Comparative Analysis of Student Performance Across Different Cohorts in Higher Education

Emilija Kisić[1*], Miroslava Raspopović Milić[1], Jovana Jović[1], Nemanja Zdravković[1] and Faruk Selimović[1]

[1] *Faculty of Information Technology Belgrade Metropolitan University, Tadeuša Košćuška 63, 11000 Belgrade, Serbia*

Abstract

Tracking student progress throughout their coursework is a common topic in educational research. While valuable insights can be gained from analyzing learner data, such analysis can sometimes be misleading, particularly when it involves predicting students' final course achievements or drawing generalized conclusions from these predictions that do not account for individual student engagement. This study analyzes learner data from two different student cohorts that attended the same course in two different academic years. The focus on the paper is placed on identifying patterns in student engagement, similarities between two cohorts and exploring individual differences of learners. The interpretation of student activities sequences was implemented using sequence plotting and heatmaps, while Ward method was used for hierarchical clustering. The study aimed to understand the extent of similarities and differences in learning behavior across the cohorts, providing insights into how students interact with course material over a semester. Results show that there are many similarities between two cohorts, however, when expressing individual differences of each learner it was concluded that none of the students had the same sequence of engagement as the cluster's mean.

Keywords

Learning analytics, higher education, student performance analysis

1. Introduction

The tracking of student engagement throughout the semester and the prediction of final exam performance based on semester activities are common topics in scientific literature [1,2]. Analyzing student engagement helps instructors better predict academic achievements, adapt teaching methods, and intervene in time to improve student performance [3,4]. Analyses of engagement are often not precise enough to capture individual learning patterns, as engagement is a complex process that includes cognitive, emotional, and behavioral dimensions [1]. Although engagement metrics can provide valuable insights, their limitations may lead to less accurate assessments of student progress and success [5,6]. For the purposes of quantitative and visual methods in studying student engagement patterns, learning analytics has become an important tool in modern education, especially in the context of online learning [7]. Learning analytics enables educational institutions to identify specific student needs, provide personalized support, and improve educational practices based on behavior and engagement patterns [8–10]. Although learning analytics offers numerous advantages in this field, it also faces challenges regarding individualization. The analyses applied often cannot precisely capture the qualitative differences among students. For instance, authors of paper [11] that quantitative methods are useful for detecting educational strategies, but note that further research using standardized instruments for measuring motivation and goal orientation would be needed to gain deeper insights into the internal motivational factors influencing engagement. These motivational factors, and the impact of various factors on education in general,

Proceedings for the 15th International Conference on e-Learning 2024, September 26-27, 2024, Belgrade, Serbia
* Corresponding author.
emilija.kisic@metropolitan.ac.rs (E. Kisić); miroslava.raspopovic@metropolitan.ac.rs (M. Raspopović Milić); jovana.jovic@metropolitan.ac.rs (J. Jović); nemanja.zdravkovic@metropolitan.ac.rs (N. Zdravković); faruk.selimovic@metropolitan.ac.rs (F. Selimović);
0000-0003-3059-2353 (E. Kisić); 0000-0003-2158-8707 (M. Raspopović Milić); 0000-0002-4204-0233 (J. Jović); 0000-0002-2631-6308 (N. Zdravković); 0000-0002-0367-9122 (F. Selimović);

are particularly significant when studying engagement patterns at the cohort level, as individual factors can become lost in the broader picture. Considering that, there is a need to examine in more detail the behavioral and engagement patterns of individuals within the same groups, as well as between different student cohorts. Studying cohorts enables the identification of shared characteristics that can aid in predicting success on final exams, as well as in identifying and suggesting support strategies that can help students progress. Previous research indicates that cohorts exhibit specific engagement patterns during different phases of a course, making it crucial to monitor performance over a defined period [12]. Variability in engagement over time can be illustrated through various visualizations, such as sequential charts and heat maps, which enable instructors and educational institutions to better monitor student engagement throughout the semester [13].

