk students [7]. Within the research, the analytical method can be broadly categorized into predictive and prescriptive analytics. The role of ChatGPT in this study is in the final step to convert the prescriptive feedback into natural language to provide at-risk students with human understandable advices [7]. Despite the inclusion of ChatGPT in this study, it remains that this text-based chatbot is used for task related text generation rather than doing data analyze, feature selection, or prediction.

Previous study has been conducted on how ChatGPT can become a student-driven education technology and how it can possible be apply to LA [8]. It mentions the strength of ChatGPT to interpret and analyze text-based data which can be a valuable technique when it comes to analyze qualitative educational records. However, this study only brought up the concept of utilizing ChatGPT to address the deficiency in qualitative analyze in existing LA technique without having further related experiments.

LA and risk student prediction in the past were mainly conducted using machine learning models or statistical analysis [9], which has the limitation in handling text data. While student's thoughts can also be an important feature in risk student prediction, integrating ChatGPT with the current LA techniques can broaden the source of data for analysis. Moreover, if ChatGPT performs comparably to current methods in analyzing numerical data, it could offer educators and researchers a powerful and convenient LA tool.

2.3. Previous use of ChatGPT to do data analyze and prediction

Machine learning models can assistant human with a variety of task, however, one common issue of it is the lack of transparency and interpretability, hence, highlight the importance of

Explainable AI such as SHAP which aims to offer explanations for the predictive methods of machine learning models [10].

In the study conducted in 2023, ChatGPT was employed to predict stock market movements using news headlines, which is text-based data, and the finding revealed that ChatGPT actually outperformed traditional sentiment analysis methods [11]. As a text-based chatbot, ChatGPT excels in understanding and delivering human-readable text messages. In the study on stock market predictions, researchers improved model interpretability by having ChatGPT provide brief interpretations of predictions. While this study focused on stock market movements using news headlines, it's crucial to recognize ChatGPT's strength in text-based data analysis and its potential for interpreting analytical results. Model interpretability is vital in risk student prediction, where the goal is to provide personalized assistance to at-risk students.

3. Methodology

3.1. LBLS467 dataset introduction

This dataset gathered the learning data from nine of programming classes with total 467 students from 2020 to 2022. The participants were all university students from non-computer science-related departments [12]. It includes two kinds of student’s learning behavior. The first one is form Bookroll, which is an online learning platform that can record student’s behaviors such as add bookmark or add marker [13]. The second one is student’s VisCode activities, which include the code length, the time they send here coding, and types of error them encountered [14]. In addition to capturing learning behavior, this dataset also includes learning strategy data as survey responses rated on a scale of 1 to 5, covering aspects such as students' strategy inventory for language learning (SILL)[15], students' self-regulated learning (SRL)[16] measurement results, and their SRL motivation.

3.2. Data preparation

The process of data preparation before getting into the experiment is depicted in Fig.1. Note that we define at risk students as students whose scores are lower than the Q1 score in their class. The Q1 score of each class is presented in Table1.

Table 1
Q1 score of each class

	a	b	c	d	e	f	g	h	i
Q1 score	62.0	73.0	72.75	77.0	85.0	75.5	76.0	71.0	81.0

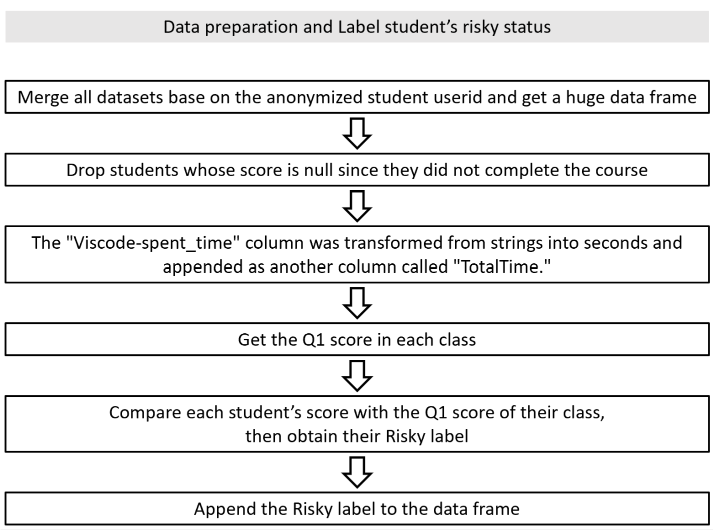

Figure 1: Data preparation and Label student’s risky status

Following data preparation, data frames were extracted base on the require number of features and data size and then split into training set and testing sets, this process is presented in Fig.2. Colum “Score” was not included in both the train test data frames because when making actual risk student prediction, the score of students is unknown.

