ORIGINAL ARTICLE

Journal of Computer Assisted Learning **WILEY**

How can valid and reliable automatic formative assessment predict the acquisition of learning outcomes?

Blaženka Divjak ⓘ | **Barbi Svetec** ⓘ | **Damir Horvat** ⓘ

Faculty of Organization and Informatics, University of Zagreb, Varaždin, Croatia

Correspondence
Blaženka Divjak, Faculty of Organization and Informatics, University of Zagreb, Pavlinska 2, Varaždin, Croatia.
Email: bdivjak@foi.hr

Abstract

Background: Sound learning design should be based on the constructive alignment of intended learning outcomes (LOs), teaching and learning activities and formative and summative assessment. Assessment validity strongly relies on its alignment with LOs. Valid and reliable formative assessment can be analysed as a predictor of students' academic performance, but the question is how significant its predictive power is, and what other elements can affect predictions.

Objectives: Our aim was to investigate the predictive power of formative assessment for summative assessment, measuring the acquisition of LOs.

Methods: We analysed formative assessment results (quizzes, homework), together with log data (video and other material use, class attendance), to determine the most influential predictors and establish a reliable predictive learning analytics model. We used the Random Forest algorithm. The model is based on the data from two university mathematical courses, delivered at different years and levels of study, incorporating 813 students in two consecutive years.

Results and Conclusions: Our results show that formative assessment, together with previous summative assessment, is a stronger predictor of summative assessment results than other data on students' engagement. The study pointed to the importance of completeness and quality of data, and clear links between assessment and LOs when making predictions of student results. It suggested that predictions are less reliable for the lowest and the highest performing students. It was noted that other factors can also affect predictions, like the level of LOs, or factors not easily extracted from digital data, like the learning environment and individual students' strategies.

KEYWORDS
assessment, learning analytics, learning outcomes, predictive model, Random Forest

1 | INTRODUCTION

Educational strategies anchored in well-defined learning outcomes (LOs) form the backbone of effective pedagogy and sound learning design (LD), leading to the quality and relevance of higher education. Approaches based on LOs are a prerequisite for student-centred teaching and learning, as LOs state what learners know and are able to do after a learning process, which needs to be confirmed in the process of assessment. LOs contain both the content and the level of performance (e.g., using Bloom's taxonomy), which enable the measurement of LOs in the assessment process. As such, LOs are the basis for criteria-based assessment and ensuring assessment validity (Divjak et al., 2023). The alignment between intended LOs, teaching and learning activities, and both formative and summative assessments (Biggs, 1999), ensures that the educational process is holistic and considers actual learning progress.

As the digital age and the use of big data are rapidly transforming education, the importance of using data from these assessments to predict and enhance student outcomes and inform LD has grown. Predictive learning analytics (LA) can significantly support the acquisition of LOs, identification of students at risk, continuously inform LD and also raise the satisfaction of students (Sghir et al., 2023). Development of predictive LA models has often been data-driven, putting more focus on the accuracy of predictions than on theory and practical pedagogical implications (Bulut et al., 2023).

Recent research has considered the prominent role of formative assessment in developing predictive analytics models, as an upgrade to earlier research focused primarily on predictors related to student engagement in online learning environments (Saqr et al., 2022). In particular, recent research has pointed out that instead of such complex variables, predictive LA models should be informed by LD and theory and based on data extracted from online formative assessments (Bulut et al., 2023), using the predictive power of 'true' assessment data (Tempelaar et al., 2016).

Formative assessment, with continuous and immediate feedback, offers insights not only into immediate progress, but also into the eventual summative performance of students. However, this only makes sense if formative and summative assessment are aligned mutually, and with the intended LOs. Moreover, assessment should be internally consistent. In order to present a meaningful predictor of student performance, assessment needs to be valid and reliable (van der Vleuten & Schuwirth, 2005). But how significant is this predictive power? And can other types of student engagement, such as interactions with digital resources or attendance, also provide additional foresight into academic performance?

Our study brings clarity to these interrelationships, based on an advanced machine learning algorithm and digital datasets from two mathematical courses. The study used the Random Forest (RF) algorithm (already recognised in predictive analytics as an accurate classifier) in a specific context of mathematical education, characterised by student-centred LD.

While preliminary insights highlight the importance of formative assessment as a predictor, this research also delves deeper, examining the specifics of data quality, the level of LOs, and the role of pedagogical and individual teaching and learning strategies that influence student success.

2 | THEORETICAL BACKGROUND

2.1 | Formative and summative assessment

If planned and conducted meaningfully, assessment not only serves for reporting on student progress, but it can also support and steer students in their learning processes, and teachers in making informed teaching decisions (Ramsden & Ramsden, 2003). Assessment programs can include two types of assessment: formative and summative. On the one hand, formative assessment refers to collecting data in order to improve students' learning. On the other hand, summative assessment means using data in order to assess students' knowledge after the completion of a particular learning sequence. (American Educational Research Association, American Psychological Association and the National Council on Measurement in Education, 2014; Dixson & Worrell, 2016) In other words, formative assessment is continuous, done throughout a course, and helps direct the teaching and learning process toward successful acquisition of LOs, while summative assessment is used to evaluate the acquisition of LOs at the end of a unit of learning. However, as pointed out by Ramsden and Ramsden (2003), "the two separate worlds of assessment called 'formative' and 'summative' in the assessment manuals do not exist in reality", as they are mutually closely related. Formative and summative assessment with feedback have been recognised in previous studies (Divjak, Žugec, et al., 2022) as one of the factors (based on factor analysis) in a model encapsulating the student perspective on e-assessment, emphasising that formative and summative assessments should be cohesively aligned within an assessment program. Other factors in the model revealed in the same study were the transparency and fairness of assessment, meaningful use of technology in assessment, and difficulty of LOs.

2.2 | Assessment validity

To be meaningful, both types of assessment need to comply with reliability and validity standards (American Educational Research Association, American Psychological Association and the National Council on Measurement in Education, 2014), while also considering other elements of assessment utility: educational impact, acceptability and costs (van der Vleuten & Schuwirth, 2005).

Validity has been described as 'the degree to which evidence and theory support the interpretations of test scores for proposed uses of tests' (American Educational Research Association, American Psychological Association and the National Council on Measurement in Education, 2014). In educational assessment, validity relates to the relationship between content standards and the content of a test, and their mutual alignment (American Educational Research Association, American Psychological Association and the National Council on Measurement in Education, 2014). These standards are commonly referred to as LOs, and the alignment between LOs and assessment has been described as constructive alignment (Biggs, 1999). So, to ensure the validity of assessment, it is crucial to ensure its alignment with the intended LOs, making sure that assessment adequately measures what it intends to. Valid assessment is essential if we want to meaningfully steer and support learning, and is therefore important when developing student-centred LD. Ensuring the validity of assessment can be supported by LA, which can consider the prioritisation of LOs (Divjak et al., 2023). In this process, the Balanced LD Planning concept and tool (Rienties et al., 2023) can be used, as they are based on LOs, constructive alignment (Biggs, 1999), and refer to the levels of LOs according to Bloom's taxonomy (Anderson & Krathwohl, 2001).

2.3 | Predictive learning analytics and formative assessment

LA can have several types and purposes, with predictive LA being focused on using past and current data patterns to forecast future outcomes. They usually utilise machine learning algorithms, learning from historic datasets and making inferences about possible outcomes in the future. (Susnjak et al., 2022) In the last several years, there has been an increase in the number of studies predicting academic outcomes based on machine learning and deep learning models, using different kinds of student data, with the aim to improve learning processes. In this context, predictive modelling has emerged as a central practice in LA (Sghir et al., 2023). As a shortcoming, it has been pointed out that there is often a lack of transparency for users in terms of how these models make predictions (Susnjak et al., 2022).

Research has been conducted using formative assessment to predict summative assessment results and academic performance (e.g., Bulut et al., 2023; Ekolu, 2022; Tempelaar et al., 2015a, 2018). Researchers (Bulut et al., 2023) have claimed that, instead of variables like event logs, timestamps of activities or clickstream data, the basis for developing predictive LA models should be online formative assessments. It has been found that features related to computer-assisted and online formative assessment (e.g., scores, completion and timestamps) are significant predictors of students' academic performance (Bulut et al., 2023), with performance in formative assessment being the key predictor, though not the timeliest (Bulut et al., 2023; Tempelaar et al., 2015a, 2015b, 2018).