Research on engagement and academic achievement also indicates that motivation and a sense of belonging have a significant impact on educational outcomes. For example, authors in [14] show that a sense of connection to the academic community can greatly contribute to students' motivation and academic success. Literature highlights the need of learning support structures in online environments, including customized assistance and mentoring, to address challenges such as isolation, technology difficulties, and limited peer interaction, which can significantly impact student engagement, retention, and academic success [15,16]. Research on student engagement in higher education focuses on identifying general patterns of behavior across students, with less attention paid to examining individual differences within these groups. The question arises: can statistical models of engagement provide sufficiently accurate data on individual behavior relative to their group? In paper [17] emphasize that counting clicks or platform is insufficient for a comprehensive analysis of engagement. They advocate for combining quantitative and qualitative methods to achieve a holistic understanding of engagement patterns. Therefore, it is important to investigate the patterns of both the group and the individual in relation to the group.

The aim of this paper is to address these questions through an analysis of student engagement from two consecutive cohorts who attended the same course during the academic years 2022/2023 and 2023/2024. Patterns of student engagement were examined in relation to course activities, including assignments, tests, and project tasks. Focusing on engagement data from the Learning Activity Management System (LAMS) and the institutional information system, this analysis seeks to identify similarities and differences between the cohorts, as well as within each group, using sequential charts, heat maps, and hierarchical clustering with Ward's method.

The goal of this work is to compare two student cohorts that have same pre-exam activities on the same course and to answer following research questions:

RQ1: What are the similarities between two cohorts?
RQ2: What are the differences between individual students and other students in the same cluster?

This paper is organized as follows. Section 2 describes the design of the course that was used for data collection of learner data. Section 3 presents the research methodology that was used to analyze student progress levels. Section 4 provides results and discussion on the results, while Section 5 concludes the paper.

2. Implemented course format

The course CS120 - Computer Organization consisted of 15 lessons, each during one week in the semester, and each covering a topic in computer architecture and organization. The course had various pre-exam assessments throughout the semester in the form of tests, homework assignments and individual projects.

Every three weeks, each student was given a unique homework assignment covering the lessons from the previous weeks, including the one in which the assignment was given. Furthermore, a total of 14 tests were given, starting from week 2 up to week 15. Each test covered the lesson in the previous week. Finally, each student was given a project in week 3, for which they had to write a 10–15 page paper, and a presentation. Students consulted the progress of their projects with professors and teaching assistants during the remainder of the semester, and defended their projects in week 15.

Course design used for this study was implemented using published lessons on Learning Activity Management System (LAMS). Chosen course Computer Organization is a course taken by the first-year undergraduate students, all from computing majors. First cohort of students taught in the academic year 2022/23 was attended by 83 students, and the second cohort taught in 2023/24 was attended by 102 students.

Teaching material was created for the semester that was taught for 15 weeks, and design was chosen based on the need to build student knowledge progressively to allow for effective tracking of student progress. Hence, each week was used to teach one lesson and each week also included one or more activities that were assigned to students for grading. As this study uses tracking of student activities during the semester and their progress, it should be noted that these activities include homework, tests and projects.

Homework assignments were used to reinforce teaching content and were due nearly each week. Homework was assigned weekly in all weeks except the last one. Course project was graded in the last week of the semester, providing students the opportunity to apply learned concepts comprehensively. Within the graded assignments, students were also graded through 5 tests. Tests were used to assess specific segments of the course allowing students to demonstrate gained knowledge over a smaller part of teaching material. Test 1 covered lessons 1-3, test 2 covered lessons 4-6, test 3 covered lessons 7-9, test 4 lessons 10-12, while test 5 covered teaching materials from lessons 13 and 14. Using such course design allowed to keep students engaged and to allow tracking their learning process more effectively.

3. Methodology

This paper analyzes data from students enrolled in the Computer Organization course that involve two different cohorts from two consecutive academic years, 2022/2023 and 2023/2024. The course content, assignments, and evaluation criteria, as described in the previous section, were consistent for both cohorts, ensuring a standardized structure for the data set. This uniformity allowed an accurate and fair comparison of learner behaviors and performance across these academic years.

