

Clustering Student Performance Across Changing Teaching Modalities: A Three-year Longitudinal Learning Analytics Study

Miran Zlatović, Igor Balaban and Marko Matus

Faculty of Organization and Informatics, University of Zagreb, Croatia

miran.zlatovic@foi.unizg.hr (corresponding author)

igor.balaban@foi.unizg.hr

marko.matus@foi.unizg.hr

<https://doi.org/10.34190/ejel.24.5.4870>

An open access article under [CC Attribution 4.0](#)

Abstract: Learning Analytics research frequently examines learner behaviour and performance within a single cohort or teaching context, leaving limited evidence on whether comparable learner profiles recur across different teaching modalities. This study addresses this gap by analysing three consecutive cohorts (N=636) of the same university course delivered under fully online, blended, and on-site teaching conditions. Data included knowledge-domain mastery estimates and LMS activity logs used for clustering, while final course grades were used post hoc to characterise the resulting learner profiles. K-means clustering was used to identify and compare learner profiles, while non-parametric tests examined differences in mastery and final grades across cohorts. Three broadly comparable profiles emerged: higher-performing students (C1), failing students (C2), and struggling students comprising failing and low-passing students (C3). While their performance-based structure remained recognisable across cohorts, LMS activity patterns varied considerably. C2 consistently exhibited the lowest activity, whereas higher activity did not consistently distinguish higher-performing from struggling students. Significant differences in knowledge-domain mastery were observed across cohorts, with higher mastery levels in later cohorts in four of seven domains, while significant differences in final grades emerged primarily for the third cohort. The findings extend previous single-cohort research and demonstrate how learner profiles can inform differentiated educational support.

Keywords: Student activity, Overlay model, Clustering, Teaching modalities, LMS logs

1. Introduction

In recent years, largely due to COVID-19 pandemic, educational institutions worldwide have rapidly transitioned from in-person teaching to online and blended teaching, often without the benefit of deliberate instructional design (Bond et al., 2021; Lowenthal et al., 2020). These changes have brought into focus the questions about how the changes to content delivery modalities impact student engagement, learning, and performance. The instructional effectiveness of e-learning in general varies with student characteristics such as age, background, and learning context (Purwoningsih, Santoso and Hasibuan, 2020), which highlights the need for instruction that considers the diversity of learners.

Teachers have relied heavily on Learning Management Systems (LMS) during this transition, which has enabled both flexibility in course delivery and ability to maintain digital traces of student activity. Therefore, researchers have readily been using Learning Analytics (LA) in order to analyse usage patterns, predict performance and enable personalized support (Jha, Jha and O'Brien, 2019; Wong, 2017).

However, most existing studies examine learner behaviour and performance within a single cohort or teaching context. Our previous study (Balaban, Filipović and Zlatović, 2023) followed this approach by combining overlay modelling and LMS activity data to identify three distinct learner profiles within a single cohort during emergency remote teaching, with final course grades subsequently used to characterise the resulting profiles. While this demonstrated the potential of post hoc Learning Analytics (LA) for identifying meaningful learner profiles, the single-cohort design could not establish whether the observed profiles were specific to the emergency remote teaching context or represented recurring patterns that would also emerge in subsequent cohorts.

The present study addresses this limitation by extending the analysis to three consecutive cohorts of the same course delivered under fully online, blended, and on-site teaching modalities, while maintaining the same course content, teachers, and assessment approach. Such a longitudinal approach enables us to examine whether comparable learner profiles recur across cohorts and how their characteristics, learning activity, knowledge-domain mastery, and academic performance vary across changing teaching contexts. In this way, the study

moves beyond the identification of learner profiles within a single cohort towards examining their stability and variation across cohorts and teaching modalities.

2. Theoretical Background

With the advent of learning analytics, there is a growing focus on evaluating the quality of educational practices and predicting student outcomes. Numerous studies (Pande, 2023; Siemens, 2013; Yan, 2022; You, 2016) suggest that predicting student academic performance enables instructors to intervene in the learning process and aid students in enhancing their learning experiences. Performance prediction is therefore essential for the early detection of students who might need more academic support. With its help, teachers can:

- Personalize materials and instruction to each student's specific needs (Hussain et al., 2019; Ifenthaler and Yau, 2020), and
- Reduce students' dropout rates by identifying those students who are at risk of dropping out (Hoyos Osorio and Daza Santacoloma, 2023).

It becomes evident that incorporating student performance prediction and modelling is a crucial strategy to address and lower student dropout rates and ultimately improve educational outcomes.

2.1 Learning Analytics as a Conceptual Framework

As already mentioned, LA uses learner-generated data to better understand and support learning processes and outcomes. On the other hand, LMS's provide particularly valuable digital traces, as students' interactions with learning resources and assessments can be combined with performance data to represent different aspects of learner behaviour and achievement (Bergdahl et al., 2020; Ifenthaler and Yau, 2020).

To conceptually structure the transformation of such data into educationally meaningful information, this study adopts Clow's (2012) Learning Analytics Cycle, which has been used as a process-oriented framework for conceptualising and implementing LA (Baker, Gašević and Karumbaiah, 2021; van Leeuwen et al., 2017) and remains relevant in recent LA research (see for example Yildirim-Erbaşlı, Dibek and Taikh, 2026). The framework conceptualises LA as an iterative process comprising four interconnected elements: Learners, Data, Metrics, and Interventions. Learners generate data through their interactions with learning environments. Next, these data are transformed into metrics and analytical insights, which may subsequently inform educational interventions and thereby influence further learner activity.

Within this framework, the present study operationalises the first three elements of the cycle. Learners interact with an LMS-supported learning environment and generate digital traces. Data from these interactions and assessments are transformed into learner representations, and Metrics are subsequently produced through overlay modelling and clustering, followed by descriptive and statistical analyses used to characterise the resulting profiles and compare cohorts. Clustering is appropriate in this context because it enables learners exhibiting similar behavioural and performance patterns to be identified as distinct profiles, an established approach in Learning Analytics research (Han et al., 2022). The fourth element, Interventions, remains outside the scope of the present study as the identified profiles instead provide an analytical basis for future intervention-oriented research.

2.2 Using Student Modelling to Identify and Describe Student Groups

Within the Metrics element of the Learning Analytics Cycle, student modelling provides a means of transforming learner data into interpretable representations of learners and their learning processes. Student modelling can be defined as the process of representing learner characteristics such as knowledge level, performance, misconceptions, and preferences (El Aissaoui, Oughdir and El Alloui, 2022).

When designing a student model, Chrysafiadi and Virvou (2013) highlight approaches such as: overlay models, stereotypes, perturbation models, machine learning techniques, constraint-based models, fuzzy models, Bayesian networks and ontology-based models. Those approaches can assist in creating an adaptive student model based on historical sequence data (Mao et al., 2020) which can predict student's performance thus presenting an important factor that helps academic institutions to take preventive strategies to decrease the rate of failure and dropping out (El Aissaoui, Oughdir and El Alloui, 2022). Among these, the overlay model is very relevant to the present study, as it enables the learner's knowledge state to be represented across predefined knowledge domains. Such learner representations can then be combined with behavioural data and analysed using data-driven techniques to identify similarities and differences among learners.

Clustering provides a complementary data-driven approach for identifying groups of learners with similar characteristics. In LA, clustering techniques have been applied to LMS-generated behavioural and performance data to identify learner profiles and reveal patterns that may not be evident from individual indicators alone. For example, Gaftandzhieva et al. (2022) have used ML (machine learning) classifiers to detect student groups and predict their final grades with moderate accuracy of 78% and have found correlations between final grades and online activities/lecture attendance. Analyses of LMS data about students' activity using clustering techniques often identified between two and four clusters:

- Quesada, Valverde and Rojas (2022) used K-Means algorithm with three clusters labelled "Active", "Absent" and "Drop Out Tasks", based on students' task-related behaviour and academic performance.
- Rajagukguk and Fudholi (2022) utilized the k-Prototype clustering to classify students into two groups: "privileged group" and "less-privileged group" based on socio-economic background and academic performance.
- Šarić et al. (2019) identified 4 clusters using K-Means algorithm, based on students' online learning behaviour: "Engaged Pre-knowers", "Pre-knowers Non-finishers", "Hard-workers", and "Non-engagers".

More recent research further demonstrates the relevance of clustering for learner profiling and early identification of students requiring support. Pecuchová and Drlík (2024), for example, used clustering of students' digital traces in a blended-learning environment to enhance early dropout prediction. Such findings illustrate how clustering can transform heterogeneous learner data into interpretable profiles that support the identification of meaningful differences in learning behaviour and performance. In the present study, this approach complements the overlay representation of students' knowledge by grouping learners according to similarities in their knowledge-domain mastery and LMS activity.

2.3 Analysis of Student Performance

There are many attempts aimed at analysing the interaction between student activity in Learning Management Systems (LMS) and final success in blended courses. Increased participation in online discussions and quizzes has often been linked to better learning outcomes (Broadbent et al, 2021; Cavanaugh, Hargis and Mayberry, 2016; Mitić and Djenic, 2023). Moreover, final course grades have been connected to the engagement in reading LMS materials (You, 2016), as well as to the quantity of logins, duration of sessions, and involvement in online activities (Nespereira et al., 2014). According to the latter source, the students with intermediate levels of engagement have shown the best performance.

The analysis of online activity in LMS has also made it possible to identify students who might be having difficulties early on, which in turn enables early interventions aimed at improving their academic performance and learning experience (Argyriou, Benamar and Nikolajeva, 2022). The same source reports that higher completion rates of weekly online quizzes predict final exam performance.