Studies have also dealt with a number of predictors other than formative assessment, for example LMS track data including time-on-task, learning disposition data, diagnostic entry tests, demographic data, and so on. (Saqr et al., 2017; Tempelaar et al., 2015a) However, a maths and statistics course study found that the strongest predictors are found in cognitive data, that is scores on entry tests and quizzes, whereas basic LMS data were not an important predictor of learning (Tempelaar et al., 2015a). Another study (Bulut et al., 2023) found that the number of clicks in the LMS and the time between first access and dates of formative assessment were valuable predictors. It was also found that the outcomes of formative assessment can be explained using trace data on student activity in e-tutorials (Tempelaar et al., 2018). Interestingly, a study in medicine found that factors reflecting engagement of students and consistency in using online resources were the most important predictors (Saqr et al., 2017). In relation to this, it has also been noted that the predictors of academic success are dependent on LD, subject area or educational institutions, which should be considered when developing predictive models (Saqr et al., 2022). One study has found that various prediction variables are affected by differences between subgroups in a sample (Tempelaar et al., 2015b).

Finally, a recent systematic review found various predictor variables used in research so far can be grouped as student-related, teacher-related and institutional features. Students' behavioural data (logs) were found to be the most frequent input for prediction, followed by students' academic data (grades). Other predictor variables included demographic and prior academic data and, rarely, students' psychological data, teachers' behavioural data or infrastructural features (Sghir et al., 2023).

The same systematic review (Sghir et al., 2023) has shown that researchers use several algorithms to select the best model for prediction, with Artificial Neural Networks being the most often used one, followed by Random Forest and Gradient Boosting algorithms. In previous research, the RF classifier has been linked with the highest accuracy and precision, with various hyperparameters presenting an advantage over other classifiers, leading to more accurate prediction (Kabathova & Drlik, 2021).

It should be noted that there are differences in the accuracy and efficiency of predictive models. For example, a study (Ekolu, 2022) used a model for probabilistic prediction, which was found to correctly estimate summative assessment results of students whose formative assessment results were between 50% and 70%. In another study, researchers were able to predict final grades with 63.5% accuracy and identify 53.9% of students at risk (Saqr et al., 2017).

3 | METHODOLOGY

We conducted a study to respond to the following research questions:

1. How can formative assessment results contribute to the prediction of summative assessment results?
2. In what ways can other contextual factors (e.g., levels of LOs, levels of study and student activity) affect predictions?
3. What elements can affect the efficiency of prediction?

3.1 | Study setting

The study was conducted at a higher education institution offering undergraduate and graduate IT programs. Two mathematical courses were included: Mathematics 1 (M1), an undergraduate course with approximately 320 full-time students enrolled each academic year, and Discrete Mathematics with Graph Theory (DMGT), a graduate course enrolling around 110 students annually. The two different mathematics courses were chosen because they are delivered at different levels of study and LOs. It was considered that students of different ages and at different levels of study might vary in terms of self-regulation, learning behaviour and strategies, including the use of learning resources. It has been found that students who are older or more experienced are more capable of differentiating learning strategies and demonstrate stronger interrelations between them (Vermunt & Vermetten, 2004). However, as the two courses are mathematical courses, they are related enough for their comparison to make sense.

The M1 course has a student workload of 5 ECTS credits (equivalent to 150 h of student work), while DMGT has 6 ECTS credits, including 10 weeks of classes and 5 weeks of project-based learning (PBL), amounting to 180 h of student work.

The M1 assessment program includes quizzes, homework assignments (formative assessment) and three periodical exams (summative assessment) as well as a mathematical essay (contributing 10% to the total grade). The DMGT assessment program includes quizzes (formative assessment) and two periodical exams (summative assessment) as well as an extensive PBL task (contributing 30% to the total grade). In this study, we focused on quizzes as formative assessment and periodical exams as summative assessment for analysis, as PBL and essays had their unique characteristics and LOs they are related with.

M1 students were divided into several groups, with some minor differences in tasks, which resulted in some formative assessment data (related to Exam 2) missing for one group. In DMGT, one LO was not covered by formative assessment. This is considered later in the text when analysing the results related to RQ3.

Formative tasks are personalised for each student. In both courses, the assignments consist of computational tasks, randomly selected from the Moodle LMS assignments database, and they are designed by the course teachers. In contrast, quizzes primarily concentrate on assessing students' grasp of concepts, fundamental terminology and their ability to solve tasks that enhance understanding. These quizzes offer automated formative assessment, employing an automated grading system coupled with feedback (Divjak, Žugec et al., 2022). The quizzes are conducted during class, while homework assignments (M1) are completed at home within predetermined time slots and then uploaded to the LMS. They contribute to the total grade with smaller percentages (between 10% and 20% in total).

To successfully complete the courses, students need to accumulate more than 50% of the total points during the semester. Failing to meet this requirement necessitates them to take exams during three additional examination periods.

Both courses were originally delivered in a blended mode prior to the pandemic, and they have transitioned to a hybrid approach, making resources and materials available through the Moodle LMS. In this hybrid format, students receive videos and reading materials via the LMS and are required to engage in classes, including lectures and seminars. While the majority of students attend on campus, some opt for remote participation by either attending lectures online in real-time or accessing the video recordings at a later point.

Sound student-centred LD was employed for both courses, utilising the Balanced Design Planning (BDP) concept and tool (Divjak, Grabar, et al., 2022), which emphasised intended LOs and

constructive alignment with teaching and learning activities (TLA), and assessment.

An example of constructive alignment of a course LO (DMGT) with corresponding TLAs, formative and summative assessment is shown in Table 1.

This alignment is crucial for ensuring the validity of the assessment program. Additionally, we utilised multiple criteria decision-making (MCDM) for prioritising the LOs and employed LA to offer insights to develop valid assessment programs. In both courses, innovative teaching methods (e.g., flipped classroom, PBL and inquiry-based learning) are used. For example, short videos or other materials are provided to students, they are supposed to investigate or practise a bit, and then during lectures, after further explanations, students are given quizzes that provide valuable feedback to both students and teachers.

To illustrate the alignment of LOs with the student workload and assessment, we used design analytics from the BDP LD tool for the DMGT course (Figure 1). While no perfect correlation exists among the elements, proper alignment and any deviations might be explained. For example, students' pre-knowledge in mathematics may influence the student workload, irrespective of the weight assigned to particular LOs. Additionally, the design of a 'perfect' assessment can be jeopardised by resource constraints and assessment costs. Moreover, the COVID-19 pandemic posed challenges, leading to the inability to conduct some assessment tasks (Divjak et al., 2023).

LA can also determine which types of TLAs were used to support students in achieving particular LOs. Six TLA types—acquisition, discussion, investigation, practice, production and assessment (Rienties et al., 2023)—were employed in the BDP tool, with TLA related to the level of LOs according to Bloom's taxonomy, as illustrated in Figure 2. Additionally, the student workload correlated with the weight of LOs, while the assessment tasks were aligned with the LO levels and contexts (Divjak et al., 2023).

3.2 | Data collection

We collected the M1 and DMGT assessment results and log data available in the LMS to analyse and compare the predictive power of formative assessment results (quizzes, homework) and other student activity (video and e-textbook consumption, class attendance), as

TABLE 1 Constructive alignment (Discrete Mathematics with Graph Theory example).

| Learning outcome	Teaching and learning activities	Formative assessment	Summative assessment
Identify the structure and type of proofs in mathematics	Flipped classroom: Students individually recap mathematical logic. The role of proofs in mathematics is discussed in class, and the fundamental types of proofs are analysed on examples	Quiz: Students classify given proofs by types and prove simple statements from number theory	Exam: Students analyse and identify the steps and find errors in the proofs of standard mathematical propositions
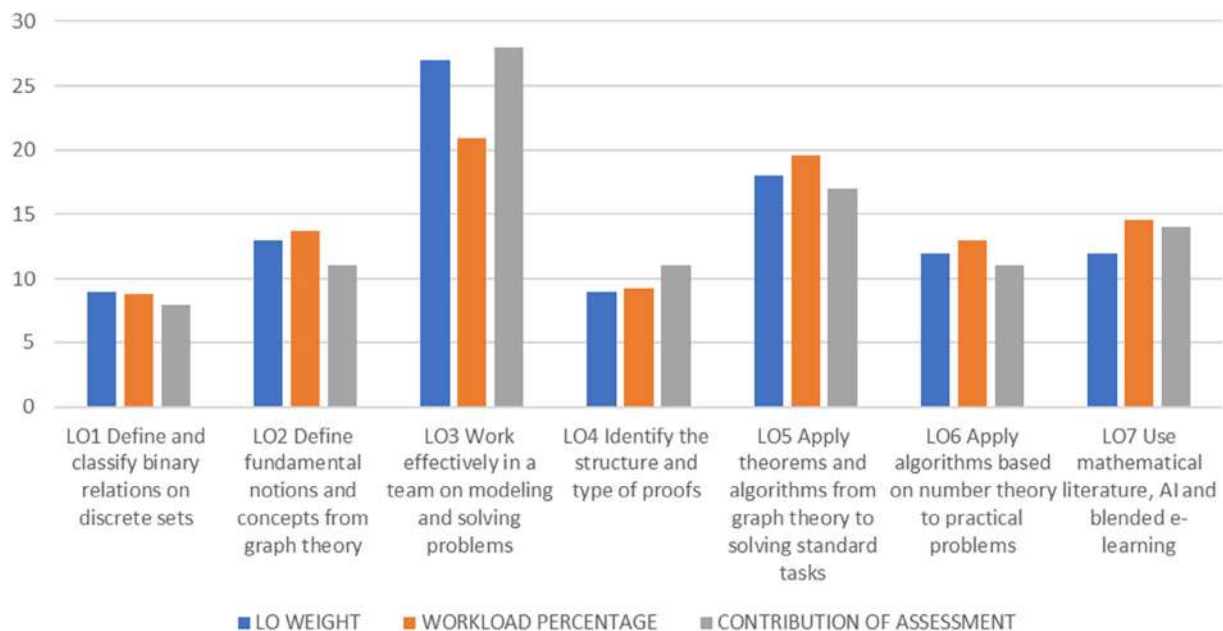
FIGURE 1 Design analytics for the Discrete Mathematics with Graph Theory course. LO, learning outcome.