All relevant data were collected from the Learning Activity Management System (LAMS) and the institutional information system. Collected data provided comprehensive records of student engagement with course materials, including all of their activities during the semester and final exam points. All data were gathered in two datasets, one for each student cohort. Both datasets have the same structure in alignment with the course format as described in the previous section: points from 5 tests, points from 14 homework, project points and final exam points.

To address the research questions, various quantitative and visual methods were employed. These techniques were chosen to identify patterns in student engagement, similarities between two cohorts, and to explore the differences between individual students and other students in the same cluster: To examine the overall data characteristics, distribution plots and basic descriptive statistics were used to assess the spread and central tendencies of student engagement metrics and performance indicators. These visualizations helped identify key trends and anomalies in the dataset, revealing insights into how different groups of students interacted with the course materials and performed in assessments. To facilitate interpretation of the complex data, visualization of students' activities was applied, combining sequence plot and heatmap. This combination of sequence plot and heatmap was used to visualize intensity in student activity across the semester in order to track student progress from the first week of semester until the last week and finally on the exam. Finally, clustering was used to categorize students into distinct groups based on their learning patterns and performance. Hierarchical clustering, specifically Ward's method, was chosen for its efficiency in minimizing variance within clusters. This method grouped students with similar engagement profiles, providing insights into common characteristics among high, middle, and low performers. It also helped identify whether specific behavioral patterns were associated with particular performance levels. After the clustering was performed, we calculated individual differences within the clusters using Euclidean distance to explore how many students were close to the assigned cluster's centroid and how many were significantly distant.

Integration of distribution analysis, advanced visualizations, and hierarchical clustering, provided a robust framework for interpreting learner data for the analyzed course.

4. Results and discussion

In this section, we present and discuss the findings from our analysis of two student cohorts, focusing on their engagement patterns, clustering, and individual variations. The study aimed to understand the extent of similarities and differences in learning behavior across the cohorts, providing insights into how students interact with course material over a semester. The results are structured to address two key research questions (RQ1 and RQ2), examining both cohort-level patterns and individual differences within clusters.

4.1. RQ1: What are the similarities between two cohorts?

Regarding RQ1, the analysis revealed several key similarities between the two cohorts. Figure 1 presents a visualization of students' semester activities from the first to the last week of the semester for the 2022/2023 cohort, while Figure 2 shows the same for the 2023/2024 cohort. We categorized students as “low progress” if they scored less than 50% on their assignments, “middle progress” if they scored between 50% and 70%, and “high progress” if they scored above 70%. “Low progress” students are labeled with red, “middle” with yellow, and “high progress” students with green color. The x-axis represents the number of a week in the semester, and the y-axis shows each student’s progress on a weekly basis. According to the course format described in Section 2, each week of the semester included at least one assignment, with some weeks containing two activities, as shown in the figures. Each achievement is represented by a rectangle, colored according to the categorization, as explained. For each week, progress is sorted so that “high progress” students appear at the top, “middle progress” in the middle, and “low progress” student sequences at the bottom of the figure.

