



Clickstream Data Schema for Learning Analytics to Understand Learner Behaviour

Mustafa Mohammad Abed¹, Dalbir Singh²

¹Center for Software Technology and Management (SOFTAM), Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Malaysia, mustafa.softam@gmail.com

²Center for Software Technology and Management (SOFTAM), Faculty of Information Science & Technology, Universiti Kebangsaan Malaysia, Malaysia, dalbir@ukm.edu.my

ABSTRACT

Learning analytics has caught numerous researchers' attention in the last decade due to its ability to analyze, report, and process information produced by learners. Several research studies have examined learning analytics in terms of learner behavior to provide quality and efficient reporting for educators and higher learning institutions. In particular, the click pattern of the learner has been examined in terms of its significant impact on determining learner performance. Yet, the clickstream data schema that has been addressed in the literature was limited to click patterns focusing on event-based click patterns and time in the logging history. However, learner interaction logging history available in clickstream data schema for major online learning management systems holds great potential to reveal diverse and complex learner behavior patterns based on various significant learner behavior elements that are waiting to be discovered. The literature has shown that great interest has been given in understanding specific behavior produced by the learner through an online learning management system. Thus, this study aims to review previous related studies that could reveal possible learner click patterns, which could be considered as an indicator to predict the learner's performance and behavior. Potential learner interaction data that could be applied to produce a clickstream data stream has been outlined. As an implication, the contribution of the study could open up new frontiers that trigger the development of a learning analytics dashboard for various online learning platforms.

Key words: Clickstream Data Schema; Learning Analytics; Learner Interaction Data; Online Learning Platforms.

1. INTRODUCTION

Learning analytics is an accurate investigation of data about the process of learning. The purpose of the analysis is to determine behavioral patterns among learners and the educational interpretation of the resulting information after

the analysis. Learning analytics presents certain issues that are worthy of further investigation, such as technology and data management issues based on relevancy and requirement.

Second, there is a lack of data quality because of the need to filter useful data from a large amount of available data, and combining them in a way that users can benefit from them. Thirdly, the educational challenges need to be met to ensure a valid method for analysis, and this process involves the participation of educators in online learning [1]. Even though stakeholders can be guided on decisions to make through perceptions, accurate perceptions can only be created through the analysis of data. Thus, applying learning analytics for online learning could develop a learner-focused strategy and provide tools that enable continuous improvement for online teaching and learning strategy in higher learning institutions [2].

The analysis of visual learning often focuses on many issues, including issues related to performance level and activity completion, while other activities focus solely on learner behavior [3]. The visual nature and diagrams of mental models provide a detailed presentation of the learner's thinking [4]. The literature has shown significant interest in specific learner behavior patterns through the click patterns produced by the learner during online learning sessions. Within such studies, the click patterns have been examined in terms of video streaming [5], [6], or position of clicks [7] [8]. However, numerous patterns could be exploited from the clicks. These patterns would aid the identification of useful information, such as learner's progress and performance. Hence, there is an imperative demand to combine multiple learner behavior patterns to provide a proper assessment for the learner's progress and performance.

Nevertheless, due to its dynamic and numerous interpretation, data schema that contains meaningful learner interaction data based on click and time centered server log needs to be further investigated in terms of novel methods of extraction, triangulation, and visualization [8]. According to

the study of [9], clicks are not enough to evaluate the learner's behavior through any online learning system. Besides, clickstream data schema that has been addressed in the literature section was limited to click patterns focusing on event-based click patterns and time in the logging history. Indirectly, common clickstream data schema are focused on Click Sequence Model and Time-Based Model. However, a hybrid model that combines click types and click interarrival times [10] is worthy of being explored in the proposed study. It is also inclined with [11] that exploit various interactions in the clickstream data schema. This proves that clickstream data schema holds a huge potential for future investigation in terms of understanding specific learner behavior.