The first stage feature extraction here was conducted by human researchers for the purpose to test the performance of ChatGPT handling data frames with different number of features. Features with more 0 values indicates fewer students are contributing data to this feature. Therefore, we assume features with lots of 0 value will have less effect on the prediction result of risk student. The feature extraction was conducted by setting up thresholds. Dropping features with more than x% of the value in the feature are 0.

The three feature numbers (9, 45, 80) were chosen for the below reasons. 9: Is the minimum number of features can be obtained with this threshold. Features that have more than 0.5% of their values as zeros were dropped. 80: Is the maximum number of features in this data frame, which included 26 features in the ‘br.csv’ file, 51 features in the ‘viscode.csv’ file, ‘TotalTime’ and ‘Risky’ column we appended and the ‘class’ column. The ’userid’ and ‘score’ columns was excluded from the training dataset as they are not relevant student’s risky status and ‘score’ is only used for labeling purpose. 45: Is approximately the number of features in between 9 and 80. If more than 81.30% of the values in a feature are zeros, that feature will be dropped. This middle point was chosen for the purpose to better demonstrate the change of accuracy among the numbers of feature.

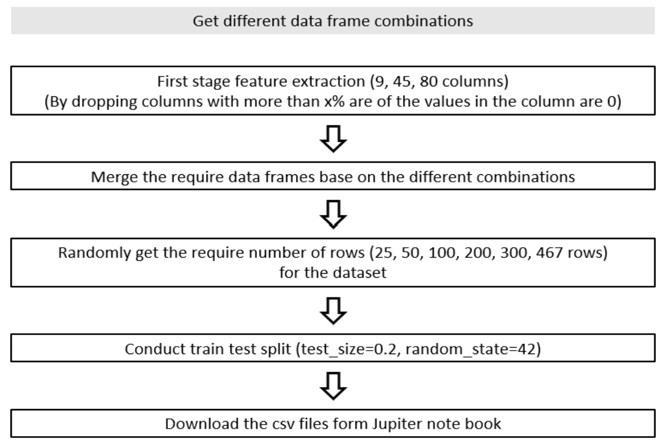

Figure 2: Get different data frame combinations

3.3. Experiment with ChatGPT

ChatGPT-4 website was chosen for conduct the experiment because it accepts bigger amount of input data compared with previous ChatGPT versions and ChatGPT-4 had released a new function for data analysis by just upload the files. In addition, using website can keep the conversation with ChatGPT in a more organized way. Detail process of the experiment and prompt is showed in Fig.3. The performance of ChatGPT doing risk student prediction is evaluated by accuracy with the below equation. With True Positives (TP) are instances when ChatGPT correctly predict a risk student as “Yes”; True Negatives (TN) are instances when ChatGPT correctly predict a non-risk student as “No”; False Positives (FP) are instances when ChatGPT incorrectly predicted a non-risky student as “Yes”; False Negatives (FN) are instances when ChatGPT incorrectly predicted a risky student as “No”.

$$\text{Accuracy} = \frac{(\text{TP} + \text{TN})}{(\text{TP} + \text{FP} + \text{TN} + \text{FN})} \tag{1}$$