FIGURE 2 Type of activities per learning outcomes (LOs) in the Discrete Mathematics with Graph Theory course.
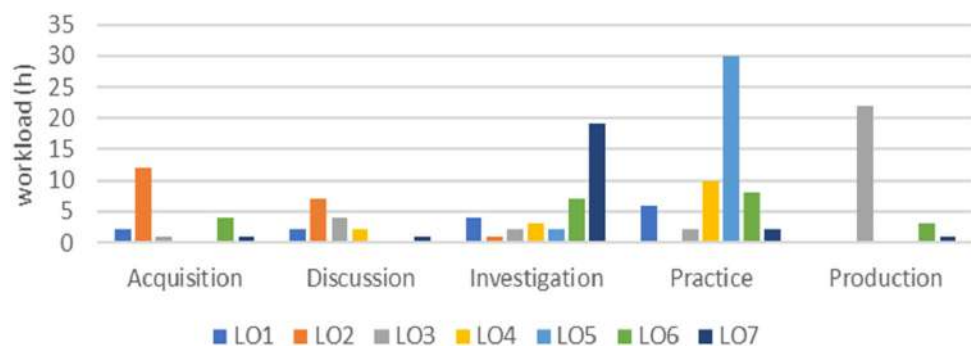

TABLE 2 Model inputs, outputs and hyperparameters.

Course	Input data		Outputs		Hyperparameters
	Assessment results	Log data	Binary model	Quaternary model	
M1	Quizzes Homework Exams 1, 2 and 3	Video E-textbook consumption Class attendance	Class 1: $s \leq 50\%$ Class 2: $s > 50\%$ of points in exam i $i = 1, 2$ DMTG $i = 1, 2, 3$ M1	Class 1: $s \leq 25\%$ Class 2: $25\% < s \leq 50\%$ Class 3: $50\% < s \leq 75\%$ Class 4: $s > 75\%$ of points in exam i	mtry: {2,3} or {2,3,4}[a] trees: random integer between 400 and 1500 min_n: random integer between 2 and 40
DMGT	Quizzes Exams 1 and 2	Video E-textbook consumption Class attendance			

[a]Only M1, when including exams 1 and 2 as predictors for Exam 3.

demonstrated in Table 2. The study included data from two academic years, 2021/2022 and 2022/2023, related to full-time students. In total, the data included 614 M1 students and 199 DMGT students (813 students in total).

Additionally, in order to gather student perspectives on, inter alia, the importance of formative assessment, as well as the difficulty of specific LOs, we also administered a survey to students via the LMS in the two academic years for both courses. Participation was voluntary,

and the questionnaire comprised both quantitative and open-ended questions. The questions varied slightly between years and student participation in the survey had a response rate between 42% and 50%.

3.3 | Data analysis

The assessment data used in this study are related to two mathematical courses in which due attention is paid to the validity and reliability of assessment. Assessment validity has been ensured based on the alignment of assessment with the intended LOs, which have been prioritised considering their relative importance (Divjak et al., 2023). To ensure reliability and other assessment characteristics, we analysed the exam structure based on the statistics provided by the Moodle LMS. Besides the internal consistency of the exams, this also includes statistics related to each of the items, primarily the facility index (the average mark related to each item), the effective weight (versus the intended weight), the discrimination index (correlation between the score for a specific item and the score for the exam) and the discriminative efficiency (how good the discrimination index is relative to the difficulty of the item).

The assessment data were analysed using the RF algorithm in R. The RF is a machine learning algorithm used for classification and regression purposes, based on building multiple decision trees and combining their results for final classification. When building decision trees, different subsets of attributes are randomly selected. Because of the Law of Large Numbers, there is no overfitting. As such, RFs are considered to be an effective tool in making predictions (Breiman, 2001; Kursa & Rudnicki, 2010). Importantly, RFs are popular due to the possibility of variable importance measures (VIMs), with the impurity importance and the permutation importance being the most widely used VIMs (Nembrini et al., 2018).

In our analysis, we first performed cross-validation to choose, among 1000 randomly chosen combinations of hyperparameters (as in Kabathova & Drlik, 2021), the appropriate combinations to make the learning algorithm as good as possible. After that, we performed the RF on the training dataset, and tested the efficiency of the algorithm on the testing dataset. For each of the courses, the data from both academic years were merged, and then randomly distributed, with 75% used for training, and 25% for testing.

The analysis was performed for two models. Basically, for each of the two courses, students were divided into classes based on their summative assessment results (periodical exams): two classes (binary model—B) and four classes (quaternary model—Q). Students were divided into classes depending on the number of points they achieved in each periodical exam—with the total number of points divided by two (50%) in the binary (B) model, and with the total number of points divided into quartiles (25%) in the quaternary (Q) model. The details regarding model inputs, outputs and hyperparameters are shown in Table 2.

The efficiency of each model was assessed based on several metrics, including the area under the Receiver Operating Characteristics curve (ROC_AUC). The ROC curve is a plot of the true positive rate (sensitivity) versus the false positive rate (1-specificity), at different threshold settings. As a graphical representation, it is used to assess the performance of binary classification models, with several approaches to extending the ROC curve to multi-class classification tasks. ROC_AUC is the area under the curve (Hand & Till, 2001; Mandrekar, 2010).

Moreover, confusion matrices were analysed for each model. Based on that, in this article, we present the Q model as the more efficient one but also more important for practical use, as described under Results.

We analysed the importance of predictors based on the mentioned VIMs, that is using the Gini index and using permutations. We used the Boruta extension for finding the initial set of predictors. Boruta adds more randomness (Kursa & Rudnicki, 2010), by creating, for each variable, a 'shadow variable' with permuted values, and training the RF on this extended dataset, containing both original and 'shadow variables'.

For the DMGT course, the analysed predictors of success in Exam 1 included quizzes, videos, e-textbooks and class attendance based on Boruta algorithm's results. For Exam 2, besides the mentioned predictors, a version of analysis was performed to test the predictive power of success in Exam 1 as well. As for the M1 course, the analysed predictors of success in Exam 1 included quizzes, homework, videos, e-textbooks and class attendance. For exams 2 and 3, besides the mentioned predictors, a version of the analysis was performed to test the predictive power of success in the previous periodical exams as well. Additionally, for the DMGT course, the importance of predictors of success in the acquisition of each LO was analysed as well.

Furthermore, log data from the Moodle LMS were analysed with respect to the four student classes, to provide a better understanding of how their activity relates to their summative assessment results. This included data on repeated access to formative assessments, access to e-textbooks and videos, as well as class attendance.

Student survey data were analysed with descriptive statistics.

4 | RESULTS

The Cronbach's alpha index showed that summative assessments included in the study were acceptable in terms of internal consistency, with values in the range from 68% to 78%. The validity of the assessments was checked based on their alignment with the intended LOs of the course (Divjak et al., 2023).

4.1 | Efficiency of the model

The efficiency of the models (both B and Q) was assessed based on, inter alia, the ROC_AUC metric. For each of the two courses, for both the training and testing dataset, the ROC_AUC values are presented in Tables 3 and 4 below.

According to literature, ROC_AUC values above 0.7 are considered acceptable, with values above 0.8 excellent and above 0.9 outstanding (Hosmer & Lemeshow, 2000). Taking this into consideration,

TABLE 3 DMGT—efficiency of the models based on the ROC_AUC metric.

Exam 1				Exam 2 (excluding Exam 1 as a predictor)				Exam 2 (including Exam 1 as a predictor)			
Be1		Qe1		Be2−		Qe2−		Be2+		Qe2+	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
0.924	0.623	0.900	0.652	0.939	0.415	0.867	0.666	0.972	0.643	0.894	0.694

it was found that the high ROC_AUC values related to the training set (generally around or above 0.9) indicate excellent discriminatory power, and show that the model has learned the patterns well. There is no underfitting, that is a high rate of bias. When the ROC_AUC values related to the testing set are considered, the efficiency of the models was generally found to be acceptable (with values around or above 0.7), except for the DMGT Be2-model. In this article, we describe in detail and discuss further the Q models, as there are better ROC_AUC values related to testing in the Q models, but also because they provide the level of granularity which enables more thorough insights and interpretation.