Therefore, in this study, it is proposed to explore clickstream data schema so that specific learner behavior pattern that can be articulated, and then visualize. Various related previous studies related to learning analytics that particularly focus on learner behavior pattern has been reviewed. The findings reveal those crucial methods and clickstream learner interaction data that can be utilized to visualize learner behavior patterns. Such fundamental understanding forms the foundation that differentiates the proposed study from others.

2. REVIEW OF LEARNER BEHAVIOUR PATTERN

In this section, the critical review was conducted on related studies: (a) Learning Analytics, (b) Learner Behaviour Pattern that leads to the discovery of the (c) research gap. For section (a) Learning Analytics, ten (10) studies, meanwhile section (b) Learner Behaviour Pattern, six (6) studies were examined. These studies are selected based on their relevancy and between the years of 2015-2020.

2.1 Learning Analytics

In the first study, a framework was proposed for learning analytics, particularly for computer science courses [12]. The proposed framework is aimed at determining the poor performance of the learners regarding their progress in terms of programming courses. The proposed framework has been adopted in a visual analytics system that includes information about the learners' progress. Besides, the system can predict the performance of learners based on their historical data and interactions with the courses.

In the second study, a learning analytics model was proposed [13]. The proposed model is aimed at analyzing the social forums and blogs in which the discussions conducted by learners are being captured. The proposed model was intended to examine the role of these discussions in terms of improving learning among the participants.

In the third study, the role of learning analytics was examined in terms of fulfilling higher education institutional strategic goals [14]. In fact, learning analytics, we're vastly engaged with the performance of learners by analyzing their progress. In this study, the researchers have addressed to what extent learning analytics would contribute toward satisfying the strategic goals of higher educational institutions. Therefore, a framework that engaged both learner performance following the strategic institutional goals has been proposed. The finding of such a study has revealed the essential importance of learning analytics toward accomplishing the institutional goals.

In the fourth study, the role of combining big data and learning analytics was addressed in terms of improving higher education in South Africa [15]. The researchers have analyzed the factors of big data and learning analytics in higher education. As a result, a framework has been proposed to describe the proper adoption of big data and learning analytics for South African higher education institutions. Similarly, the same researchers have proposed a conceptual framework (Matsebula, & Mnkandla 2017a) for adopting big data analytics in improving the quality of higher education. The researchers have acquired and analyzed educational data such as forums, social media, learning management portals, and other electronic data.

In the fifth study, the role of machine learning was studied in terms of learning analytics [16]. The researchers have proposed a model that describes the contribution of machine learning algorithms toward improving learning analytics. Such a model contains aspects that could be optimized by machine learning, such as data gathering, whereby learning analytics could depend mainly on historical data. Besides that, for the reporting aspect whereby machine learning would facilitate in determining the exact information that could be reported. Finally, the prediction aspect whereby machine learning would help in predicting outliers or interesting patterns related to learners' progress.

In the sixth study, a decision-making model was proposed based on learning analytics to help higher education institutions in terms of academic decision making [17]. The researchers have presented an architecture for their model, whereby learning analytics would be implemented. The proposed model has been tested in a public university and demonstrated its efficacy.

In the seventh study, the relation was studied between learning analytics of learners' behavior following the learner's success [18]. The researchers have analyzed several online activities such as learner access and their interactions. Although the finding of such a study has revealed that there is a minor correlation of learning analytics and learner's

success, yet it implies the importance of employing learning analytics in various aspects of higher education topology.

In the eighth study, a framework was proposed for workplace learning analytics in the engineering field [19]. The proposed framework has been depicted in a visual dashboard that analyzes the learner activities without exposing their private information. Such activities would be visualized and reported to the program managers who can take proper actions toward evaluating the learners' progress.

In the ninth study, a visualization model was proposed for learning analytics to enrich the interactions between educators and learners [20]. The proposed model has been depicted in a visualized dashboard that enables the

communications, discussions, and dialogue between the learners and their educators. Besides, the proposed model contains visual progress and comparisons between previous and current interactions, which facilitate personalization and continuous dialogue.