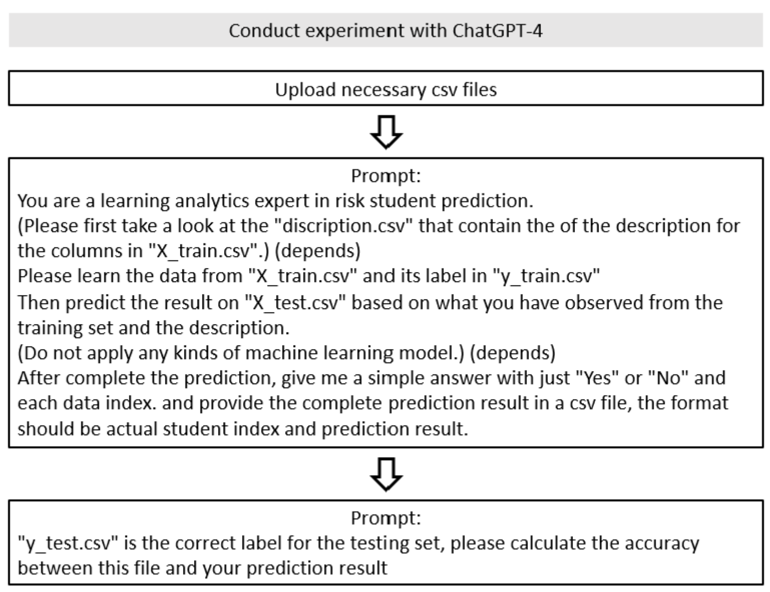

Figure 3: Conduct experiment with ChatGPT-4

4. Results and Discussion

4.1 Reply RQ-1: How do ChatGPT make prediction?

Upon received the below prompt and the uploaded files, ChatGPT will go through the process depicted in Fig.4. In this process, depending on the prompt, predictions would be conducted with or without the use of ML model, as illustrated in Fig. 5 and Fig. 6 respectively.

ChatGPT really insist that the prediction should be conducted using ML model, if didn't specified in the prompt, it will always apply machine learning model, with the Random Forest Classifier being the preferred choice in over 95% of cases. As ChatGPT explained, the choice of the Random Forest Classifier is based on its popularity and its robustness against overfitting, as well as its ability to effectively handle a diverse feature type. Occasionally, ChatGPT would choose other ML models such as Logistic Regression model for its binary classify characteristic or Gradient Boosting Classifier because it can handle a mix of continuous and categorical variables.

On the other hand, with the limitation of not to use ML model, ChatGPT would use heuristic approach, simple statistical methods or logical reasoning instead to complete the prediction. Two method it applies frequently are comparing the mean value of certain features between risky and non-risky student or analyze the correlation between each features and student's risky status. In addition, ChatGPT will also take the distribution of risky and non-risky students in the training set into consideration. Mentioning there are lots of student being label as risky or non-risky.