The efficiency of the models was also analysed based on confusion matrices, with the ones related to the models demonstrating the highest efficiency according to the ROC_AUC presented below.

In relation to the M1 course, it was found that, concerning exams 1 and 2, the (Q) model performed well in predicting the students in the middle quartiles, that is in classes 2 and 3. For example, regarding Exam 2 (Figure 3), the model Qe2+ correctly predicted 27 class 2 students and 50 class 3 students, although it also misclassified 22 students actually belonging to class 2 as class 3. The model performed less well in predicting the lowest and the highest quartiles, struggling to accurately predict the students in classes 1 and 4. For example, regarding Exam 2, it underpredicted the number of students in class 4, correctly detecting 10 students, but classifying additional 11 as the neighbouring class 3.

The results were different when it came to Exam 3 (Figure 4). Here, the model performed well in predicting the lower quartiles, or lower performing students. For example, the model Qe3-correctly predicted 42 students in class 1 and 47 students in class 2. It also misclassified additional 24 class 1 students as class 2 and 22 class 2 students as class 1.

As for the DMGT course, the confusion matrices suggested that, for Exam 1, the model performed well in predicting the middle quartiles, in particular class 3. For Exam 2, it also performed well in predicting class 3, but with class 4 coming in second, suggesting better performance in classifying the better performing students.

4.2 | The importance of predictors

4.2.1 | Discrete mathematics with graph theory

For Exam 1, class attendance came out as the least important predictor based on the Gini index, permutations and the Boruta extension equally. As for other predictors, the three VIMs gave different results,

putting the predictors in the following order (from the most important): quizzes, e-textbook, videos (Gini); e-textbook, quizzes, videos (permutation); e-textbook, videos, quizzes (Boruta).

For Exam 2, the results were more uniform, with the Gini index and the Boruta pointing to the following order (from the most important): first periodical exam, quizzes, e-textbooks, class attendance and videos. With permutation, quizzes came out as second least important.

The analysis of predictors (B model) of success in the acquisition of LOs (those that are covered by both formative and summative assessment) generally pointed to the importance of quizzes, followed by e-textbooks and class attendance. Videos were again not found to be an important predictor. It should be noted that the ROC_AUC values were not acceptable in relation to all the tested LOs. However, for the LOs for which they were acceptable, and which turned out to be at the *Apply* level (LO1, LO5) of Bloom's taxonomy (Anderson & Krathwohl, 2001), quizzes came in the first place looking at two of the three considered indices (whether Gini, permutations or Boruta), followed by e-textbooks or class attendance.

4.2.2 | Mathematics 1

For Exam 1, quizzes and homework were found to be the most important predictors according to all the three measures. As for the other three predictors, the order was the following (from the most important): e-textbook (reading material), videos, class attendance (Gini); video, class attendance, e-textbook (permutation); class attendance, video, e-textbook (Boruta).

For Exam 2, the success in Exam 1 was found to be the most important predictor according to all the three measures. Quizzes, e-textbook and homework were moderately important (though in different order considering the three measures), while videos and class attendance (in varying order) were the least important.

For Exam 3, the success in exams 2 and 1 (respectively) was found to be the most important predictor based on all the three measures, followed by quizzes. E-textbook, class attendance and homework (in varying order) were moderately important, while videos came out as the least important.

4.3 | Contextual factors affecting predictions

In addition to identifying the predictors of success in summative assessments (exams), we examined the distribution of students

TABLE 4 M1—efficiency of the models based on the ROC_AUC metric.

Exam 1				Exam 2 (excluding Exam 1 as a predictor)				Exam 2 (including Exam 1 as a predictor)				Exam 3 (excluding Exams 1 and 2 as predictors)				Exam 3 (including Exams 1 and 2 as predictors)			
Be1		Qe1		Be2−		Qe2−		Be2+		Qe2+		Be3−		Qe3−		Be3+		Qe3+	
Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test	Train	Test
0.901	0.730	0.914	0.710	0.910	0.752	0.925	0.748	0.972	0.643	0.956	0.821	0.942	0.602	0.883	0.753	0.971	0.670	0.921	0.701

across classes in two consecutive academic years encompassed by this study. We assessed the reliability of summative assessments using Cronbach's alpha in the Moodle LMS and analysed students' activity within the Moodle LMS through log data for each activity, as well as fluctuations in activity throughout the semester. It is important to note that our analysis is limited to capturing activities within the LMS and does not include potential student activities outside of it.

Our findings reveal similar distribution patterns across four classes over the two academic years, as illustrated by the M1 Exam 1 and DMGT Exam 2 graphs in Figures 5 and 6. This uniformity supports the aggregation of data from the 2 years into a single sample. Furthermore, the results indicate that, as anticipated, students in the graduate-level DMGT course exhibit higher success rates compared with the freshmen in the undergraduate-level M1 course. Notably, a pattern emerges in M1, where success on Exam 3 lags significantly behind that on exams 1 and 2 (with similar result patterns). This transition towards lower scores is also depicted in the braided graph in Figure 7. In contrast, students perform better on Exam 2 (graph theory) than on Exam 1 (discrete mathematics and number theory) in DMGT, as shown in Figure 8.

Students' online activities are concentrated around the weeks of the periodical exams (three for M1 in weeks 6, 11 and 16, and two for DMGT in weeks 6 and 12), as can be seen in Figures 9 and 10. Better-performing students evidently invested more time in (online) learning. It is worth mentioning that we categorised classes based on Exam 1 outcomes.

Examining Figures 11 and 12, we observe that in both courses, the most frequently consulted resources were previously taken formative assignments (quizzes for M1 and DMGT, and homework for M1). Intriguingly, more successful students engaged with more reading materials (e-textbooks with exercises, additional PDFs, etc.) than their less successful peers.

4.4 | Student perspective

In both the 2021/2022 and 2022/2023 academic years, students of the M1 and DMGT courses had the chance to provide feedback after Exam 1. For the 2021/2022 academic year in the M1 course, 127 out of the 305 students who took Exam 1 responded to the survey, resulting in a 42% response rate. For the DMGT course, 50 out of 103 students who took the exam completed the survey, giving a 49% response rate. In the 2022/2023 academic year for M1, 150 out of 320 students who took Exam 1 answered the questionnaire, equating to a 47% response rate. For the DMGT course in the same year, 51 out of 103 students who took Exam 1 provided feedback, representing a 50% response rate.

In both years and across both courses, content related to mathematical proofs and rigours mathematical reasoning was identified as the most challenging.

Feedback from M1 students in the 2021/2022 academic year revealed that the majority recognised the connection between formative and summative assessments. They also indicated that formative
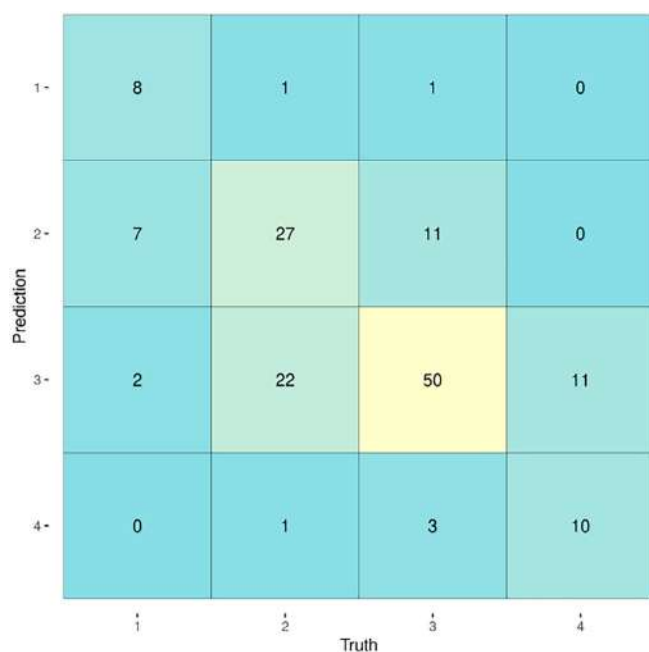
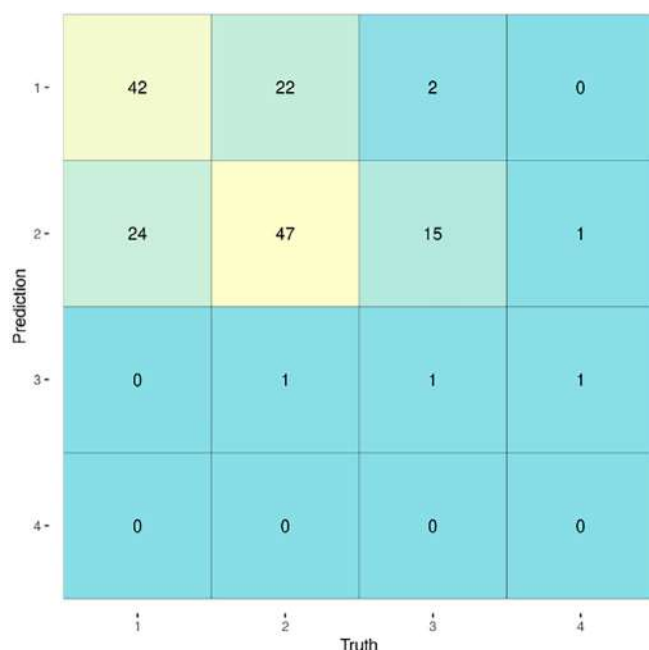**FIGURE 3** M1 Qe2+ confusion matrix.