Finally, in the tenth study, a visualization model was proposed for learning analytics to improve the communications and interactions between the learners and their educators [21]. The proposed model has concentrated on the extent of how the educators would understand the flow of interactions performed by the learners. The proposed model has included significant indicators such as the ease of use, intention to use, and utility.

Table 1: Summary of learning analytics studies

Previous Study	Learning analytics task	Method	Uses of Data	Issues
[12]	Detecting the poor performance of learners	A framework for learning analytics	Learner progress, learner behavior, and historical data	<ul style="list-style-type: none"> - The use of data has not been evaluated in terms of its significance - This study has not examined the visualization in terms of learning analytics
[13]	Discussions in social forums	A model to examine the role of discussion following improving learning	Educational dialogue	<ul style="list-style-type: none"> - No visualization has been examined - Other information should be included, such as learner progress.
[14]	Relationship between learning analytics and accomplishing institutional strategic goals	A framework to combine both learner performance and institutional strategic goals	Learner progress and grades	<ul style="list-style-type: none"> - Visualization has not been included - Educator perspective has not been included
[22][15]	Relationship between big data and learning analytics in improving higher education	A conceptual framework for the proper adoption of big data in learning analytics	A vast amount of educational information including forums discussion, social media content and learning management	<ul style="list-style-type: none"> - Learner performance has not been examined - Educator information has not been included
[16]	The role of machine learning in learning analytics	A model combines machine learning and learning analytics	Learner information, interactions, and reports	Although this study has presented a prediction model for the poor performance of learners, yet, it did not have educator involvement in

				terms of notifying the poor performance.
[17]	Learning analytics in decision making	A model that depicts the factors of learning analytics in academic decision making	Learner progress, program managers' actions	Visualization has not been addressed in terms of decision making
[18]	Relationship between learning analytics and learner success	A model combines learning analytics and learner's accomplishments	Learner's behavior including learner's access and interactions, grade	Educator aspect has not been addressed
[19]	The gap of learning in education and practice for engineering courses	A framework for workplace learning analytics	Learner activities and program managers actions	Visualization has not been addressed
[20]	Improper communication and interactions between instructors and learners	A visualized model that facilitate the communications	Interactions between learners and instructors	<ul style="list-style-type: none"> - This model did not focus on the learner progress - This model did not consider an important factor regarding what kind of information to be visualized.
[21]	Improper communication and interactions between instructors and learners	A model that addresses the factors of the instructor's understanding through visualization	Interactions between learners and instructors	<ul style="list-style-type: none"> - This model did not focus on the learner's progress - This model did not consider an important factor regarding what kind of information to be visualized.

Based on the summary of learning analytics studies in table 1, it can be concluded that various types of learner interaction data have been utilized to predict and determine learner performance and behavior. Secondly, emphasis on visualization based on these learner interaction data has not been fully addressed, especially from the educator perspective, to assess learner performance and behavior. Thus, no uniform learning analytic approach exists that utilizes a standard set of learner interaction data and visualization techniques.

2.2 Learner Behaviour Pattern

Recently, more attention has been caught by researchers toward utilizing learner behavior in the context of learning analytics. According to the study of [23], the role of learning analytics was examined in the context of providing in-time and quality feedback for educators based on their learners'

progress. Such feedback is completely relying on learner behavior through the online learning environment. In particular, this has been conducted using a technique called clickstream, which is a web-based approach and intended to observe the behavior of online visitors. The researchers have used such a technique to observe the learner's click patterns in an online programming learning environment. The finding of this study has demonstrated a significant relationship between the click patterns of learners and their performance.

The same researchers have also proposed an extension of the previous study [8]. The researchers have added a tool to visualize the clicks patterns in which the educators would take advantage of the big data analytics to improve their feedback.

Learner's click patterns have revealed several factors that might have impacted the learner's performance through

blended or online learning environments [6]. The researchers have focused on the clicks that learners might accommodate while they were watching tutorial videos through the online learning system. Such clicks included 'Play' click, 'Forward seek' click, 'Backward seek' click, 'Pause' and 'Stop' clicks. Using the principle of component analysis, the researchers have concluded that the click patterns would contribute toward predicting learner performance.