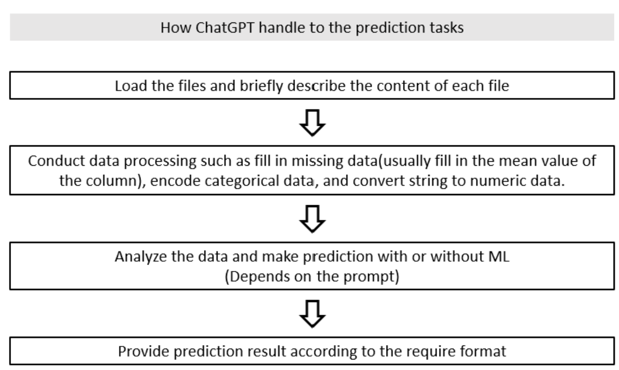

Figure 4: How ChatGPT handle to the prediction tasks

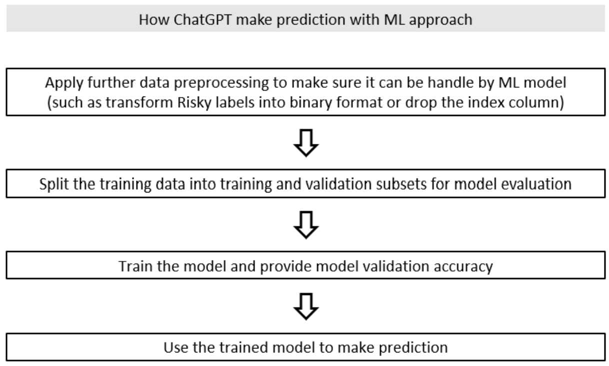

Figure 5: How ChatGPT make prediction with ML approach

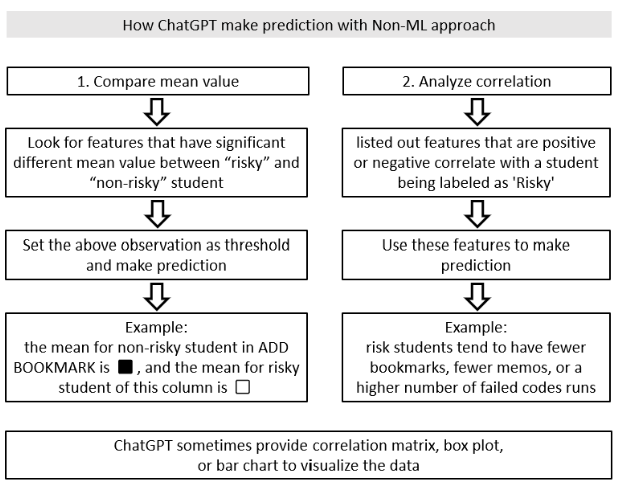

Figure 6: How ChatGPT make prediction with Non-ML approach

4.2. Reply RQ-2: How accurate is the prediction result from ChatGPT?

The prediction results presented in Table 2 focused on how the change in number of features and data size would affect the prediction accuracy. Because of the different characteristic of data frames in the LBLS467 dataset, the input data frames in Table 2 only include learning behavior data. The inclusion of both learning behavior and learning strategy data frames are presented in Table 3. The prediction results in both Table 2and 3 were analyzed along with the column name description file which contain the description for each feature. Furthermore, the data presented in both Tables were obtained through a repetitive process of executing Figures 2 and 3, each repeated five times, with the average accuracy recorded.

A discernible pattern from both Table 2 and is that ML approaches in general outperform Non-ML approaches, with the accuracy of the ML approach surpassing the Non-ML approach in all instances. In addition, the t-test results reveal statistically significant differences in mean accuracy between the ML and non-ML approaches across the 9, 45, and 80 features data frames. This outcome suggests that ChatGPT might not be good at logical reasoning and use heuristic approach or simple statistical methods to accurately make prediction.

Furthermore, the mean accuracy with ML approach across the data frames with 9, 45, and 80 features in Table 2 is relatively similar, with each hovering around the 75% mark. In contrast, the mean accuracy with Non-ML approach gradually increases as the number of features increases. However, there’s no clear pattern regarding how the data size would affect the accuracy.

When the learning strategy data frame is included, all the average accuracy values in Table 3 surpasses the mean accuracy in Table 2. This indicates that adding the learning strategy information can help improve the prediction accuracy, especially with the non-ML approach.

However, it’s crucial to note that ChatGPT's performance in both ML and Non-ML approaches can be variability, occasionally resulting in either exceptionally high or low accuracy especially with Non-ML approaches. This variability may be attributed to the selection of different ML models or heuristic approaches.

Table 2
Prediction results with different numbers of columns and data size

	9 features		45 features		80 features	
	ML	Non-ML	ML	Non-ML	ML	Non-ML
Size=25	84.00	48.00	80.00	28.00	76.00	64.00
Size=50	72.00	54.00	74.00	46.00	80.00	68.00
Size=100	83.00	44.00	71.00	51.00	77.00	50.00
Size=200	77.50	41.50	71.00	61.50	72.50	59.50
Size=300	75.67	56.98	75.67	60.00	77.33	61.00
Size=467	76.59	41.70	77.87	65.11	76.18	61.28
Mean	78.13	47.70	74.92	51.94	76.50	60.63
t-test result	P-value<0.05 statistically significant		P-value<0.05 statistically significant		P-value<0.05 statistically significant	

Table 3
Learning behavior and learning strategy features with all 467 rows

	ML	Non-ML
learning strategy + 9 features	79.57	72.34
learning strategy + 45 features	79.36	64.68
learning strategy + 80 features	83.83	67.45

5. Conclusion

Despite its occasional performance fluctuations, ChatGPT proves capable of serving as a basic and convenient tool for LA and fundamental risk student prediction. It offers flexibility by enabling the application of both ML and non-ML methods for prediction, which opens up the possibility for further research to explore different kinds of input data for LA. With traditional ML approach, ChatGPT typically achieves accuracy levels of around 70-80%. It simplifies the process by handling data processing, model training, and code execution automatically. While manual ML model training may yield higher and more stable accuracy, it demands time and expertise. When course instructors find ChatGPT's predictive performance acceptable, it becomes a convenient option for implementing a learning risk classifier.