FIGURE 4 M1 Qe3− confusion matrix.

assessment was beneficial for their exam preparation (as seen in Figure 13).

In the subsequent year, students from both courses were asked to evaluate the helpfulness of formative assessment in preparing for the summative exam. A significant majority from both courses reported that formative assessment was useful in at least one aspect, whether content-related (align with LOs) or technical (such as familiarity with e-exams), if not both. As shown in Figure 14, only a minority (16% of respondents) felt that the formative assessment was not

helpful. Among first-year undergraduate students (M1), 53% found formative assessment beneficial in both the technical and content-related aspects, whereas graduate students appreciated the content-related assistance more.

5 | DISCUSSION

5.1 | RQ1: How can formative assessment results contribute to the prediction of summative assessment results?

The analysis of the importance of predictors showed that formative assessment is an important predictor of success in summative assessment, that is, in the acquisition of LOs. In particular, quizzes have been found to be an important predictor (in both courses), followed by homework (in M1) as a form of formative assessment. In general, other student activity, such as e-book consumption, class attendance and especially watching videos, were not found to be as important predictors of student success in summative assessment. The results of the analysis of predictors per LO in the DMGT course point to a similar conclusion, confirming the importance of formative assessment in predicting the acquisition of LOs. These results are in line with some previous findings (Bulut et al., 2023), showing that online formative assessment scores are the key predictor of students' course performance.

Our study also found that previous summative assessment (periodical exams) is an important predictor of success in the following summative assessment. This is evident in the context of both analysed courses: the DMGT course, where Exam 1 was found to be the best predictor of success in Exam 2, and the M1 course, where exams 1 and 2 were shown to be the best predictors of success in Exam 3 (according to all the three analysed measures—the Gini index, permutations and Boruta). In relation to the latter, it can be discussed that in M 1 the first two exams are the strongest predictors of success in Exam 3 because of students' learning strategies: if their success in the first two exams does not allow them to complete the course via periodical exams no matter how successful they are in Exam 3 (and they have to sit the full exam at the end of the semester anyway), they do not make the effort to pass Exam 3. If we look at the log data related to student activity around Exam 3, we can also see the lowest activity rate in relation to the other two exams. Furthermore, the fact that Exam 2 is a stronger predictor than Exam 1 can be explained in the context of LOs, as Exam 2 is more related to Exam 3 than Exam 1. This pattern is very different in DMGT, because after Exam 2, students work on a team project and do not have another exam as in M1. Furthermore, graph theory as the content covered by Exam 2 is much closer to practice (our students study IT) than mathematical proofs, relations and number theory covered by Exam 1. That was also confirmed in students' feedback in the questionnaire. M1 students also expressed difficulty with proofs, but since they are first year undergraduate students, the outcomes and consequently the performance expected from them is much less demanding than from the DMGT students.
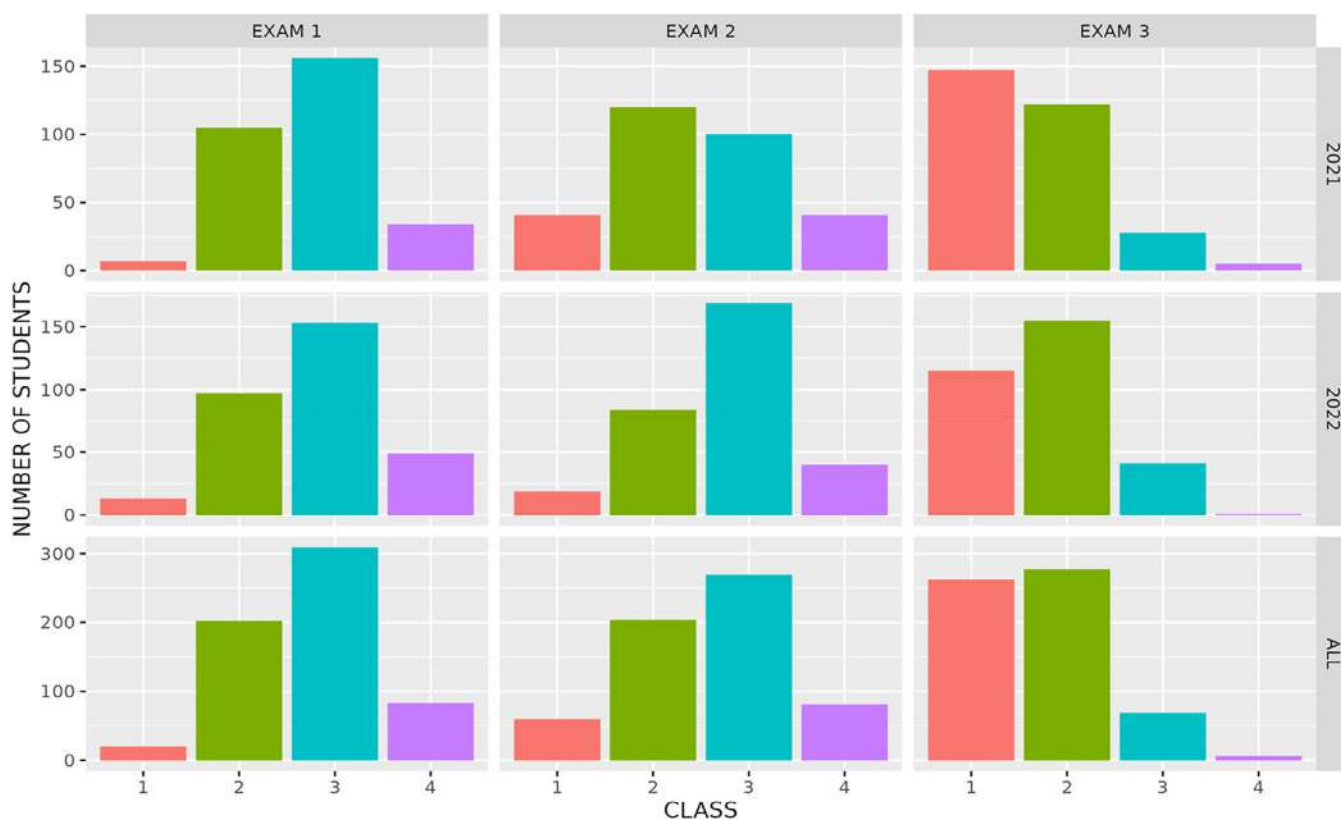
FIGURE 5   Distribution of Mathematics 1 students into classes.

Furthermore, our results also reveal some differences in the importance of predictors between the two courses, pointing to the role of context. For example, when comparing the predictive power of e-textbook consumption, the results show that their predictive power is generally higher in the context of the graduate course (DMGT). This may be related to the learning habits and styles of more senior students and their ability to learn independently. This comes as no surprise, as it has already been found that predictors of student achievement depend on contextual factors related to LD, subject area, and educational institutions (Saqr et al., 2022).

The essential premise in the background of this predictive model is that formative assessment needs to be aligned with summative assessment. This is related to the notion of constructive alignment (Biggs, 1999) of LOs, TLAs and assessment, as well as to the validity of assessment (Divjak et al., 2023).

In particular, to ensure the validity of assessment, it is essential to link assessment (both formative and summative) with the intended LOs. While doing so, it is crucial to note that not all LOs are equally important. Therefore, for the courses used in this study, this was done in line with the approach presented in Divjak et al. (2023): the assessment programs were not only based on LOs, but also reflected their prioritisation, which can be done by utilising various multi-criteria decision methods. Ensuring valid assessment is an important part of LD and its continuous improvement (Divjak et al., 2023).

In other words, for this predictive model to make sense, and to be able to use the predictive power of formative assessment, the precondition is to ensure that assessment is constructively aligned with the intended LOs, and that formative and summative assessment are mutually linked (Figure 15). In practical terms, that means that, for example, for a particular LO, there should be dedicated tasks in quizzes, as well as in the relevant periodical exams. At the same time, students' performance in summative assessment needs to be considered in the quality assurance and consequent revision of LOs.