Buckley et al. (2021) used group-based trajectory modelling (GBTM) to accurately classify students based on page view counts. They found out that activity in LMS, such as page views and attendance at synchronous sessions, can predict final success in blended courses and identify at-risk students. However, raw activity data (such as page view counts) alone was not sufficient to predict student achievement – an additional stratification by attendance at virtual synchronous learning sessions improved predictive accuracy.

Recent empirical studies provide further evidence that LMS activity should be interpreted together with other learner indicators. Abdulkareem Shafiq et al. (2025), analysing Moodle digital traces alongside demographic, academic-history, family-background and attendance data, demonstrated the value of combining behavioural traces with broader student information when examining academic performance. Similarly, Xu et al. (2025) identified distinct learner engagement profiles and found that higher LMS engagement did not uniformly correspond to higher academic performance. This counter-intuitive relationship further illustrates that observable activity frequency cannot be treated as a direct proxy for learning success.

Previous findings therefore reflect the potential of LMS activity for identifying students at risk and examining academic outcomes, while also showing the limitations of interpreting behavioural traces in isolation. Similar levels of observable LMS activity may be associated with different academic outcomes, supporting the integration of behavioural indicators with performance-related learner representations and other relevant student characteristics.

On the other hand, predictors of academic success, such as lecture attendance, time management, and motivation, may also vary across different generations of students, indicating that educational strategies and student engagement methods may need to evolve over time (De Philippis and Rossi, 2016). Differences in student performance across generations further suggest that cohorts should not be assumed to be equivalent in terms of performance or learning behaviour (Roberson, 2011).

In the end, the evidence indicates that the relationship between LMS activity and academic performance is complex and may vary across learners and cohorts (see for example Cavanaugh, Hargis and Mayberry, 2016; Charbonneau-Gowdy, 2018; Van Goidsenhoven et al., 2020). This reinforces the value of identifying distinct learner profiles rather than interpreting LMS activity in isolation. Such profiling can provide a basis for more targeted educational support by distinguishing students who show very low engagement from those who remain active but nevertheless experience learning difficulties. Recent Learning Analytics research has increasingly emphasised this transition from student-performance prediction and at-risk identification toward actionable information and targeted pedagogical interventions (Alalawi, Athauda and Chiong, 2025). In terms of Clow's Learning Analytics Cycle, learner profiling can therefore provide an analytical basis for the subsequent intervention stage, although the design and evaluation of such interventions remain outside the scope of the present study.

3. Research Objectives

Multiple studies have shown that it is possible to detect student groups and analyse their performance, by using post hoc analysis of LMS data in conjunction with grouping techniques, i.e., through analysis of their self-assessment scores (Watson et al., 2017) or through analysis of their activities such as number of logins and content reads (Tseng et al., 2016). However, the studies identified in the literature review (see previous section) were focused only on a single generation of students. In order to further explore the ways in which the stored LMS data can be used to detect and describe student groups, as well as to study and explain possible differences between them, this research goes one step forward and uses the data from the same course, but throughout the period of 3 consecutive academic years.

While the course itself hasn't changed during the observed three-year period, it is important to notice that one external factor has been different in each year:

- During the baseline year 2020/21, the entire teaching process had to be conducted online, due to enforced Covid-19 pandemic full lockdown measures in Croatia.
- During the year 2021/22, partial and more relaxed lockdown measures were in place, with a mix of synchronous on-site teaching and (a)synchronous online teaching.
- In academic year 2022/23 there were no lockdown measures and the teaching process returned to its traditional synchronous on-site roots.

The aim is to examine whether comparable learner profiles can be identified in the same course across three consecutive academic years and how students' mastery and final grades vary over this three-year period, with the following research objectives:

RO1. To classify students into groups based on knowledge-domain mastery estimates and LMS activity using clustering across three consecutive cohorts.

RO2. To describe and compare the resulting student groups across three consecutive cohorts experiencing different teaching modalities, with implications for improving the course structure and activities.

RO3. To identify differences in student performance across three consecutive cohorts in terms of final grades and mastery of knowledge domains.

As noted in the Introduction, this study extends our earlier analysis of the 2020/21 cohort (Balaban, Filipović and Zlatović, 2023) by incorporating data from 2021/22 and 2022/23. The first two research objectives (RO1 and RO2) are therefore aligned with the previous work to preserve consistency of the analysis process from the original single-cohort study. This enables a longitudinal analysis of whether comparable learner profiles can be observed in this course across three consecutive academic years. The new research objective RO3 extends the longitudinal analysis by examining differences in students' performance across cohorts experiencing different course delivery modes associated with changing lockdown conditions.

Based on the research objectives, two research questions have been created. The first research question RQ1 supports the research objectives RO1 and RO2, while the second question RQ2 supports the objective RO3:

RQ1: What types of student groups can be detected based on students' knowledge-domain mastery and LMS activity across three consecutive academic years, and to what extent do comparable profiles recur across cohorts?

RQ2: What differences in student performance can be observed across three consecutive cohorts in terms of knowledge-domain mastery and final grades?

Consistent with the Learning Analytics Cycle, the present study focuses on the analytical stage of transforming LMS-generated digital traces into meaningful learner profiles. Rather than evaluating interventions, this study establishes learner profiles that can support intervention-oriented research in future work. The main purpose is to examine whether comparable learner profiles recur across cohorts experiencing changing teaching modalities.

Although the previous study of the 2020/21 cohort demonstrated the feasibility of identifying learner profiles within a single academic cohort, it remained unknown whether comparable profiles would recur across different cohorts experiencing substantially different teaching modalities. Addressing this gap is important because Learning Analytics aims to identify robust behavioural patterns that remain meaningful across changing educational contexts. Therefore, the present study extends the original single-cohort analysis into a longitudinal investigation involving three consecutive cohorts exposed to online, blended and on-site instruction.

4. Research Methodology

4.1 Research Workflow

The methodological workflow of this study is conceptually grounded in the Learning Analytics Cycle proposed by Clow (2012). As outlined in Section 2.1, the cycle comprises four interconnected elements: Learners, Data, Metrics, and Interventions. Since the objective of the present study is to identify and analyse learner profiles rather than to implement or evaluate educational interventions, the methodological workflow operationalises the first three elements of the cycle through three corresponding stages (see Figure 1).

Stage 1: Learning process and learner interactions (Learners) - in the first stage, corresponding to the Learners element of the Learning Analytics Cycle, the course was designed according to the principles of programmed learning, with its online component implemented in Moodle LMS. To provide a consistent learning path, the course combined learning materials, self-assessment quizzes, and conditional activities that guided students through the instructional content. Across three consecutive academic years, students interacted with the LMS-supported learning environment and generated digital traces reflecting their learning behaviour. A detailed description of the course design and teaching modalities is provided in Section 5.

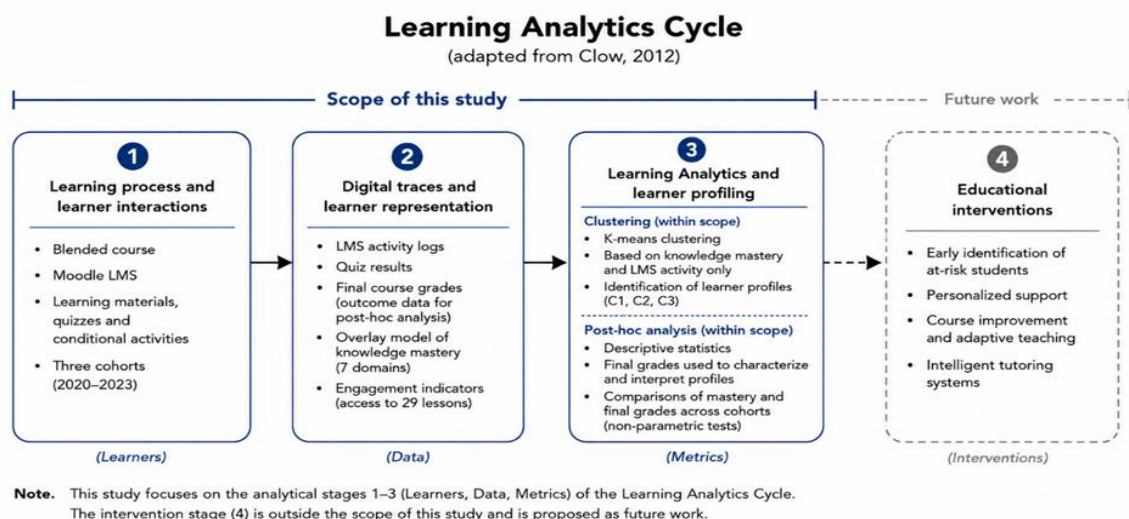


Figure 1: Mapping of the research stages onto Learning Analytics Cycle elements

Stage 2: Digital traces and learner representation (Data) - at the second stage, learner-generated data were collected from Moodle LMS, while final course grades were obtained from the National University Information System (NUIS). The analytical dataset consisted of three components: (1) overlay-model estimates representing students' mastery of seven knowledge domains, (2) frequencies of access to learning materials recorded in

Moodle, and (3) final course grades. During dataset preparation, real user identifiers were used to match data from two separate sources (LMS logs data and grades from NUIS). Dataset was then anonymised, in order to protect the identity of participants. To support data integration and preparation, a dedicated SQLite database was developed and processed using Python and SQLAlchemy.

Each student was represented by mastery estimates for seven (7) knowledge domains. Engagement indicators described students' access frequencies for 29 Moodle lessons (L1.1–L1.5, L2.1–L2.18, L3.1–L3.4, and L4.1–L4.2). Final course grades were included as an additional outcome indicator for subsequent cluster characterisation and cross-cohort analysis. More details about knowledge domains and LMS activity indicators are provided in Section 5.