Nevertheless, it is important to acknowledge certain limitation in the research. Firstly, the LBLS467 dataset was collected form programming course, characterized by a substantial presence of numerical coding learning records, which can significant differ from other subjects. Moreover, the prompts being use for both ML and Non-ML approaches are almost the same. Tailoring prompts for different approaches might provide a more precise description of the tasks and potentially lead to improved performance. Further research endeavors could explore the applicability of ChatGPT to predict at-risk students using data from different subjects, enhance accuracy level and performance stability through prompt modification, and conduct experiments with text-based data.

Acknowledgments

This study is supported in part by the National Science and Technology Council of Taiwan under contract numbers NSTC 112-2410-H-004 -063 – and NSTC 112 - 2636 - H - 008 - 005 -.

References

[1] Flanagan, B., Ogata, H., Learning analytics platform in higher education in Japan, Knowledge Management & E-Learning 10(4) (2018) 469-484.

[2] Siemens, G., Baker, R. S. d., Learning analytics and educational data mining: towards communication and collaboration, in: Proceedings of the 2nd international conference on learning analytics and knowledge, 2012, pp. 252-254.

[3] Ray, P. P., ChatGPT: A comprehensive review on background, applications, key challenges, bias, ethics, limitations and future scope, Internet of Things and Cyber-Physical Systems (2023).

[4] Sullivan, M., Kelly, A., McLaughlan, P., ChatGPT in higher education: Considerations for academic integrity and student learning (2023).

[5] Mills, A., Bali, M., Eaton, L., How do we respond to generative AI in education? Open educational practices give us a framework for an ongoing process, Journal of Applied Learning and Teaching, 6(1) (2023).

[6] AlAfnan, M. A., Dishari, S., Jovic, M., Lomidze, K., Chatgpt as an educational tool: Opportunities, challenges, and recommendations for communication, business writing, and composition courses, Journal of Artificial Intelligence and Technology, 3(2) (2023) 60-68.

[7] Susnjak, T., Beyond Predictive Learning Analytics Modelling and onto Explainable Artificial Intelligence with Prescriptive Analytics and ChatGPT, International Journal of Artificial Intelligence in Education (2023) 1-31.

[8] Dai, Y., Liu, A., Lim, C. P., Reconceptualizing ChatGPT and generative AI as a student-driven innovation in higher education, Procedia CIRP (2023) 84-90.

[9] Marwaha, A., Singla, A., A study of factors to predict at-risk students based on machine learning techniques, Intelligent Communication, Control and Devices: Proceedings of ICICCD 2018 (2020) 133-141.

[10] Lu, O. H., Li, A. L. L., Min-Jia, Bobea, Matthew, Huang, Anna Y.Q., Yang, Stephen J.H., Analyzing Student Programming Propensity with SHAP to Classify Future Performance (2023).

[11] Lopez-Lira, A., Tang, Y., Can chatgpt forecast stock price movements? return predictability and large language models, arXiv preprint arXiv:2304.07619 (2023).

[12] Lu, O. H., Huang, A. Y., Flanagan, B., OGATA, H., YANG, S. J., A Quality Data Set for Data Challenge: Featuring 160 Students' Learning Behaviors and Learning Strategies in a Programming Course, in: Proceedings of the 30th International Conference on Computers in Education, 2022.

[13] Ogata, H., Yin, C., Oi, M., Okubo, F., Shimada, A., Kojima, K., Yamada, M., E-Book-based learning analytics in university education, in: Proceedings of the International conference on computer in education (ICCE 2015), Asia-Pacific Society for Computers in Education, 2015, pp. 401-406.

[14] Lu, O. H. T., Huang, A. Y., Huang, J. C., Huang, C. S., Yang, S. J., Early-Stage Engagement: Applying Big Data Analytics on Collaborative Learning Environment for Measuring Learners' Engagement Rate, in: Proceedings of the 2016 International Conference on Educational Innovation through Technology (EITT), IEEE, 2016, pp. 106-110.

[15] Oxford, R. L., Burry-Stock, J. A., Assessing the use of language learning strategies worldwide with the ESL/EFL version of the Strategy Inventory for Language Learning (SILL), System, 23(1) (1995) 1-23.

[16] Zimmerman, B. J., Schunk, D. H., Self-regulated learning and performance: An introduction and an overview, Handbook of self-regulation of learning and performance (2011) 15-26.