Such coherence should be ensured while developing and improving LD, and can be supported by LD analytics, like those provided by the BDP LD tool (Divjak, Grabar, et al., 2022).

Student feedback (questionnaire), in which nearly half of the students across both courses and years actively participated, indicating a strong level of student engagement in feedback mechanisms, the majority of students recognised the value of formative assessments in preparing for their summative exams, underlining the significance of these assessments in curriculum design. It is obvious that students seemed to benefit from both the content and the technical guidance provided by the formative assessments. Additionally, there were evident differences in preferences between undergraduate and graduate students. Undergraduates appeared to value the technical aspects of formative assessments more, while graduates leaned towards content-related assistance. This divergence can be because graduate students have already reached the technical level needed for e-exams. Lastly, while the majority found formative assessments beneficial, the fact that a minority did not find them helpful points to potential areas for refinement in future curriculum iterations.
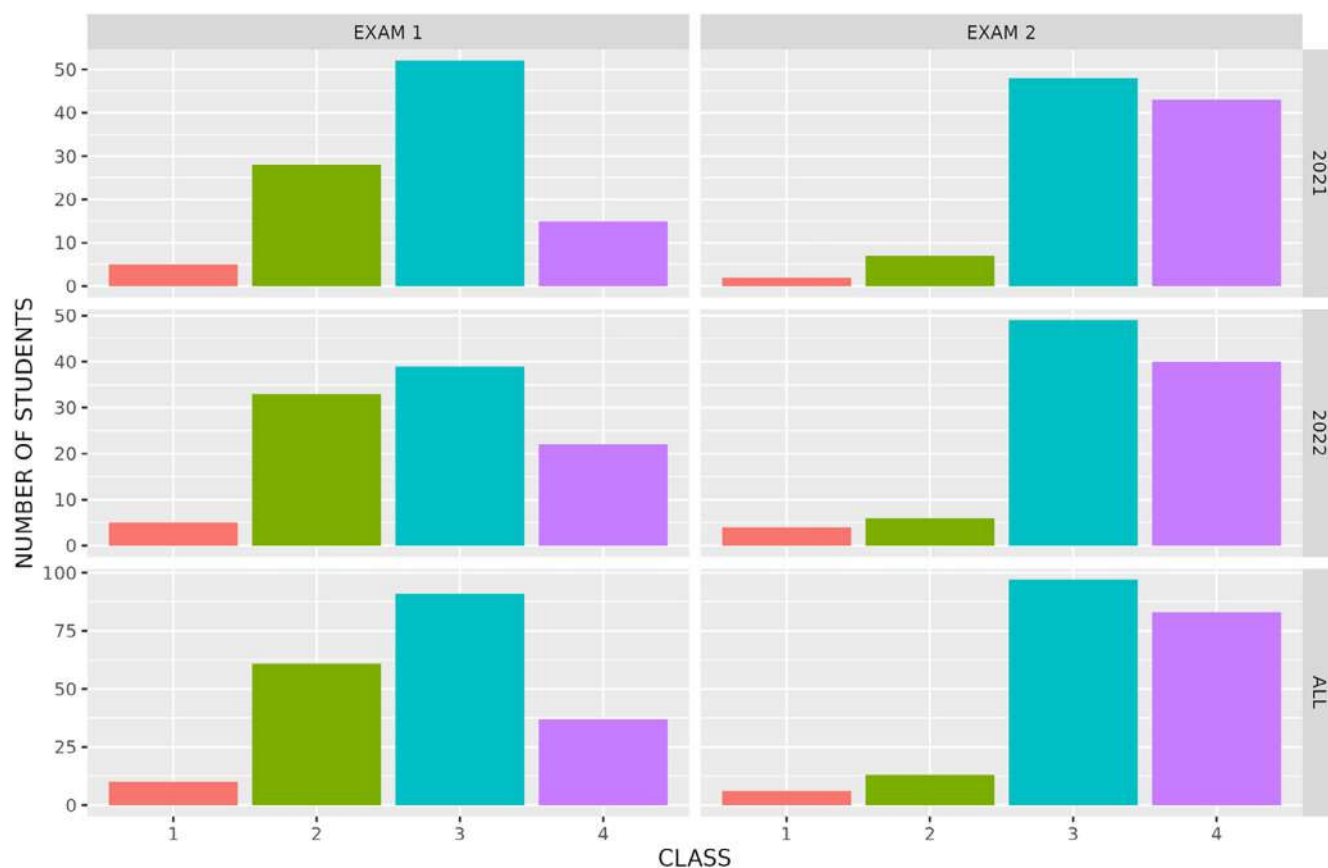
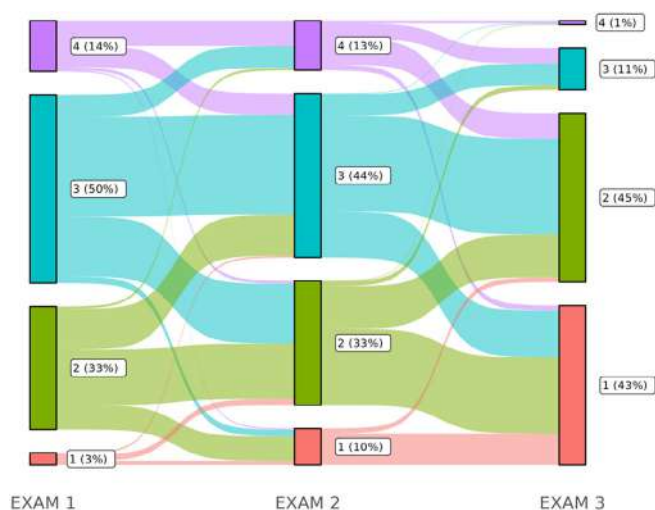**FIGURE 6**   Distribution of Discrete Mathematics with Graph Theory students into classes.

FIGURE 7   Mathematics 1 students' transitions between classes.

FIGURE 8   Discrete Mathematics with Graph Theory students' transitions between classes.

5.2 | RQ2: In what ways can other contextual factors (e.g., levels of learning outcomes, levels of study and student activity) affect predictions?

During curriculum implementation, data on student activity in their online learning environment (log data) can be used as the basis for LA. So, while discussing the predictive power of various types of student activity in the LMS, it is also worth considering the intensity with which students took part in these activities.
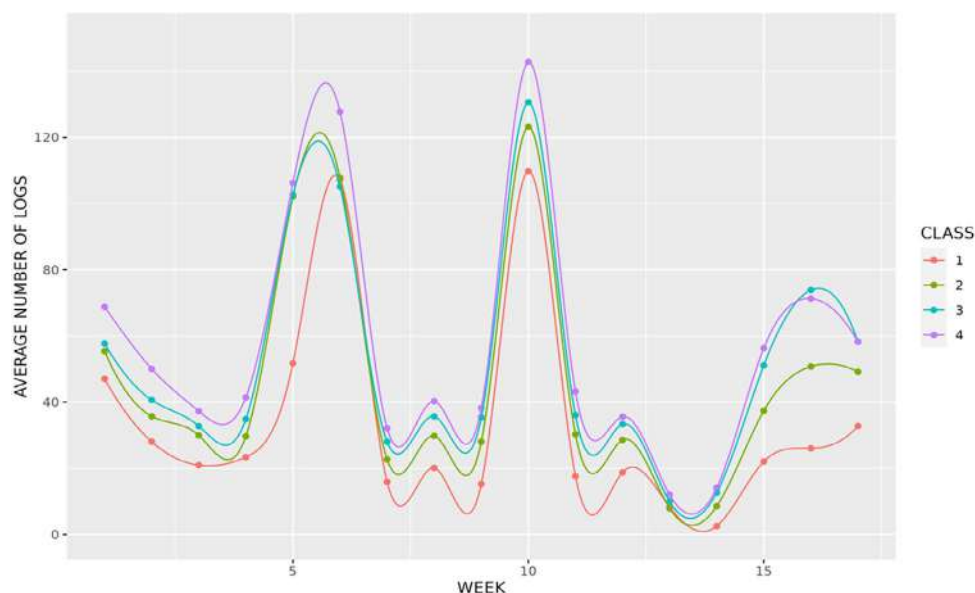
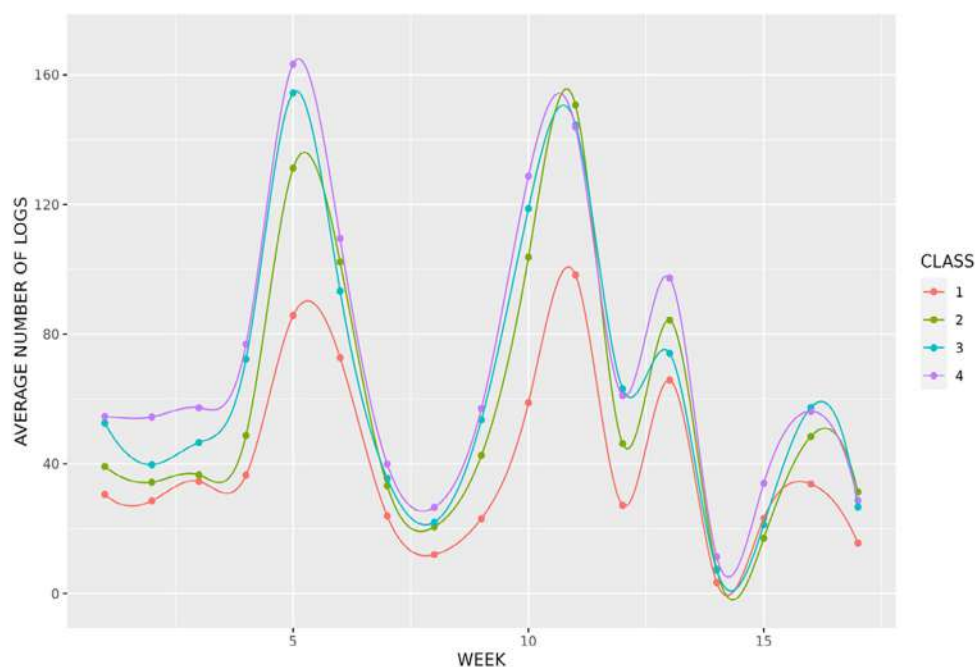**FIGURE 9** M1 students' activity per week.