The clustering analysis was performed using knowledge-domain mastery estimates derived from the overlay model and LMS activity measures. Final course grades were not included as clustering features; instead, they were used post hoc to describe and interpret the resulting clusters, particularly in terms of the distribution of failing and passing students.

Stage 3: Learning Analytics and learner profiling (Metrics) - in the third stage, corresponding to the Metrics element of the cycle, the prepared learner representations were analysed to identify and characterise learner profiles and to examine differences across the three student cohorts. Specifically, k-means clustering was employed to identify groups of students with similar characteristics, descriptive statistics were used to characterise the resulting profiles, and non-parametric statistical tests were conducted to examine differences in knowledge-domain mastery and final course grades across cohorts.

The fourth element of Clow's cycle, Interventions, is not operationalised in the present study. The identified learner profiles may, however, provide an analytical basis for the design and evaluation of educational interventions in future research.

4.2 Clustering Procedure

With one of the objectives being to partition students into mutually exclusive groups according to similarities in their knowledge-domain mastery and LMS activity patterns, a partitioning algorithm is an appropriate choice. K-means clustering was selected because it identifies compact clusters represented by centroids, is computationally efficient, and has been widely applied for exploratory grouping of numerical educational data (Dutt, Ismail and Herawan, 2017; Jain, 2010; Steinley, 2006). Applying the same clustering procedure to each academic year enables longitudinal comparison of the resulting cluster structures over time.

4.2.1 Feature selection

As already noted in Section 4.1, the analysed dataset contained three major components: students' mastery of seven knowledge domains, lesson-access frequencies, and final course grades. The clustering procedure was conducted using only the first two components: knowledge-domain mastery and lesson-access frequencies.

The clustering features were selected to represent complementary dimensions of student learning, according to the findings presented in Section 2.3, which emphasise the connections between formative assessment, observable LMS activity and final grades. Knowledge-domain mastery was represented by seven overlay-model mastery estimates derived from students' self-assessment quiz performance, with each estimate corresponding to a distinct course knowledge domain (see Section 5 for more details on course structure). LMS activity was represented by the frequency of accessing each of the 29 instructional lessons.

The final course grade was used for cluster characterisation, to help identify how the existing clusters built on domain-specific knowledge (formative self-assessment quizzes) and learning behaviour (assessment-independent reading activity) differ in terms of academic achievement (final grades).

The clustering analysis retained the original feature granularity rather than aggregating measures across course topics. Seven self-assessment quiz scores were included individually because each quiz assessed mastery of a distinct course domain. Likewise, reading activity was represented by the frequency of accessing each of the 29 instructional lessons, preserving differences in students' engagement with specific learning materials. Aggregating these measures into overall mastery or activity scores would have obscured topic-specific learning patterns (Elejalde and Wildemann, 2026; Trittenbach, Bach and Böhm, 2019) that were central to the study's objective of identifying detailed student learning profiles.

4.2.2 Calculation of overlay mastery scores

Overlay-model mastery estimates were derived from students' performance on the seven domain-specific self-assessment quizzes recorded in the LMS. For each knowledge domain, a student's mastery estimate was calculated, based on LMS logs data, as the proportion of the maximum score obtained on the corresponding self-assessment quiz. For example, a value of 0.8571 represents an overlay-model mastery estimate of 85.71% for the corresponding knowledge domain.

Since such mastery scores are already normalised to the [0,1] interval, additional normalisation during data pre-processing stage is not necessary. These scores can be used directly by k-means clustering algorithm, avoiding typical pitfalls of using raw or non-normalised data in clustering (implicit weighting and extreme maximum values when calculating Euclidean distance, scale bias, etc.).

4.2.3 Data pre-processing

Prior to clustering, all clustering features were represented on the [0,1] scale to ensure comparable contribution to the Euclidean distance metric used by the k-means algorithm. Knowledge-domain mastery estimates were already expressed on this scale through the overlay-model calculation and therefore required no additional normalization. In contrast, lesson-access frequencies were normalized using min-max scaling separately for each of the 29 lessons and separately within each academic-year cohort, according to $x' = (x - x_{min}) / (x_{max} - x_{min})$, where x_{min} and x_{max} represent the minimum and maximum observed access frequencies for the respective lesson within that cohort. This cohort-specific normalisation was consistent with the clustering procedure, which was performed separately for each academic-year cohort. Consequently, the normalised activity values represent relative activity patterns within each cohort and should not be interpreted as directly comparable measures of absolute LMS activity levels across cohorts.

4.2.4 Validation of clusters

The optimal number of clusters was determined using the Silhouette method (Chiang and Mirkin, 2010; De Amorim and Hennig, 2015; Rousseeuw, 1987). For each academic year, the dataset was clustered nine times, with the number of clusters ranging from 2 to 10. The Average Silhouette Width (ASW) scores were calculated for each clustering attempt using Python's scikit-learn library and plotted to visually assess the optimal number of clusters. The highest ASW score was consistently observed with three clusters (see Figure 2), leading to the selection of this scenario for the clustering analyses. These clusters were labelled as C1, C2, and C3.

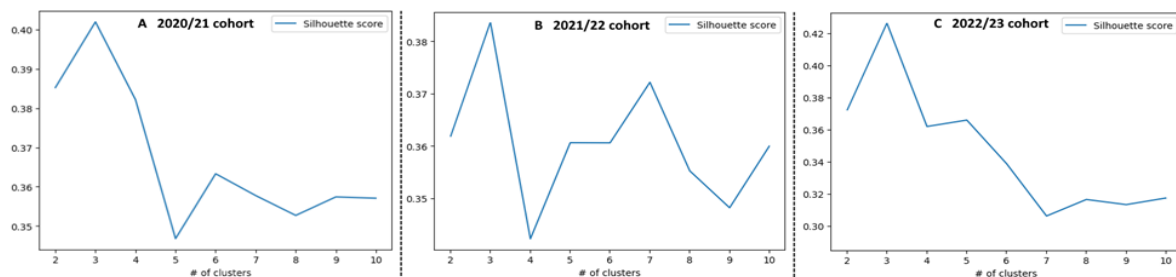


Figure 2: Silhouette method plots for three consecutive cohorts

5. Course Description and Design

The course "Business Informatics" is a first-year, opening semester course within a three-year undergraduate transdisciplinary (fields of informatics and economics) professional study program held at the University of Zagreb, Faculty of Organisation and Informatics. Key content areas covered in the "Business Informatics" curriculum include fundamentals of information systems and their business applications, understanding of computer hardware and software and the introduction to security of information systems. The breakdown of student enrolment per academic year is presented in Table 1.

Table 1: Course enrolment

Academic year	Number of enrolled students
2020/21	222
2021/22	196
2022/23	218

Onsite classes have been supplemented with asynchronous online activities within LMS (reading lectures, and quizzes, both formal and self-assessments). Without strict deadlines or time limits, students could partake in these activities at their own pace. The learning design of the course was driven by basic principles of programmed learning paths, as summarized by Seel (2012) where learning content is broken down into smaller parts, which are complemented by one or more comprehension questions. If students give correct answers, they are able to move on to the next part. Otherwise, they are required to redo the content until they give correct answers. Such learning design can enable the teachers to create better structured teaching activities, which in turn leads to more effective learning (see for example Britain, 2004).

From the technical perspective, the programmed learning paths were facilitated by Moodle's Lesson activities that allow teachers to blend the usual contents (pages of text combined with multimedia) with feedback-centred questions in a unit of content. The topics within each knowledge domain have been broken down into a series of lessons and have been prepared for the online portion of classes as asynchronous materials. The asynchronous content should be used by students at their own pace - either as main learning source about the topics that have been scheduled for the asynchronous self-study, or to relearn the topics already taught in the course during synchronous online/onsite lectures.

All knowledge domains were organized as shown in Figure 3 and made available in LMS to all students enrolled in the course. The beforementioned figure visualizes the sequential flow of activities within a knowledge domain. In order to supervise student's navigation through the elements of a knowledge domain, conditional activities (a feature of Moodle LMS) have been utilized. In general, Moodle allows for restriction of access to most of the learning resources within the course (e.g., lessons, quizzes, etc.) based on the results achieved on other resources (e.g., another lesson, quiz, assignment, etc.). Therefore, conditional activities were implemented with intention of preventing students from accessing more advanced contents of a knowledge domain, i.e., not before they have acquired the necessary knowledge of previous contents. Achieving the prescribed score of correct answers to control questions, which were always included at the end of a lesson, was a condition for advancing to the next lesson. Finally, when the required score in all the lessons within a domain was achieved, the final self-assessment quiz for that domain is unlocked, which could have been attempted an unlimited number of times.

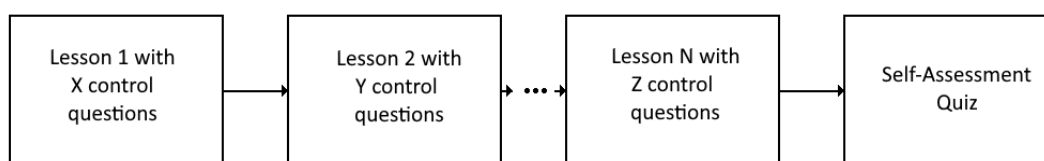


Figure 3: Flow of the lessons within a single knowledge domain

The course contents span over 7 domains of knowledge, which were already mentioned in the previous section. These domains are further arranged into two separate chains of learning. Having these two chains mutually independent, students were able to start their asynchronous studying from either of them. As shown in Figure 4, subsequent domains within each chain are dependent upon previous domains (full arrows). To start studying the dependent domain student must first show required levels of masteries in all the preceding domains. For example, to access the "Input/Output Unit", students must start by studying the "Basic Computer Principles" first and then proceed to the "Central Unit of a Computer".