Similarly, [5] have examined the role of click patterns of the learners in terms of the grades through the online learning systems. The researchers have focused on the click pattern categories such as the number of clicks related to reading per week, and the number of clicks related to video streaming per week. Using a case study, the researchers have identified a direct relationship between the click patterns and learner's grades.

Apart from the click patterns, different studies have focused on other learner behavior. For example, logging history has been addressed by [24], where a conceptual framework has been provided for analyzing learner behavior through online assessment systems. This study has focused on logging history and its impact on determining at-risk learner performance. Another online behavior has also been addressed by [25], whereby the researchers have examined the role of timely completion of training and assignments and their relationship toward learner success or failure through the online learning environment. The researchers have proposed an intervention model that enables the educator to alert and award learners based on the completion of their learning. As a result of this study, a significant relationship has been identified between the timely completion of assignments and learner success.

Table 2: Summary of learner behavior pattern

Previous Studies	Learning analytics task	Method	Uses of Data	Issues
[7]	In-time and quality feedback of learner performance	A Clickstream model to analyze the positions of learner clicks	Positions of clicks and number of times for such clicks	Click patterns have been limited to the positions
[8]	Visualize click pattern feedback	A model to visualize big data of click patterns	Positions of clicks and number of times for such clicks	Click patterns have been limited to the positions
[6]	Feedback of learner performance	Clickstream model to observe the patterns of video clicks	Video related clicks including Play, Forward seek, Backward seek, Pause and Stop	Click patterns have been limited to the video related clicks
[5]	Relationship between click patterns and learner grade	A visualized model based on Clickstream	Two main click patterns including reading clicks and video streaming clicks	Click patterns can be extended to include new behavior
[24]	Validating learner performance	A visualized model based on learner behavior	Logging history	Although logging history is a good indicator, other behaviors can be combined
[25]	Relationship between timely completion of assignments and learner success	A model to observe the learners behavior regarding assignment completion	Time spent to complete assignments	Other factors of learner behavior can be related to the learner's success

Based on the summary of the learner behavior pattern in table

2, it can be concluded that clickstream data schema has been utilized to visualize learner behavior patterns. However, the clickstream data schema is limited to learner interaction data,

which are derived from event-based click patterns and time. Thus, the clickstream data schema can be further exploited to extract the various type of learner interaction data that could be used to articulate specific learner behavior patterns.

3. CONCLUSION

As described in section (a) and (b), it has been proven that learner's click patterns can be considered as an indicator to predict the learner's performance and behavior. However, according to the study of [9], clicks are not enough to evaluate specific learner behavior through any online learning system. Besides that, other studies have also examined various factors of learner behavior, such as logging history and timely completion of assignments. Hence, there is an imperative demand to combine multiple learner interaction data to create clickstream data schema that could provide proper assessment for the learner's progress and behavior.

Therefore, in this study, it is proposed to examine clickstream data schema so that specific learner behavior pattern that can be articulated, and then visualize. Indirectly, it can reveal learner behavior elements which can be articulated to learner behavior pattern during online learning that eventually could benefit the learners, educators, and higher learning institutions. Such a concept forms the foundation for the formulation learning analytics model, as proposed in this study. It may also trigger the integration of learning analytic in other studies related the online learning [26] [27] [28] that could eventually benefit the educators and learners [29] [30], especially in the COVID-19 pandemic era [31] [32].

ACKNOWLEDGEMENT

The researchers would like to express their highest gratitude towards Universiti Kebangsaan Malaysia's research grant: GGP-2019-026.