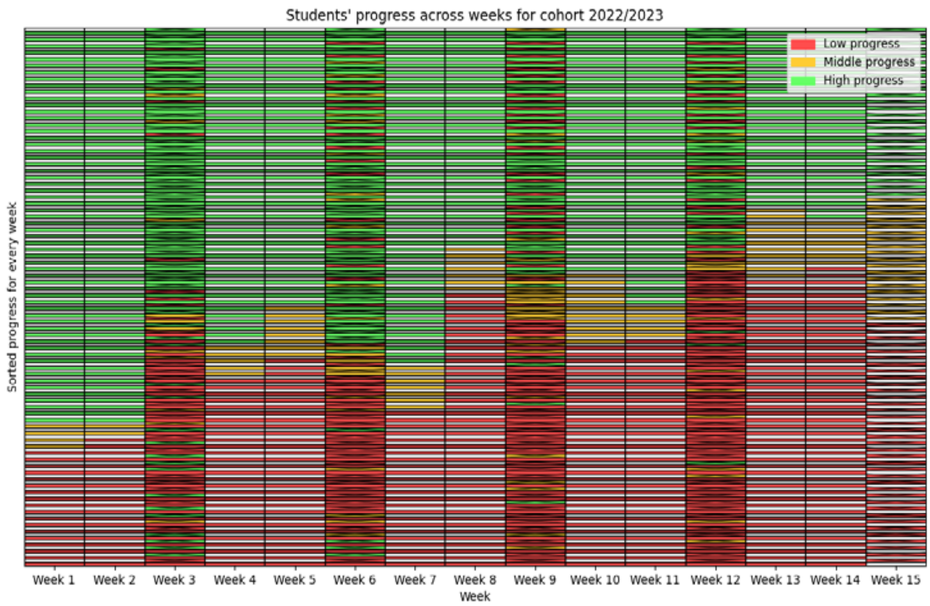

Figure 1: Visualization of students’ progress across weeks (Cohort 2022/23)

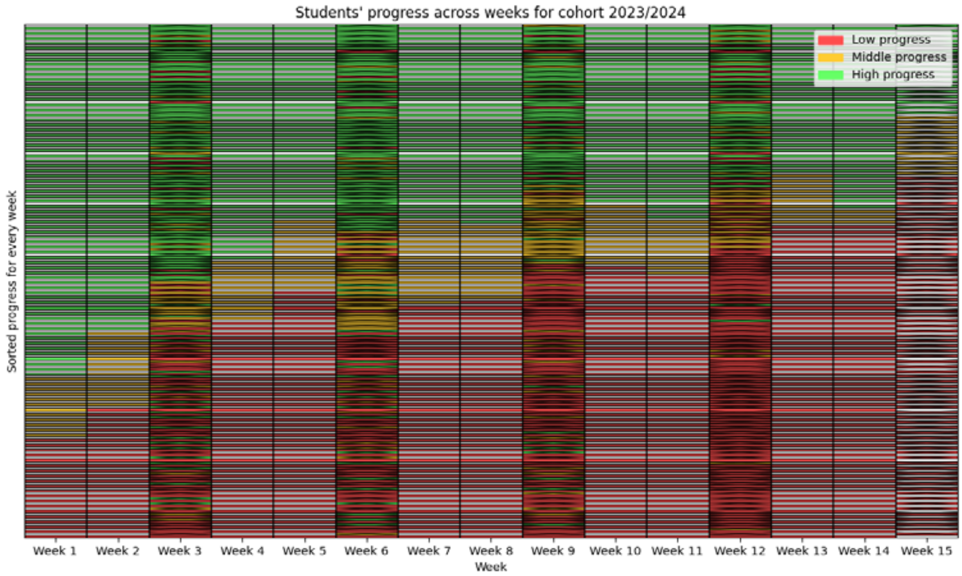

Figure 2: Visualization of students' progress across weeks (Cohort 2023/24)

Figures 1 and 2 demonstrate a consistent pattern of behavior across both cohorts' semester activities, suggesting that the two generations have a similar distribution of low, middle, and high progress throughout the semester. These similarities indicate that students, regardless of cohort, tend to engage with coursework and assessments at comparable levels and frequencies from the start to the end of the semester. Figures 3 and 4 further confirm, through statistical analysis, significant similarities in the distribution of semester activities between the cohorts, despite some differences in the final exam point distributions.