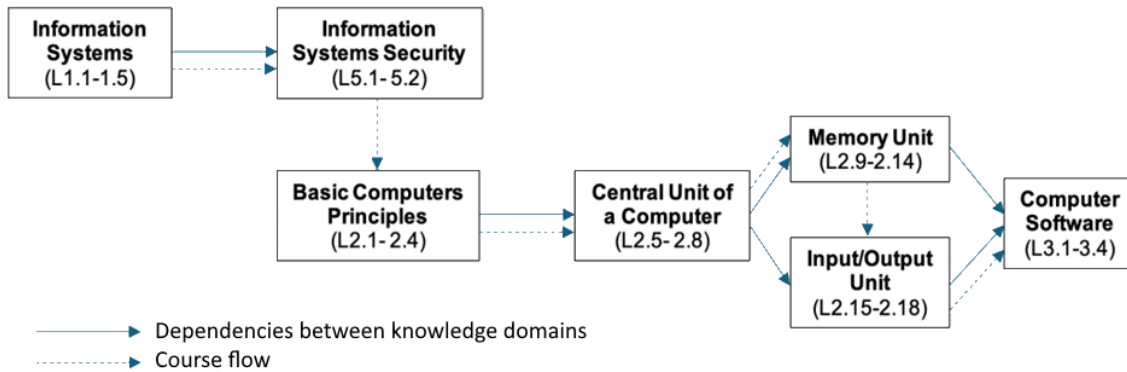


Figure 4: Course structure and flow

In order to assess students' knowledge and evaluate their effort in the course, besides the already mentioned final per-domain self-assessment quizzes, students also had to attend formal online midterms and the final exam. Both the midterms (during the semester) and the final exam (when the semester had ended, only for the students that failed to pass the course during the semester) were conducted within a controlled environment as formal online quizzes. New Moodle quizzes were prepared for each midterm and each exam. Students' grades from midterms and exams were used to determine their final course grades. In the present study, these final grades were used only for post hoc cluster characterisation and cross-cohort performance analysis and were not included as clustering features.

6. Results

6.1 Cluster Analysis of Learner Representations

For each cohort, clustering was performed using the knowledge-domain mastery estimates and LMS activity measures described in Section 4. Final grades were excluded from the clustering procedure.

6.1.1 Final grades per cluster

After the three clusters had been identified independently of final course grades, grade distributions were examined post hoc to characterise and interpret the resulting learner profiles. The frequencies of each possible grade within each cluster are presented in Tables 2, 3 and 4.

Data from Table 2 reveal that, during the baseline year 2020/21 (full lockdown), cluster C1 is primarily populated by students with positive grades, with one exception. Cluster C2 exclusively includes students who failed the course, while cluster C3 comprises the remaining students who either passed or failed the course.

Table 2: Final grades per cluster in academic year 2020/21

Grade	Clusters			Total
	C1	C2	C3	
1	1	125	22	148
2	37	0	7	44
3	23	0	4	27
4	2	0	1	3
5	0	0	0	0
No. of students	63	125	34	222

The grade counts for the three clusters in the academic year 2021/22 (partial lockdown) in Table 3 show similar patterns. Cluster C1 primarily includes students with the highest and passing grades, with three exceptions of failed students. Cluster C2 remains predominantly composed of students who failed the course, except for two students with the lowest passing grades. Cluster C3 consists of the remaining students, encompassing both those who passed and those who failed.

Table 3: Final grades per cluster in academic year 2021/22

Grade	Clusters			Total
	C1	C2	C3	
1	3	74	35	112
2	12	2	43	57
3	17	0	4	21
4	5	0	0	5
5	1	0	0	1
No. of students	38	76	82	196

The grade counts for the final year 2022/23 (no lockdowns, traditional on-site face-to-face classes) in Table 4 follow the same pattern. Cluster C1 is populated entirely with students having the highest positive grades, with no failing students. Cluster C2 is almost exclusively composed of failing students, except for two students with the lowest passing grades. Cluster C3 contains a mix of the remaining passing and failing students.

Table 4: Final grades for each cluster in academic year 2022/23

Grade	Clusters			Total
	C1	C2	C3	
1	0	85	16	101
2	0	2	62	64
3	46	0	0	46
4	6	0	0	6
5	1	0	0	1
No. of students	53	87	78	218

To facilitate interpretation throughout the remainder of the paper, the three clusters are therefore referred to as higher-performing students (C1), failing students (C2), and struggling students (C3). These labels are descriptive rather than absolute: C1 predominantly comprises students with passing grades, C2 predominantly comprises students who failed the course, while C3 represents a mixed profile of failing and low-passing students.

6.1.2 Activity per cluster

In addition to final grades, two aspects of student activity were analysed: the frequency of self-assessment quiz submissions and the frequency of reading provided lessons. This analysis was conducted for each academic year and further broken down by cluster.

The average number of self-assessment quiz submissions per student is presented in Figure 5. Cluster C2 consistently demonstrated the lowest quiz submission rates across all three cohorts. On the other hand, the relative activity of C1 and C3 varied substantially: in 2020/21, C3 was considerably more active than C1 (17.5 vs. 6.7 attempts per student), this pattern reversed in 2021/22 (11.5 vs. 26.2), while their activity was nearly identical in 2022/23 (20.7 vs. 21.5).

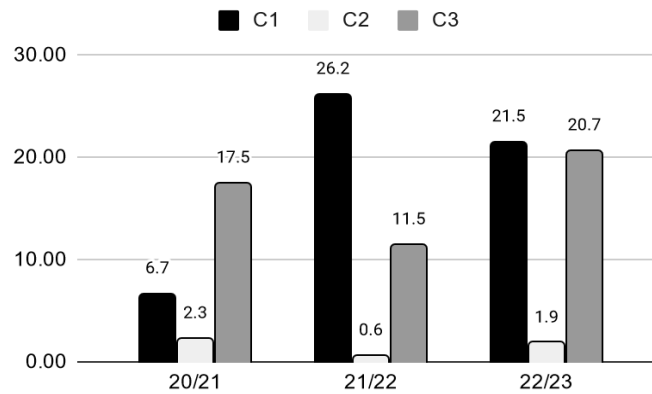


Figure 5: Comparison of the average self-assessment quizzes submission attempts per student

On the other hand, average reading attempts for all lessons within the course are shown in Figure 6. As described in Section 4.2.3, lesson-access frequencies were min–max normalized separately within each academic-year cohort. The figure therefore supports comparison of relative activity patterns among clusters within each cohort rather than direct comparison of absolute activity levels across cohorts.

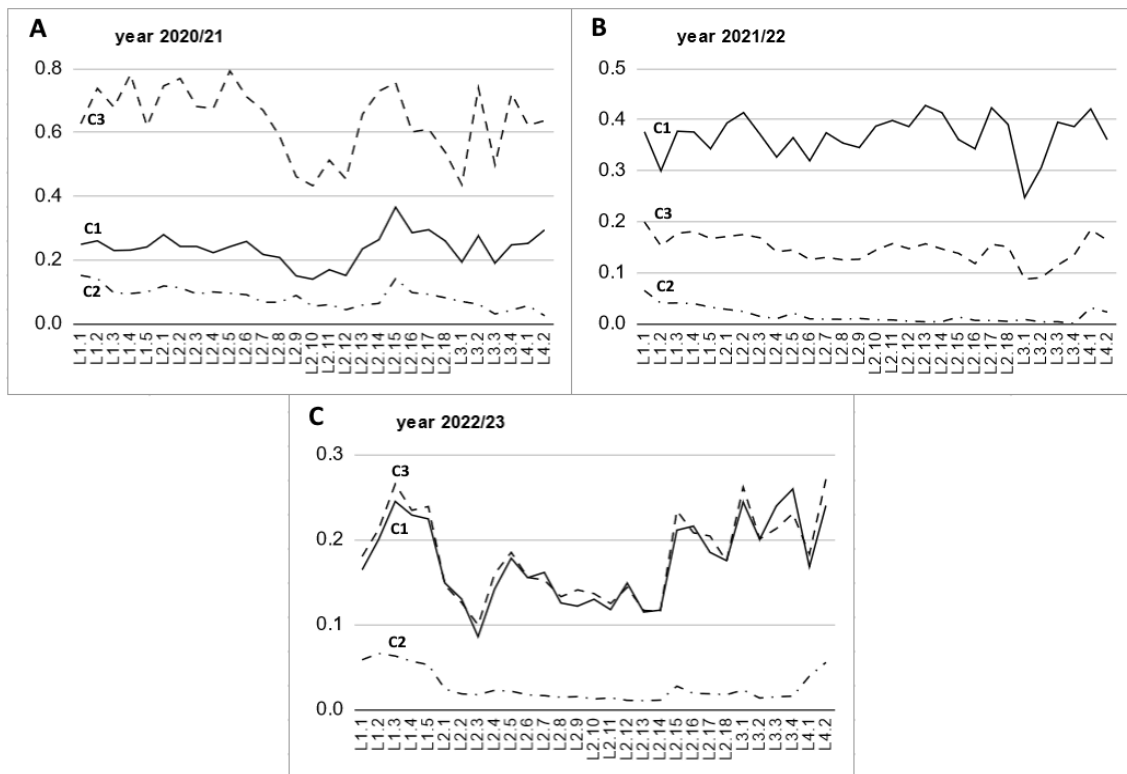


Figure 6: Comparison of relative lesson-access activity by cluster within each academic-year cohort (cohort-specific min–max scaling to [0,1])

Within each cohort, Cluster C2 showed the lowest relative lesson-reading activity. In 2020/21, C3 showed higher relative activity than C1; this within-cohort pattern reversed in 2021/22, while C1 and C3 showed similar relative activity in 2022/23. Because lesson-access frequencies were normalized separately within each cohort, these patterns describe the relative positioning of clusters within each academic year and do not indicate absolute differences in LMS activity between academic years.