REFERENCES

- [1] M. Guenaga and P. Garaizar, "From Analysis to Improvement: Challenges and Opportunities for Learning Analytics," *Rev. Iberoam. Tecnol. del Aprendiz.*, vol. 11, no. 3, pp. 146–147, 2016.
- [2] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telemat. Informatics*, vol. 37, pp. 13–49, Apr. 2019.
- [3] J. Kay and S. Bull, "New opportunities with open learner models and visual learning analytics," in *International Conference on Artificial Intelligence in Education*, 2015, pp. 666–669.
- [4] S. Mcneil, "Visualizing mental models : understanding cognitive design and development," pp. 73–96, 2015.
- [5] M. C. Goulden *et al.*, "CCVis: Visual analytics of student online learning behaviors using course clickstream data," *Electron. Imaging*, no. 1, 2019.
- [6] O. H. T. Lu, A. Y. Q. Huang, J. C. H. Huang, A. J. Q. Lin, H. Ogata, and S. J. H. Yang, "Applying learning analytics for the early prediction of students' academic performance in blended learning," *Educ. Technol. Soc.*, vol. 21, no. 2, pp. 220–232, 2018.
- [7] Á. Serrano-Laguna *et al.*, "Information Science and Applications (ICISA) 2016," *Comput. Human Behav.*, vol. 92, no. 4, pp. 27–32, 2018.
- [8] D. A. Filvà, M. A. Forment, F. J. García-Peñalvo, D. F. Escudero, and M. J. Casañ, "Clickstream for learning analytics to assess students' behavior with Scratch," *Futur. Gener. Comput. Syst.*, vol. 93, pp. 673–686, Apr. 2019.
- [9] E. Fincham, A. Whitelock-Wainwright, V. Kovanović, S. Joksimović, J.-P. van Staaldunen, and D. Gašević, "Counting Clicks is Not Enough: Validating a Theorized Model of Engagement in Learning Analytics," in *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, 2019, pp. 501–510.
- [10] and B. Y. Z. Gang Wang, Xinyi Zhang, Shiliang Tang, Christo Wilson, Haitao Zheng, "Clickstream user behavior models," *ACM Trans. Web*, vol. 11, no. 4, p. 37, 2017.
- [11] Q. Li, R. Baker, and M. Warschauer, "Using clickstream data to measure, understand, and support self-regulated learning in online courses," *Internet High. Educ.*, vol. 45, p. 100727, Apr. 2020.
- [12] D. Azcona, I. H. Hsiao, and A. F. Smeaton, "Personalizing computer science education by leveraging multimodal learning analytics," *Proc. - Front. Educ. Conf. FIE*, vol. 2018-October, pp. 1–9, 2019.
- [13] M. Adraoui, A. Retbi, M. K. Idrissi, and S. Bennani, "Social learning analytics to describe the learners' interaction in online discussion forum in Moodle," *2017 16th Int. Conf. Inf. Technol. Based High. Educ. Training, ITHET 2017*, 2017.
- [14] J. Jonathan, S. Sohail, F. Kotob, and G. Salter, "The Role of Learning Analytics in Performance Measurement in a Higher Education Institution," *Proc. 2018 IEEE Int. Conf. Teaching, Assessment, Learn. Eng. TALE 2018*, no. December, pp. 1201–1203, 2019.
- [15] F. Matsebula and E. Mnkandla, "A big data architecture for learning analytics in higher education," *2017 IEEE AFRICON Sci. Technol. Innov. Africa, AFRICON 2017*, pp. 951–956, 2017.
- [16] F. Sciarone, "Machine learning and learning analytics: Integrating data with learning," *2018 17th Int. Conf. Inf. Technol. Based High. Educ. Training, ITHET 2018*, pp. 1–5, 2018.