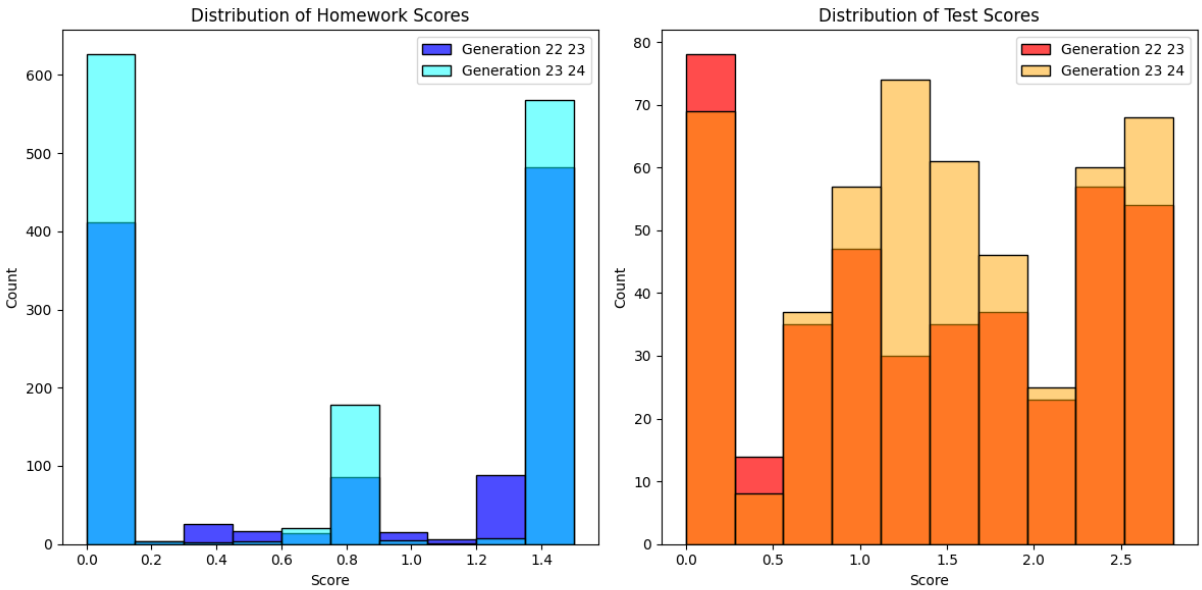

Figure 3: Distribution of homework and test scores

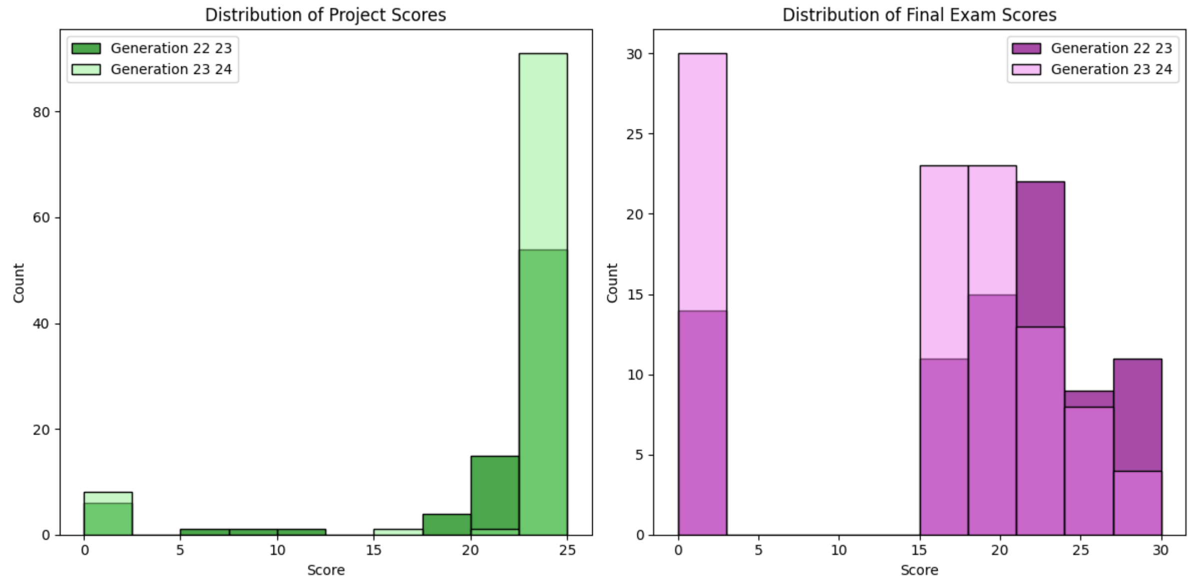

Figure 4: Distribution of project and final exam scores

4.2. RQ2: What are the differences between individual students and other students in the same cluster?

Regarding RQ2, hierarchical clustering using Ward's method was applied to categorize students into distinct groups based on their learning patterns and performance. Silhouette scores were calculated to determine the optimal number of clusters, which was found to be 3. These clusters correspond to three groups of students: low progress, middle progress, and high progress. From this analysis, we can conclude that learner data on semester activities along with scores on the final exams can be effectively grouped into clusters, with similar behavioral patterns evident within each cluster.