6.2 Analysis of Knowledge-domain Mastery Estimates

To further explore students' academic performance, their overlay-model mastery estimates, derived from self-assessment quiz performance, were compared across all seven knowledge domains over the three-year period.

This analysis was conducted on the entire yearly populations rather than by clusters, to provide a broader perspective. Tests for normality, using both the Kolmogorov-Smirnov and Shapiro-Wilk tests, indicated that the mastery data was not normally distributed. For all 21 data subsets (7 knowledge domains $\times$ 3 years), the final significance statistics from both tests were $p = 0.000$ (at $\alpha = 0.05$), with degrees of freedom (df) of 222, 196, and 218 for the years 2020/21, 2021/22, and 2022/23, respectively. Given these results, a non-parametric Kruskal-Wallis H test was applied to assess year-to-year differences in mastery levels.

Visual inspection of the mastery distributions (Figure 7) revealed two general patterns across cohorts. Four domains – Information Systems (IS), Memory Unit (MU), Information Systems Security (ISS) and Input/Output Unit (IOU) – showed progressively higher median mastery levels from 2020/21 to 2022/23. The remaining three domains – Basic Computer Principles (BCP), Computer Software (CS) and Central Unit of a Computer (CUC) – showed higher mastery in 2021/22 compared with 2020/21, followed by relative stability or a slight decline in 2022/23.

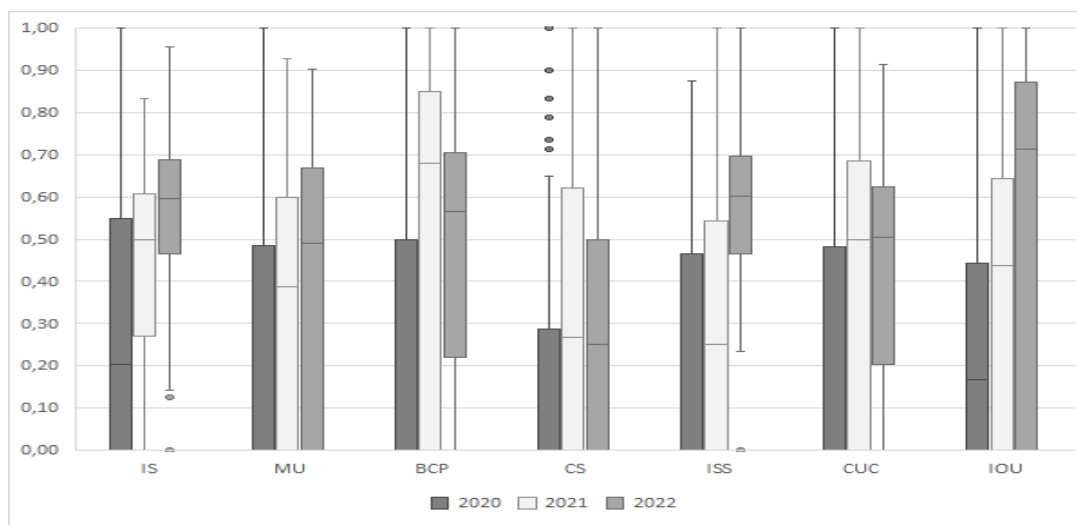


Figure 7: Box plots of students' overlay-model mastery estimates for each knowledge domain and academic year

To determine whether the visually observed differences are statistically significant, Kruskal-Wallis H tests were conducted, resulting in the following χ^2 values: IS = 76.504, MU = 34.326, BCP = 74.797, CS = 46.207, ISS = 117.458, CUC = 45.456, and IOU = 105.997 ($\alpha = 0.05$, $df=2$, all p -values < 0.001), indicating statistically significant differences in students' mastery across academic years for each knowledge domain.

Since more than two academic years were compared for each domain, additional pairwise comparisons were performed for all possible pairings of academic years, using Dunn-Bonferroni tests. For the IS, IOU and ISS domains, significant differences in students' mastery were found between all three academic year pairings ($p < 0.05$, adjusted using the Bonferroni correction), as shown in Table 5.

Table 5: Pairwise comparisons of mastery estimates in domains IS, IOU and ISS across academic years

Year1-Year2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
Knowledge Domain IS					
2020/21-2021/22	-68.877	17.790	-3.872	0.000	0.000*
2020/21-2022/23	-151.224	17.306	-8.738	0.000	0.000*
2021/22-2022/23	-82.347	17.866	-4.609	0.000	0.000*
Knowledge Domain IOU					
2020/21-2021/22	-76.645	17.594	-4.356	0.000	0.000*
2020/21-2022/23	-175.832	17.116	-10.273	0.000	0.000*
2021/22-2022/23	-99.188	17.669	-5.614	0.000	0.000*

Year1-Year2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
Knowledge Domain ISS					
2020/21-2021/22	-44.605	17.604	-2.534	0.011	0.034*
2020/21-2022/23	-179.032	17.126	-10.454	0.000	0.000*
2021/22-2022/23	-134.427	17.680	-7.603	0.000	0.000*

* $p < 0.05$

For the MU, BCP, CS, and CUC domains, significant differences in students' mastery were found between the academic years 2020 and 2021, as well as between 2020 and 2022 ($p < 0.05$, adjusted using the Bonferroni correction). However, no significant difference was found between the years 2021 and 2022 (see Table 6).

Table 6: Pairwise comparisons of mastery estimates in domains MU, BCP, CS and CUC across academic years

Year1-Year2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
Knowledge Domain MU					
2020/21-2021/22	-61.646	17.499	-3.523	0.000	0.001*
2020/21-2022/23	-98.732	17.024	-5.800	0.000	0.000*
2021/22-2022/23	-37.086	17.574	-2.110	0.035	0.105
Knowledge Domain BCP					
2020/21-2021/22	-109.012	16.917	-6.444	0.000	0.000*
2020/21-2022/23	-141.896	17.389	-8.160	0.000	0.000*
2021/22-2022/23	32.884	17.464	1.883	0.060	0.179
Knowledge Domain CS					
2020/21-2021/22	-90.395	16.695	-5.414	0.000	0.000*
2020/21-2022/23	-106.584	17.161	-6.211	0.000	0.000*
2021/22-2022/23	16.188	17.235	0.939	0.348	1.000
Knowledge Domain CUC					
2020/21-2021/22	-100.635	17.143	-5.870	0.000	0.000*
2020/21-2022/23	-101.056	17.622	-5.735	0.000	0.000*
2021/22-2022/23	0.421	17.697	0.024	0.981	1.000

* $p < 0.05$

The pairwise comparisons largely confirmed these descriptive patterns. Significant differences were observed across all three cohort comparisons for IS, IOU, and ISS. For MU, BCP, CS, and CUC, significant differences were found between 2020/21 and each of the two later cohorts, whereas differences between 2021/22 and 2022/23 were not statistically significant.

6.3 Analysis of Final Grades

Complementing the mastery analysis in the previous section, a similar analysis was conducted on students' final grades over the three-year period. Normality tests (Kolmogorov-Smirnov and Shapiro-Wilk) revealed that the grades were not normally distributed. For all three academic years, the reported significance statistic from both tests was 0.000 (at $p = 0.05$, with degrees of freedom of 222, 196, and 218 for the years 2020, 2021, and 2022, respectively). As a result, a non-parametric Kruskal-Wallis H-test was used to assess year-to-year differences in students' final grades.

A visual inspection of the box plot for students' final grades (Figure 8) reveals that the median final grade (1) was identical in both 2020 and 2021, corresponding to the 1st quartile. The average final grades for these years were

1.48 in 2020 and 1.60 in 2021. In 2022, the median grade increased to 2, aligning with the 3rd quartile, and the average final grade for 2022 was 1.82.

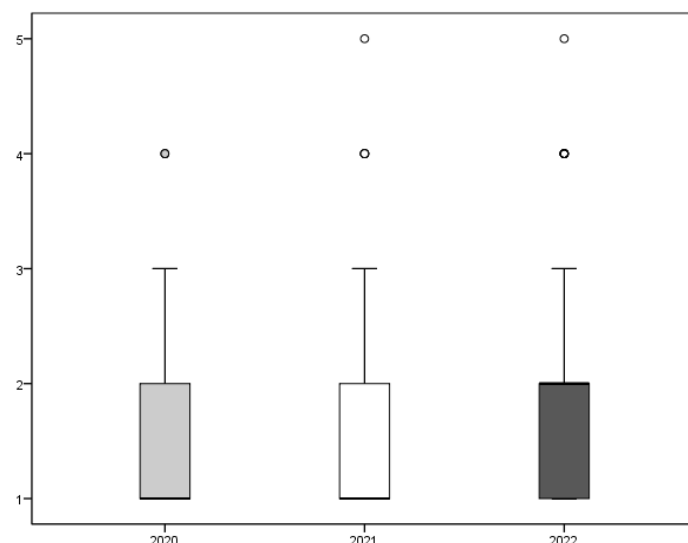


Figure 8: Box plots of students' final grades in each academic year

To assess whether these visually observed differences were statistically significant, Kruskal-Wallis H tests were conducted. Its results have revealed a significant difference in students' final grades across academic years ($\chi^2 = 19.491$, $df = 2$, $p = 0.000$, $\alpha = 0.05$). Dunn-Bonferroni pairwise comparisons (Table 7) showed significant differences between 2020/21 and 2022/23 and between 2021/22 and 2022/23, while the difference between 2020/21 and 2021/22 was not statistically significant.