FIGURE 10 Discrete mathematics with graph theory students' activity per week.

To do so, we analysed the number of students' logs into e-textbooks, formative assessment quizzes and videos, as previously described. These results may shed some more light on *why* the order of predictors looks the way it does. Namely, in the DMGT course, regardless of the student class (based on their success in Exam 1), re-visiting formative assessment (quizzes) around the time of summative assessment (Exam 1) was generally found to be more intense than using e-textbooks, and especially than watching videos. This reflects the importance of predictors of success in Exam 1 according to the Gini index (though the other two measures point to different importance values). Similarly, in the context of the M1 course, the most intense activity around periodical exams was related to re-visiting formative assessment quizzes, followed by e-textbooks, and, lastly, for-mative assessment quizzes, followed by e-textbooks, and, lastly,

videos, roughly reflecting the results related to the importance of pre-dictors. As videos turn out to be the least popular resource, it is also worth considering why they come at the end of the spectrum when it comes to the importance of predictors. As for why such materials are less popular, this phenomenon could be attributed to the fact that many videos are recordings of live classes or their summaries, and most students attended classes in person and took notes. This is because the courses were delivered in a hybrid mode, with the major-ity of students participating face-to-face and engaged in TLAs, and videos are a passive learning resource. There was less need to watch videos at home and videos were most often used as recapitulation material immediately before exams. Additionally, students may not have been accustomed to exploring other types of videos available in
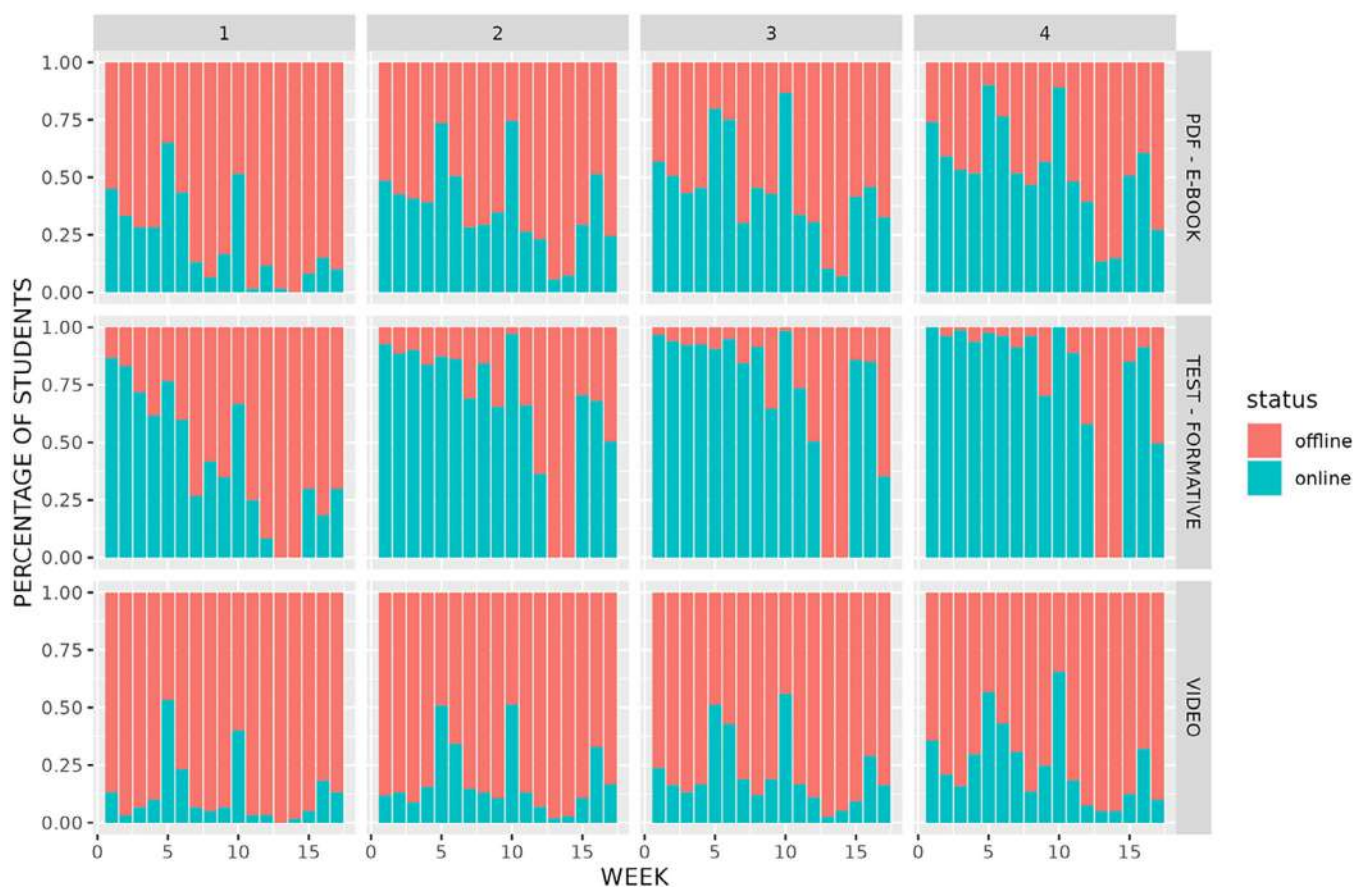
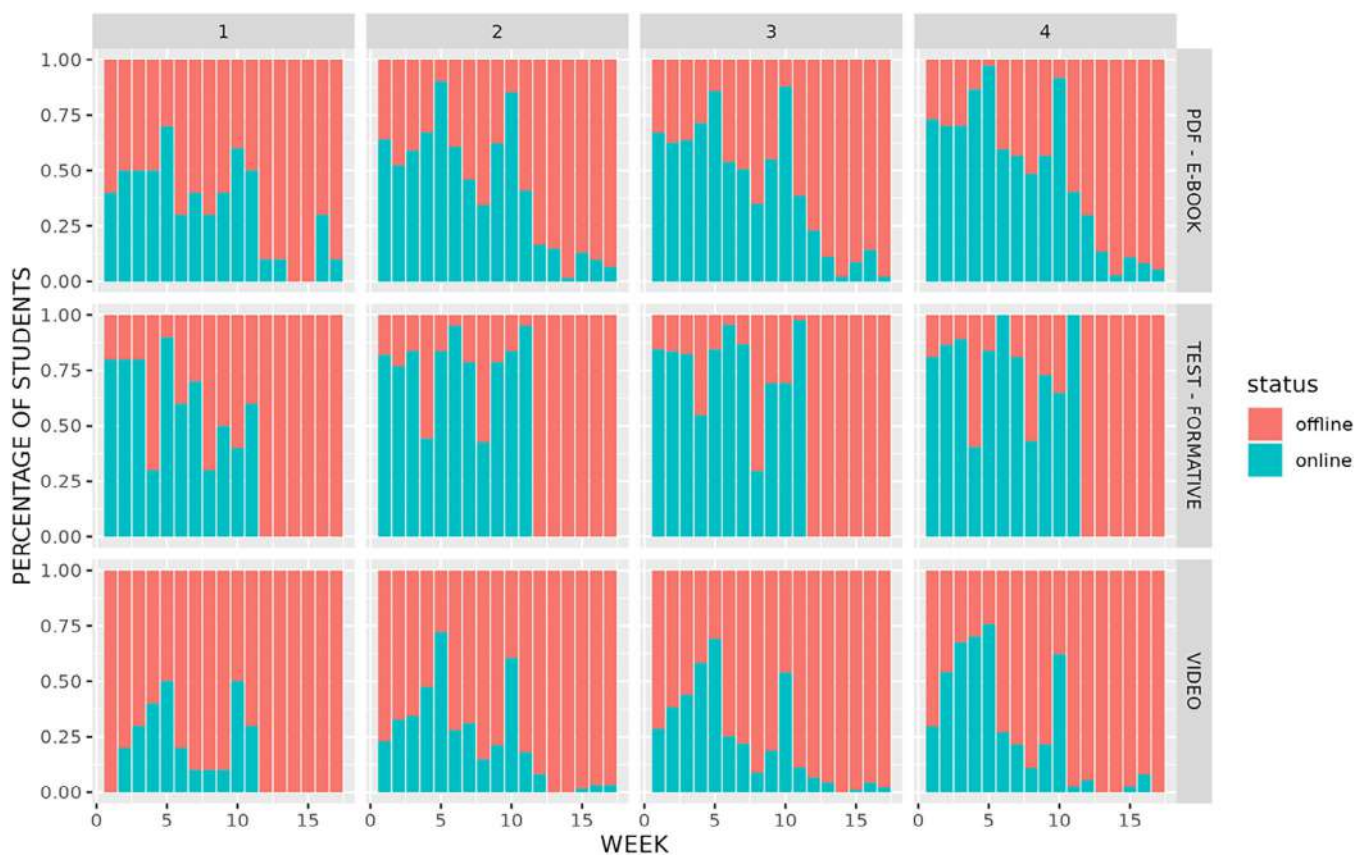**FIGURE 11**   M1 students' weekly activity per resource/predictor.