- [17] Y. Vanessa Niet, V. G. Diaz, and C. E. Montenegro, “Academic decision making model for higher education institutions using learning analytics,” *2016 4th Int. Symp. Comput. Bus. Intell. ISCBI 2016*, pp. 27–32, 2016.
- [18] B. P. Marques, J. E. Villate, and C. V. Carvalho, “Analytics of student behaviour in a learning management system as a predictor of learning success,” *Iber. Conf. Inf. Syst. Technol. Cist.*, 2017.
- [19] E. van der Stappen, “Workplace learning analytics in higher engineering education,” in *2018 IEEE Global Engineering Education Conference (EDUCON)*, 2018, pp. 15–20.
- [20] S. Charleer, A. Vande Moere, J. Klerkx, K. Verbert, and T. De Laet, “Learning analytics dashboards to support adviser-student dialogue,” *IEEE Trans. Learn. Technol.*, vol. 11, no. 3, pp. 389–399, 2018.
- [21] R. Paiva, I. I. Bittencourt, W. Lemos, A. Vinicius, and D. Dermeval, “Visualizing learning analytics and educational data mining outputs,” in *International Conference on Artificial Intelligence in Education*, 2018, pp. 251–256.
- [22] F. Matsebula and E. Mnkandla, “Information systems innovation adoption in higher education: Big data and analytics,” *Proc. - 2016 3rd Int. Conf. Adv. Comput. Commun. Eng. ICACCE 2016*, pp. 326–329, 2017.
- [23] D. Amo, M. Alier, F. J. García-Peñalvo, D. Fonseca, and M. J. Casañ, “Learning analytics to assess students’ behavior with scratch through clickstream,” *CEUR Workshop Proc.*, vol. 2188, pp. 74–82, 2018.
- [24] M. Wasif, H. Waheed, N. R. Aljohani, and S.-U. Hassan, “Understanding Student Learning Behavior and Predicting Their Performance,” pp. 1–28, 2019.
- [25] L. Horodyskyj, C. Mead, Z. Pardos, and A. D. Anbar, “Improving Student Outcomes Through Informed Use of Learning Analytics,” in *AGU Fall Meeting Abstracts*, 2018.
- [26] N. A. M. Z. Mercy Trinovianti Mulyadi, “MMORPG GAME FRAMEWORK BASED ON LEARNING STYLE FOR LEARNING COMPUTER NETWORKING,” *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 8, no. 1, pp. 63–77, 2019.
- [27] Z. M. Nursyahida Mokhtar, Amirah Ismail, “PRELIMINARY STUDY: FLOOD AWARENESS TRAINING PREPARATION USING SERIOUS GAMES,” *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 7, no. 2–2, pp. 13–26, 2018.
- [28] T. S. M. T. W. Hairulliza Mohd. Judi, Hanimastura Hashim, “KNOWLEDGE SHARING DRIVING FACTORS IN TECHNICAL VOCATIONAL EDUCATION AND TRAINING INSTITUTE USING CONTENT ANALYSIS,” *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 7, no. 2, pp. 11–28, 2018.
- [29] M. M. Mohd Salihan Ab Rahman, Nazlena Mohamad Ali, “COMELGETZ PROTOTYPE IN LEARNING PRAYERS AMONG CHILDREN,” *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 6, no. 1, pp. 115–125, 2017.
- [30] N. A. A. Salyani Osman, Noraidah Sahari@Ashaari, Nor Azan Mat Zin, Ratna Zuarni Ramli, “COURSEWARE USING COGNITIVE APPRENTICESHIP STRATEGY,” *Asia-Pacific J. Inf. Technol. Multimed.*, vol. 5, no. 2, pp. 1–18, 2016.
- [31] M. Diena Dwidienawati, Sri Bramatoro Abdinagoro, David Tjahjana, Dyah Gandasari, “Forced Shifting to e-Learning during the COVID-19 Outbreak: Information Quality, System Quality, Service Quality, and Goal Orientation influence to e-Learning Satisfaction and Perceived Performance,” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 2, pp. 1518–1525, 2020.
- [32] S. N. Wan Hassan, W.A.S., Ariffin, A., Ahmad, F., Sharberi, S.N.M., Nor Azizi, M. I., Zulkiflee, “COVID-19 Pandemic: Langkawi Vocational College Student Challenge in Using Google Classroom for Teaching and Learning (T&L),” *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3299–3307, 2020.