Table 7: Pairwise comparisons of final grades across academic years

Year1-Year2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj. Sig.
2020-2021	-26.351	16.072	-1.640	0.101	0.303
2020-2022	-68.510	15.635	-4.382	0.000	0.000*
2021-2022	-42.159	16.141	-2.612	0.009	0.027*

* $p < 0.05$

In the end, it is worth noting that the quantitative results (Kruskal-Wallis H-test and additional Bonferroni pairwise comparisons in Table 7) align with the qualitative visual observations in Figure 8. Specifically, the differences in students' final grades are not statistically significant when comparing years 2020/21 and 2021/22, but are statistically significant when comparing 2020/21 and 2022/23, as well as 2021/22 and 2022/23.

7. Discussion

This study sets out to explore how student profiles and academic outcomes evolve across three consecutive years with varying teaching modalities (fully online, blended, and on-site) while keeping course content, assessment, and instructional staff consistent. This longitudinal setting enabled the examination of behavioural and performance patterns across three consecutive cohorts experiencing different teaching contexts, addressing all three research objectives and both research questions.

From the perspective of the Learning Analytics Cycle, the identified learner profiles represent the analytical outcome of transforming learner-generated digital traces into actionable educational knowledge. Although the present study does not evaluate instructional interventions, the resulting profiles establish a basis for future intervention-oriented Learning Analytics research.

7.1 Learner Profiles Across Cohorts and Teaching Modalities

In response to RO1, RO2 and RQ1, the research findings suggest that students can be consistently categorised into three broadly comparable clusters across the three consecutive academic years, while comparison of their

characteristics reveals both recurring and some cohort-specific patterns. Cluster C1 primarily consists of higher-performing students with passing grades, Cluster C2 is mostly composed of failing students, while Cluster C3 represents struggling students and includes both failing students and a minority of students who pass with low grades. It should be noted however that these labels were assigned only after clustering, based on the post hoc examination of final-grade distributions while final grades were not used to form the clusters. Although the relative composition of these clusters varies across years, their overall structure remains generally comparable across all three cohorts. These findings extend our previous single-cohort study (Balaban, Filipović and Zlatović, 2023) by showing that broadly comparable learner profiles also emerge in subsequent cohorts experiencing different teaching contexts.

This is also in line with similar research examining student performance and final scores. For instance, Quesada, Valverde and Rojas (2022) analysed Moodle activity logs using Learning Analytics techniques and clustering algorithms, finding that k-means clustering with three clusters yielded the best results, grouping students based on task-related behaviour and academic performance. Brown, White and Power (2016) conducted a two-step cluster analysis on the summative assessment scores of 773 undergraduate students, identifying three groups: High Achievers, Standard Achievers, and Low Achievers. The findings of the present study add a longitudinal perspective by showing that broadly comparable performance-based learner profiles can recur across consecutive cohorts within the same course.

However, students' activity levels per cluster were not as consistent across the three years as their performance-based profiles. Failing students (Cluster C2) consistently exhibited the lowest activity relative to the other profiles within each cohort. In contrast, the relative activity levels of higher-performing students (Cluster C1) and struggling students (Cluster C3) varied considerably. During the first year (2020/21, full lockdown), struggling students were more active than higher-performing students in both lesson-reading attempts and self-assessment quiz submissions. This pattern reversed in the second year (2021/22, partial lockdown), while by the final year (2022/23, no lockdowns), the activity levels of C1 and C3 were nearly identical. For lesson-access activity, these comparisons refer specifically to the relative positioning of clusters within each cohort because lesson-access frequencies were normalized separately by academic year - they should not be interpreted as differences in absolute lesson-access frequency between cohorts. Therefore, while Cluster C2 consistently showed the lowest relative activity within each cohort, higher activity alone did not distinguish higher-performing from struggling students. Some very active students still failed the course, while less active students managed to pass with high grades. This apparent activity-performance paradox suggests that the frequency of LMS activity alone does not capture how effectively students engage with learning resources and may reflect differences in prior knowledge, motivation, learning habits, or the strategies students use when interacting with course materials. This is consistent with Cavanaugh, Hargis and Mayberry (2016), Charbonneau-Gowdy (2018) and Van Goidsenhoven et al. (2020), who similarly reported that high levels of online activity do not necessarily correspond to academic success.

From a Learning Analytics perspective, these findings illustrate how learner-generated data can be transformed into interpretable learner profiles within the Metrics element of Clow's (2012) Learning Analytics Cycle. Importantly, the identified profiles show that broadly comparable performance-based group structures can coexist with substantially different patterns of LMS activity across cohorts.

7.2 Student Performance Across Cohorts and Teaching Modalities?

In addressing RO3 and RQ2 ("What differences in student performance can be observed across three consecutive cohorts in terms of knowledge-domain mastery and final grades?"), the findings indicate noticeable variability in student performance over the 2020/21 to 2022/23 period. At the level of individual knowledge domains, the analysis revealed two distinct patterns. Four out of seven domains demonstrated continuous yearly improvements in mastery, while the remaining three showed improvement between the first and second years followed by stagnation or a slight decline between the second and third years. At the level of final grades, no statistically significant difference was found between the first (2020/21) and second (2021/22) cohorts. However, significant differences were noted when the third (2022/23) cohort was compared with both preceding cohorts. Based on such findings, it can be concluded that changes in knowledge-domain mastery were already observable in the second year, while more substantial differences in final course performance became evident only in the third year.

Variability in academic performance between different cohorts of first-year students can be related to several factors, as highlighted in the literature. These include individual adjustment to the higher education environment (López, Santelices and Taveras, 2023) and rather selective nature of first-year courses (Jury,

Smeding and Darnon, 2015), as well as the alignment of course design with the expectations of new student generations (Roberson, 2011). In the present study, another important contextual factor systematically differed across the observed years: the teaching conditions associated with COVID-19 lockdown measures. Over the three-year period, the learning environment changed from fully online teaching during full lockdown, through blended teaching under relaxed measures, to traditional on-site teaching without lockdown restrictions.

Previous research has documented differences in academic performance associated with the COVID-19 pandemic and emergency remote teaching. Vanderstraeten et al. (2023) reported that academic performance among first-year students was indirectly affected by declines in well-being and self-efficacy during the pandemic, while Burkholder and Salehi (2023) found that the shift to remote teaching contributed to existing demographic gaps and affected student groups unevenly. Decreased academic performance during the pandemic has also been reported in other studies (e.g., Wang et al., 2024). The pattern observed in the present study is consistent with this broader literature: higher levels of knowledge-domain mastery and final course performance were observed in the later cohorts as teaching shifted from fully online to blended and subsequently to on-site delivery. However, because the three cohorts comprised different groups of students and individual and cohort-level characteristics were not controlled, these differences cannot be attributed exclusively to teaching modality.

7.3 Study Contribution and Practical Implications

The main contribution of this study lies in extending our previous single-cohort analysis (Balaban, Filipović and Zlatović, 2023) to three consecutive cohorts of the same course experiencing different teaching modalities. While the previous study demonstrated that meaningful learner profiles could be identified during emergency remote teaching, the present longitudinal analysis shows that broadly comparable performance-based profiles recur across subsequent cohorts, despite changes in teaching context. At the same time, the activity patterns associated with these profiles varied considerably across cohorts. This combination of relative stability in performance-based profiles and variability in LMS activity provides a more nuanced understanding of learner profiles than could be obtained from a single-cohort analysis.

While the specific C1–C3 profiles identified in this study are context-dependent and should not be assumed to generalise directly to other courses, the analytical approach itself is transferable to other LMS-supported learning contexts in which comparable behavioural and performance data are available. In this sense, the study demonstrates how the Learners–Data–Metrics sequence of the Learning Analytics Cycle can be operationalised to examine whether meaningful learner profiles recur across cohorts and teaching contexts.

From a practical perspective, the identified profiles demonstrate how LMS-generated data can support the recognition of different groups of learners who may require different forms of educational support. Continuously low activity, as observed for C2, may help identify students requiring additional attention, while the C3 profile shows that high activity alone should not be interpreted as evidence of successful learning. In this sense, students exhibiting a C3-type profile may represent a particularly relevant group for monitoring and targeted support at appropriate points during the course. In terms of Clow's (2012) Learning Analytics Cycle, these learner profiles represent an analytical output of the Metrics element that may inform the design of differentiated support within the subsequent Interventions element. However, such interventions were not implemented or evaluated in the present study and should therefore be examined in future research.

7.4 Limitations

Several limitations should be considered when interpreting the findings. First, although the course content, instructional staff, assessment methods, and overall course structure remained consistent across the three academic years, each cohort comprised a different group of first-year students. Individual and cohort-level characteristics, including prior knowledge, motivation, demographics, and learning habits, were not controlled. Therefore, the observed differences cannot be attributed exclusively to changes in teaching modality.

Second, the study was conducted within a single course and institutional context, which limits the generalisability of the identified learner profiles and performance patterns. While the analytical approach may be transferable to other LMS-supported learning contexts, the recurrence and characteristics of the specific learner profiles identified in this study should be examined in other courses and educational settings.

Finally, LMS activity was operationalised through the frequency of access to course materials and self-assessment activities. These measures capture the quantity of observable LMS activity but not the quality,

duration, purpose, or effectiveness of students' engagement with learning resources. Consequently, similar levels of recorded activity may reflect substantially different learning behaviours and strategies.