To further investigate individual differences, we calculated the Euclidean distance between each student and the centroid of their assigned cluster. The results are presented in Table 1.

Table 1
Percentage of students by distance category for cohort 2022/2023 and cohort 2023/2024

Categorization	Number of students [%] Cohort 2022/23	Number of students [%] Cohort 2023/24
Low	59.8%	59.4%
Moderate	19.5%	20.8%
High	20.7%	19.8%

Table 1 reveals that approximately 60% of students are close to their corresponding cluster centroids, with these distances categorized as low. About 20% of students are moderately distant from their cluster centroids, categorized as moderate, while the remaining 20% are far from their cluster centroids, categorized as high. The distribution of distances from cluster centroids is very similar across both cohorts, confirming again the observed similarities between the two groups for different student categories.

We can conclude that around 60% of students exhibit behavioral patterns similar to others in their cluster, while around 20% demonstrate behaviors that differ more significantly from their cluster peers. The analysis also revealed that no student had an identical engagement sequence to the cluster mean, highlighting unique variations even within grouped patterns.

5. Conclusions and future works

This research compared two different student cohorts that attended the same course in two different academic years. Learner sequence was visualized for each student, and the findings were that consistent patterns exist between both cohorts, despite some differences in final exam point distributions. Two cohorts displayed following similarities (RQ1): (i) similar pattern in behavior and (ii) significant similarities regarding distributions of semester activities. Based on these findings, further analysis was conducted to identify if learner data can be grouped in clusters using Ward's method. Hierarchical clustering analysis (RQ2) distinguished 3 groups "low", "middle" and "high" progress students. Even within the clusters similar patterns in behavior were identified. However, when analyzed the individual differences of learner sequence for each learner as compared to the characteristics of each clusters following findings were concluded: (i) around 60% of students have similar patterns in behavior as other students in the cluster, (ii) around 20% of students are different from other students from the same cluster, and (iii) none of the students had the same sequence of engagement as the cluster's mean.

After this analysis the following questions were posed. Can these findings be used for early intervention during the semester when needed? About 60% of students in both generations more or less follow the pattern of behavior as the students within the same cluster, but there are 20% of students which do not follow this pattern. When we want to make predictions and make conclusions considering the student population, a justified question arises: can we make conclusions applicable to all students? Future work should consider individualized approaches more and be careful with early intervention during semester because there are students which show "good achievement" during semester and "bad achievement" on the exam, so motivating this group of students is very important. In this regard, a hybrid model of individual approach and student modeling should be proposed to track student progress and evaluated on several groups of students.

Acknowledgements

This work is partially supported by the Erasmus + project "Improving the quality and sustainability of learning using early intervention methods based on learning analytics (ISILA)" (project. no. 2023-1-FI01-KA220-HED-000159757).

Declaration on Generative AI

The author(s) have not employed any Generative AI tools.

References

[1] K. Schnitzler, D. Holzberger, T. Seidel, "All better than being disengaged: Student engagement patterns and their relations to academic self-concept and achievement," Eur J Psychol Educ, vol. 36, pp. 627–652, 2021. doi:10.1007/s10212-020-00500-6.

[2] D. Oreški, E. Kisić, J. Jovic, M. Raspopović Milić, "Student modeling with clustering: comparative analysis of case studies in two higher educational institutions from different countries," in Proceedings of the Fourteenth International Conference on E-Learning, 2023.

[3] W. Tomaszewski, N. Xiang, Y. Huang, M. Western, B. McCourt, I. McCarthy, "The impact of effective teaching practices on academic achievement when mediated by student engagement: Evidence from Australian high schools," Educ Sci (Basel), vol. 12, p. 358, 2022. doi:10.3390/educsci12050358.