8. Conclusion

This study provides a longitudinal Learning Analytics perspective on student behaviour and performance across three consecutive cohorts of the same course delivered under fully online, blended, and on-site teaching conditions. Building on our previous single-cohort study (Balaban, Filipović and Zlatović, 2023), the findings show that broadly comparable learner profiles can be identified across cohorts despite changes in teaching context. Three recurring profiles were observed: higher-performing students (C1), failing students (C2), and struggling students (C3). While their performance-based structure remained broadly recognisable across cohorts, their LMS activity patterns varied considerably.

A particularly relevant finding concerns the relationship between LMS activity and academic performance. Failing students (C2) consistently showed lower activity relative to the other profiles within each cohort, suggesting that comparatively low LMS activity may provide a useful signal for identifying students requiring additional attention. However, higher levels of activity did not consistently distinguish higher-performing students from struggling students, indicating that activity frequency alone provides an incomplete representation of student engagement and learning. From a practical Learning Analytics perspective, such profiles may help instructors distinguish between students showing very low engagement and those who remain active but nevertheless experience difficulties, thereby providing an analytical basis for more differentiated educational support.

The analysis of knowledge-domain mastery and final grades further revealed differences across the three cohorts. Improvements in mastery were already observable in the second cohort, while statistically significant differences in final grades emerged primarily for the third cohort. Although these patterns coincided with the transition from fully online to blended and subsequently to on-site teaching, they cannot be attributed exclusively to teaching modality because the study involved different cohorts and did not control for individual and cohort-level characteristics.

The main contribution of the study therefore lies in extending a single-cohort learner profiling approach to a longitudinal setting and demonstrating that broadly comparable performance-based profiles can recur across consecutive cohorts while their LMS activity patterns vary across teaching contexts. While the specific profiles identified here should not be assumed to generalise directly beyond the course and institutional context studied, the Learning Analytics approach used to transform learner data into interpretable profiles is potentially transferable to other LMS-supported educational settings.

Future studies should examine whether comparable learner profiles emerge across other courses and educational contexts. They should also incorporate additional learner-level characteristics, such as prior knowledge, demographics, motivation, and learning habits, to better explain differences between cohorts and, in particular, why high levels of LMS activity may be associated with successful performance for some learners but not for others.

Ethics Statement: The study used routinely collected educational data generated as part of regular course activities. Student identifiers were used only to match data from the LMS and the National University Information System during dataset preparation, after which the analytical dataset was anonymised. No additional data were collected directly from students for research purposes. In accordance with institutional practice for research based on anonymised educational data, formal ethics approval was not required.

AI tools Statement: AI tools have not been used in development of this paper's main content. ChatGPT (v5) was used to improve the language quality of certain manuscript sections. Authors hereby confirm that they have thoroughly reviewed and approved the final text.

Competing Interests: The authors of this publication declare there are no competing interests.

Funding: This research was supported by the Croatian Science Foundation (Hrvatska Zaklada za Znanost) [project IP-2020-02-5071].

References

- Abdulkareem Shafiq, D., Marjani, M., Ahamed Ariyaluran Habeeb, R. and Asirvatham, D., 2025. Digital Footprints of Academic Success: An Empirical Analysis of Moodle Logs and Traditional Factors for Student Performance. *Education Sciences*, 15(3), 304. <https://doi.org/10.3390/educsci15030304>
- Alalawi, K., Athauda, R. and Chiong, R., 2025. An Extended Learning Analytics Framework Integrating Machine Learning and Pedagogical Approaches for Student Performance Prediction and Intervention. *International Journal of Artificial Intelligence in Education*, 35(3), pp 1239-1287. <https://doi.org/10.1007/s40593-024-00429-7>
- Argyriou, P., Benamar, K. and Nikolajeva, M., 2022. What to blend? Exploring the relationship between student engagement and academic achievement via a blended learning approach. *Psychology, Learning and Teaching*, 21(2), pp 126-137. <https://doi.org/10.1177/14757257221091512>
- Baker, R.S., Gašević, D. and Karumbaiah, S., 2021. Four paradigms in learning analytics: Why paradigm convergence matters. *Computers and Education: Artificial Intelligence*, 2, 100021. <https://doi.org/10.1016/j.caeai.2021.100021>
- Balaban, I., Filipović, D. and Zlatović, M., 2023. Post hoc identification of student groups: Combining user modeling with cluster analysis. *Education and Information Technologies*, 28(6), pp 7265-7290. <https://doi.org/10.1007/s10639-022-11468-9>
- Bergdahl, N., Nouri, J., Karunaratne, T., Afzaal, M. and Saqr, M., 2020. Learning analytics for blended learning: A systematic review of theory, methodology, and ethical considerations. *International Journal of Learning Analytics and Artificial Intelligence for Education (iJAI)*, 2(2), pp 46–79. <https://doi.org/10.3991/ijai.v2i2.17887>
- Bond, M., Bedenlier, S., Marín, V. and Händel, M., 2021. Emergency remote teaching in higher education: Mapping the first global online semester. *International Journal of Educational Technology in Higher Education*, 18(1), 50. <https://doi.org/10.1186/s41239-021-00282-x>
- Britain, S., 2004. A Review of Learning Design: Concept, Specifications and Tools. A report for the JISC E-learning Pedagogy Programme, 2006. <https://staff.blog.ui.ac.id/harrybs/files/2008/10/learningdesigntoolsfinalreport.pdf>
- Broadbent, J., Sharman, S.J., Panadero, E. and Fuller-Tyszkiewicz, M., 2021. How does self-regulated learning influence formative assessment and summative grade? Comparing online and blended learners. *The Internet and Higher Education*, 50, 100805. <https://doi.org/10.1016/j.iheduc.2021.100805>
- Brown, S., White, S. and Power, N., 2016. Cluster analysis of assessment in anatomy and physiology for health science undergraduates. *International Journal of Teaching and Learning in Higher Education*, 28(1), pp 102-109. <https://www.isetl.org/ijthe/pdf/IJTLHE2114.pdf>
- Buckley, K., Fairman, K.A., Pogge, E.K. and Raney, E.C., 2021. Use of Learning Management System Data to Predict Student Success in a Pharmacy Capstone Course. *American Journal of Pharmaceutical Education*, 86(4), 8594. <https://doi.org/10.5688/ajpe8594>
- Burkholder, E.W. and Salehi, S., 2023. Effects of the COVID-19 pandemic on academic preparation and performance: a complex picture of equity. *Frontiers in Education*, 8, 1126441. <https://doi.org/10.3389/feduc.2023.1126441>
- Cavanaugh, C., Hargis, J. and Mayberry, J., 2016. Participation in the virtual environment of blended college courses: An activity study of student performance. *International Review of Research in Open and Distributed Learning*, 17(3), pp 263-275. <http://dx.doi.org/10.19173/irrodl.v17i3.1811>
- Charbonneau-Gowdy, P., 2018. Beyond stalemate: Seeking solutions to challenges in online and Blended Learning programs. *Electronic Journal of E-learning*, 16(1), pp 56-66. <https://academic-publishing.org/index.php/ejel/article/view/1856/1819>
- Chiang, M. and Mirkin, B., 2010. Intelligent Choice of the Number of Clusters in K-Means Clustering: An Experimental Study with Different Cluster Spreads. *Journal of Classification*, 27, pp 3-40. <https://doi.org/10.1007/s00357-010-9049-5>
- Chrysafiadi, K. and Virvou, M., 2013. Student modeling approaches: A literature review for the last decade. *Expert Systems with Applications*, 40, pp 4715-4729. <https://doi.org/10.1016/j.eswa.2013.02.007>
- Clow, D., 2012. The learning analytics cycle: Closing the loop effectively. In: *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12)*. Vancouver, BC, Canada, 29 April - 2 May. New York: ACM, pp 134-138. <https://doi.org/10.1145/2330601.2330636>
- De Amorim, R.C. and Hennig, C., 2015. Recovering the number of clusters in data sets with noise features using feature rescaling factors. *Information Sciences*, 324, pp 126–145. <https://doi.org/10.1016/j.ins.2015.06.039>
- De Philippis, M. and Rossi, F., 2016. Parents, schools and human capital differences across countries. *Social Science Research Network*. <https://doi.org/10.2139/ssrn.2896780>
- Dutt, A., Ismail, M.A. and Herawan, T., 2017. A Systematic Review on Educational Data Mining. *IEEE Access*, 5, pp 15991–16005. <https://doi.org/10.1109/ACCESS.2017.2654247>
- Elejalde, E. and Wildemann, S., 2026. Information loss in aggregated social media discussions: studying the migration discourse within Europe. *EPJ Data Science*, 15, 31. <https://doi.org/10.1140/epjds/s13688-026-00641-6>
- El Aissaoui, O., Oughdir, L. and El Alloui, Y., 2022. A Literature Review on Student Modeling Purposes. In: *Advanced Intelligent Systems for Sustainable Development (AI2SD'2020): Advances in Intelligent Systems and Computing*, vol 1417. Tangier, Morocco, 21-26 December. Cham: Springer, pp 758-784. https://doi.org/10.1007/978-3-030-90633-7_64
- Gaftandzhieva, S., Talukder, A., Gohain, N., Hussain, S., Theodorou, P., Salal, Y.K., and Doneva, R., 2022. Exploring online activities to predict the final grade of students. *Mathematics*, 10(20), 3758. <https://doi.org/10.3390/math10203758>