[4] J. Jović, M. Raspopović Milić, S. Cvetanović, "Multidimensional concept map representation of the learning objects ontology model for personalized learning," J Internet Technol, vol. 24, pp. 1043–1054, 2023. doi:10.53106/160792642023092405003.

[5] H. Lei, Y. Cui, W. Zhou, "Relationships between student engagement and academic achievement: A meta-analysis," Soc Behav Pers, vol. 46, pp. 517–528, 2018. doi:10.2224/sbp.7054.
[6] G. A. Frishkoff, K. Collins-Thompson, L. Hodges, S. Crossley, "Accuracy feedback improves word learning from context: evidence from a meaning-generation task," Read Writ, vol. 29, pp. 609–632, 2016. doi:10.1007/s11145-015-9615-7.
[7] M. Á. Conde, A. Georgiev, S. López-Pernas, J. Jovic, I. Crespo-Martínez, M. Raspopović Milić, et al., "Definition of a learning analytics ecosystem for the ILEDA project piloting," in Lecture Notes in Computer Science, Cham: Springer Nature Switzerland, 2023, pp. 444–453. doi:10.1007/978-3-031-34411-4_30.
[8] D. Gašević, J. Jovanović, A. Pardo, S. Dawson, "Detecting learning strategies with analytics: Links with self-reported measures and academic performance," J Learn Anal, vol. 4, pp. 113–128, 2017. doi:10.18608/jla.2017.42.10.
[9] C. R. Henrie, L. R. Halverson, C. R. Graham, "Measuring student engagement in technology-mediated learning: A review," Comput Educ, vol. 90, pp. 36–53, 2015. doi:10.1016/j.compedu.2015.09.005.
[10] E. Kisić, M. Raspopović Milić, N. Zdravković, M. Á. Conde, "The Effects of Implementing Project-Based Learning in the Programming Course," in Proceedings of the 14th International Conference on eLearning (eLearning-2023), Belgrade, Serbia, 2023.
[11] V. Kovanović, D. Gašević, S. Joksimović, M. Hatala, O. Adesope, "Analytics of communities of inquiry: Effects of learning technology use on cognitive presence in asynchronous online discussions," Internet High Educ, vol. 27, pp. 74–89, 2015. doi:10.1016/j.iheduc.2015.06.002.
[12] A. M. Pazzaglia, M. Clements, H. J. Lavigne, "An Analysis of student engagement patterns and online course outcomes in Wisconsin," REL Midwest, 2016. Available: http://files.eric.ed.gov/fulltext/ED566959.pdf.
[13] E. Gaudino-Goering, "Using a heat map to visualize academic outcomes," Learning Outcomes Assessment, 2021. Available: https://www.learningoutcomesassessment.org/wp-content/uploads/2021/02/AiP-Gaudino-Goering-2.pdf.
[14] D. R. Johnson, "College students' sense of belonging: A key to educational success for all students by Terrell L. Strayhorn," J Coll Stud Dev, vol. 54, pp. 662–663, 2013. doi:10.1353/csd.2013.0088.
[15] O. Rotar, "Online student support: a framework for embedding support interventions into the online learning cycle," Res Pract Technol Enhanc Learn, vol. 17, 2022. doi:10.1186/s41039-021-00178-4.
[16] C. R. Graham, J. Borup, S. Tuiloma, A. Martínez Arias, D. M. Parra Caicedo, R. Larsen, "Institutional support for academic engagement in online and blended learning environments," Online Learn, vol. 27, 2023. Available: https://olj.onlinelearningconsortium.org/index.php/olj/article/view/4001.
[17] E. Fincham, A. Whitelock-Wainwright, V. Kovanović, S. Joksimović, J.-P. van Staalduinen, D. Gašević, Counting clicks is not enough, in: Proceedings of the 9th International Conference on Learning Analytics & Knowledge, ACM, New York, NY, USA, 2019. URL: http://dx.doi.org/10.1145/3303772.3303775. doi:10.1145/3303772.3303775.P. S. Abril, R. Plant, The patent holder's dilemma: Buy, sell, or troll?, Communications of the ACM 50 (2007) 36–44. doi:10.1145/1188913.1188915.