- Han, F., Vaculíková, J. and Juklová, K., 2022. The relations between Czech undergraduates' motivation and emotion in self-regulated learning, learning engagement, and academic success in blended course designs: Consistency between theory-driven and data-driven approaches. *Frontiers in Psychology*, 13, 1001202. <https://doi.org/10.3389/fpsyg.2022.1001202>
- Hoyos Osorio, J.K. and Daza Santacoloma, G., 2023. Predictive model to identify college students with high dropout rates. *Revista Electrónica de Investigación Educativa*, 25(13), pp 1-10. <https://doi.org/10.24320/redie.2023.25.e13.5398>
- Hussain, S., Muhsin, Z., Salal, Y., Theodorou, P., Kurtoğlu, F. and Hazarika, G., 2019. Prediction model on student performance based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning*, 14(8), pp 4-22. <https://doi.org/10.3991/ijet.v14i08.10001>
- Ifenthaler, D. and Yau, J. Y.K., 2020. Utilising learning analytics to support study success in higher education: a systematic review. *Educational Technology Research and Development*, 68, pp 1961-1990. <https://doi.org/10.1007/s11423-020-09788-z>
- Jain, A.K., 2010. Data clustering: 50 years beyond K-means, *Pattern Recognition Letters*, 31(8), pp 651-666. <https://doi.org/10.1016/j.patrec.2009.09.011>
- Jha, S., Jha, M. and O'Brien, L., 2019. Analysing Computer Science Course Using Learning Analytics Techniques. In: *2019 IEEE Asia-Pacific Conference on Computer Science and Data Engineering (CSDE)*. Melbourne, VIC, Australia, 9-11 December. New York: IEEE, pp 1-6. <https://doi.org/10.1109/CSDE48274.2019.9162369>
- Jury, M., Smeding, A. and Darnon, C., 2015. First-generation students' underperformance at university: The impact of the function of selection. *Frontiers in Psychology*, 6, 710. <https://doi.org/10.3389/fpsyg.2015.00710>
- López, M.J., Santelices, M.V. and Taveras, C.M., 2023. Academic performance and adjustment of first-generation students to higher education: a systematic review. *Cogent Education*, 10(1), 2209484. <https://doi.org/10.1080/2331186X.2023.2209484>
- Lowenthal, P., Borup, J., West, R. and Archambault, L., 2020. Thinking Beyond Zoom: using asynchronous video to maintain connection and engagement during the COVID-19 pandemic. *Journal of Technology and Teacher Education*, 28, pp 383-391. <https://doi.org/10.70725/331114ctnuti>
- Mao, Y., Marwan, S., Price, T.W., Barnes, T. and Chi, M., 2020. What time is it? Student modeling needs to know. In: *Proceedings of the 13th International Conference on Educational Data Mining (EDM 2020)*. Online, 10-13 July. n.p.: International Educational Data Mining Society, pp 171-182. <http://files.eric.ed.gov/fulltext/ED608054.pdf>
- Mitić, J. and Djeniĉ, S., 2023. Improvement of learning outcomes in traditional learning model by introducing online activities and big data analytics. *Interactive Learning Environments*, 32(7), pp 3783-3798. <https://doi.org/10.1080/10494820.2023.2190362>
- Nespereira, C.G., Dai, K., Redondo, R.P.D. and Vilas, A.F., 2014. Is the LMS access frequency a sign of students' success in face-to-face higher education?. In: *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM '14)*. Salamanca, Spain, 1-3 October. New York: ACM, pp 283-290. <https://doi.org/10.1145/2669711.2669912>
- Pande, S.M., 2023. Machine learning models for student performance prediction. In: *2023 International Conference on Innovative Data Communication Technologies and Application (ICIDCA)*. Uttarakhand, India, 2023, pp 27-32. <https://doi.org/10.1109/icidca56705.2023.10099503>
- Pecuchová, J. and Drlík, M., 2024. Enhancing the Early Student Dropout Prediction Model Through Clustering Analysis of Students' Digital Traces. *IEEE Access*, 12, pp 159336-159367. <https://doi.org/10.1109/ACCESS.2024.3486762>
- Purwoningsih, T., Santoso, H.B. and Hasibuan, Z.A., 2020. Data Analytics of Students' Profiles and Activities in a Full Online Learning Context. In: *2020 Fifth International Conference on Informatics and Computing (ICIC)*. Gorontalo, Indonesia, 3-4 November. New York: IEEE, pp 1-8. <https://doi.org/10.1109/icic50835.2020.9288540>
- Quesada, A.M., Valverde, F.L. and Rojas, S.G., 2022. Clustering students' online behavior and relation with academic performance. In: *2022 International Conference on Electrical, Computer and Energy Technologies (ICECET)*. Prague, Czech Republic, 20-22 July. New York: IEEE, pp 1-6. <https://doi.org/10.1109/ICECET55527.2022.9872960>
- Rajagukguk, S.A. and Fudholi, D.H., 2022. Rich kid, poor kid: Students clustering using K-Prototype algorithm. In: *2022 IEEE 7th International Conference on Information Technology Development and Applications (ICITDA)*. Yogyakarta, Indonesia, 4-5 November. New York: IEEE, pp 1-4. <https://doi.org/10.1109/ICITDA55840.2022.9971445>
- Roberson, C., 2011. Aligning generations to improve retention in introductory computing courses. *Journal of Computing Sciences in Colleges*, 26(6), pp 30-36.
- Rousseeuw, P.J., 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, pp 53-65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Seel, N., 2012. Programmed Learning. In: Seel N.M. (Ed.), *Encyclopedia of the Sciences of Learning*. Springer, Boston, MA, USA. https://doi.org/10.1007/978-1-4419-1428-6_671
- Siemens, G., 2013. Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), pp 1380-1400. <https://doi.org/10.1177/0002764213498851>
- Steinley, D., 2006. K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), pp 1-34. <https://doi.org/10.1348/000711005X48266>
- Šarić, I., Grubišić, A., Šerić, L. and Robinson, T.J., 2019. Data-Driven Student Clusters Based on Online Learning Behavior in a Flipped Classroom with an Intelligent Tutoring System. In: *Intelligent Tutoring Systems (ITS 2019): Lecture Notes in Computer Science, vol 11528*. Kingston, Jamaica, 3-7 June. Cham: Springer, pp 72-81. https://doi.org/10.1007/978-3-030-22244-4_10

- Trittenbach, H., Bach, J. and Böhm, K., 2019. Understanding the effects of temporal energy-data aggregation on clustering quality. *IT - Information Technology*, 61 (2-3), pp 111-123. <https://doi.org/10.1515/itit-2019-0014>
- Tseng, S.F., Tsao, Y.W., Yu, L.C., Chan, C.L. and Lai, K.R., 2016. Who will pass? Analyzing learner behaviors in MOOCs. *Research and Practice in Technology Enhanced Learning*, 11, 8. <https://doi.org/10.1186/s41039-016-0033-5>
- Van Goidsenhoven, S., Bogdanova, D., Deeva, G., Broucke, S.V., De Weerd, J. and Snoeck, M., 2020. Predicting student success in a blended learning environment. In: *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge (LAK '20)*. Frankfurt, Germany, 23-27 March. New York: ACM, pp 17-25. <https://doi.org/10.1145/3375462.3375494>
- Van Leeuwen, A., van Wermeskerken, M., Erkens, G. and Rummel, N., 2017. Measuring teacher sense making strategies of learning analytics: A case study. *Learning: Research and Practice*, 3(1), pp 42-58. <https://doi.org/10.1080/23735082.2017.1284252>
- Vanderstraeten, L., Opdecam, E., Everaert, P. and Beyers, W., 2023. The effect of COVID-19 on the well-being of first-year university students. *Educational Psychology*, 43(5), pp 407-429. <https://doi.org/10.1080/01443410.2023.2223787>
- Wang, Y., Zhang, Y., Wang, J., Ge, W., Wang, L., Jia, N., Li, S. and Li, D., 2024. Impact of campus closure during COVID-19 on lifestyle, educational performance, and anxiety levels of college students in China. *BMC Public Health*, 24, 2218. <https://doi.org/10.1186/s12889-024-19744-8>
- Watson, S.L., Watson, W.R., Yu, J.H., Alamri, H. and Mueller, C., 2017. Learner profiles of attitudinal learning in a MOOC: An explanatory sequential mixed methods study. *Computers & Education*, 114, pp 274-285. <https://doi.org/10.1016/j.compedu.2017.07.005>
- Wong, B.T.M. (2017). Learning analytics in higher education: an analysis of case studies. *Asian Association of Open Universities Journal*, 12(1), pp 21-40. <http://dx.doi.org/10.1108/AAOUJ-01-2017-0009>
- Xu, Z., Choudhury, F.H., Ma, S., Murphrey, T.P. and Dooley, K.E., 2025. Leveraging Learning Analytics to Model Student Engagement in Graduate Statistics: A Problem-Based Learning Approach in Agricultural Education. *Behavioral Sciences*, 15(10), 1360. <https://doi.org/10.3390/bs15101360>
- Yan, C., 2022. Research on Student Academic Performance Prediction Methods. *Highlights in Science, Engineering and Technology*, 24, pp 257-263. <https://doi.org/10.54097/hset.v24i.3940>
- Yildirim-Erbasli, S., Dibek, M.I. and Taikh, A., 2026. Cognitive learning analytics. *Encyclopedia*, 6(3), 69. <https://doi.org/10.3390/encyclopedia6030069>
- You, J.W., 2016. Identifying significant indicators using LMS data to predict course achievement in online learning. *The Internet and Higher Education*, 29, pp 23-30. <https://doi.org/10.1016/j.iheduc.2015.11.003>