FIGURE 12   Discrete mathematics with graph theory students' weekly activity per resource/predictor.
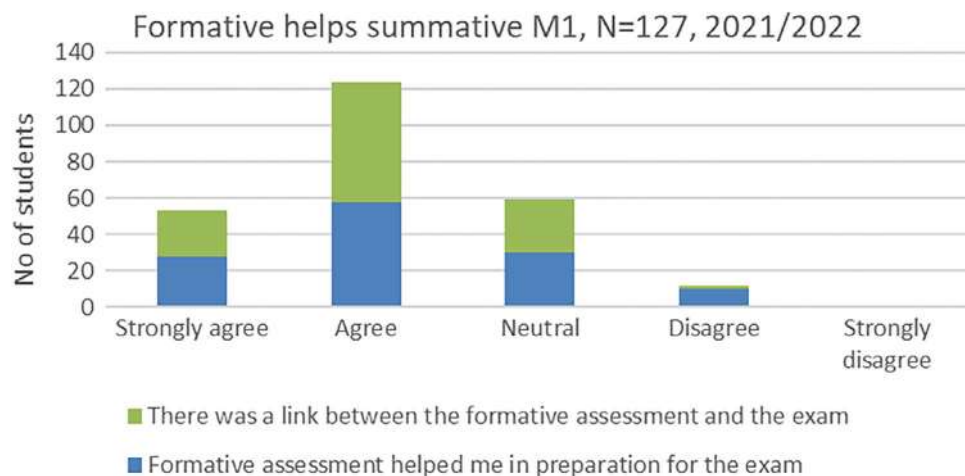
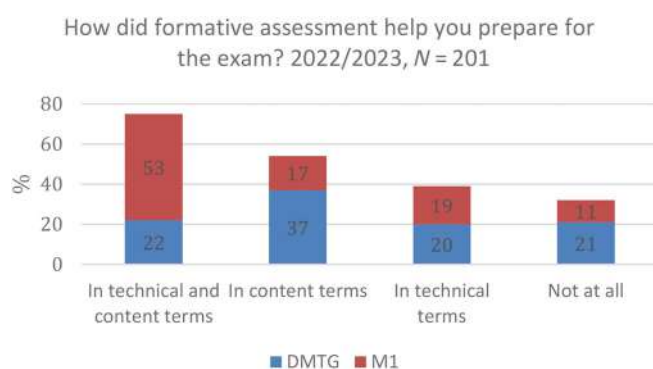**FIGURE 13** Links between formative and summative assessment.

FIGURE 14 Formative assessment supports summative—student perspective.
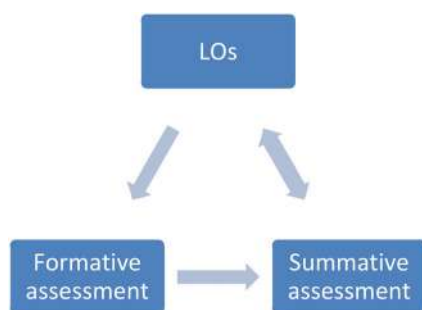

FIGURE 15 Linking learning outcomes (LOs) with formative and summative assessment.

the LMS intended to prepare them for classes, particularly within the context of the flipped classroom approach.

Besides some difference in the importance of predictors, M1 and DMGT students also showed different patterns of progression, pointing to a difference between undergraduate (M1) and graduate (DMGT) students. But generally, students who performed well on Exam 1 maintained their standing in subsequent exams, indicating that success in Exam 1 can be a strong predictor for future performance

(if performance in previous exams is analysed as a predictor). However, in our case, M1 students showed a decrease in success towards the end of the semester (Exam 3, probably for previously discussed reasons), while DMGT students showed an increase in success (with a substantial portion of students shifting from class 3 to class 4). Students who did not perform well in Exam 1, belonging to the lower quartiles, demonstrated some improvement over time, with DMTG students from the second quartile showing notable potential for improvement. These variations in outcomes may be tied to students' activity levels during the course and their individual learning approaches, strategies and self-regulation.

This analysis might have a particularly important influence on LD. Generally, insights into what types of student activity in the LMS has the highest frequency (based on log data) might support educators and learning designers in rethinking and possibly improving course LD in a way that better responds to the needs of their students. In particular, the analysis provides valuable insights when it comes to assessment planning, enabling educators to tailor support for students based on their results in Exam 1. Moreover, the results of this analysis provide valuable insights as grounds for feedback to students (Banihashem et al., 2022), which can support them in the development of effective learning practices. Furthermore, the results highlight the importance of encouraging continuity and persistence in learning throughout the course.

In general, student perspective, based on the survey in both courses, but also on previous analysis (Divjak, Žugec, et al., 2022) indicated a consistent observation that LOs related to mathematical proofs and abstract mathematical reasoning were continuously challenging, hinting at a potential area where enhanced instructional support might be beneficial.

Due to the number of students involved in the analysis as well as the number of course LOs, we came to some predictors on the level of LOs only for LOs on the *Apply* level, showing that regular practice (represented here by quizzes) supports students to better perform related to these LOs, which does not come as a surprise. But, it is interesting to note that we tried to use the flipped classroom approach and ask students to look at some materials before classes

and this result indicates that the combination of acquisition and practice activities, followed by formative assessment, is beneficial for students.

5.3 | RQ3: What elements can affect the efficiency of prediction?

The validity of assessment is a prerequisite for predicting assessment results, providing recommendations for continuous improvement in LD, and supporting students' self-regulated learning, as well as making meaningful predictions (Divjak et al., 2023).

When discussing the performance of a predictive model, our study points to several issues.

Our results suggest, based on the ROC_AUC values, that all the models—whether binary or quaternary, whether including previous summative assessments or not—perform excellent on the training data, but with a drop in performance on the test data. This suggests a need to collect additional data, to provide the model with more learning material. Namely, the models generally do best in classifying the middle quartiles, with some difficulty in classifying the lowest and the highest performing students, suggesting difficulty in distinguishing between extremes of performance. This is related to class imbalance, so more data, with an emphasis on the lowest and highest performing classes, could enable the models to learn the characteristics of these students better, leading to better predictions. At the same time, due to the fact that very often we have normal (Gauss) distribution of performance data, the lower number of students in extremes is expected. Nevertheless, for higher education, it is important to collect and analyse their own data and not just accept conclusions based on MOOC data (very large samples) that are more reliable but not always relevant for higher education. Another approach would be to collect data several years in a row (if teaching and assessment models are similar) and aggregate them. Unfortunately, data collected during the peak of the pandemic (online teaching and learning) differ considerably from regular delivery (mainly blended or hybrid).

Moreover, it should be noted that the efficiency of prediction is related to the way students' summative assessment results are categorised. In the context of our study, this is related to the differences between the B and the Q model. As demonstrated in Tables 3 and 4 above, if we look at both analysed courses, the ROC_AUC value on the testing dataset is most often higher for the Q than the B model. In relation to the DMGT course, this applies to all summative assessments (Qe1, Qe2+ and Qe2−), while in relation to the M1 course, it applies to the majority (Qe2+, Qe3− and Qe3+). This points to the importance of considering (and testing) various possible categorizations in order to receive the best possible results. Even though binary classification is more researched and applied, it is not fine enough to support changes in LD.

Our study pointed to the importance of completeness of data. In particular, this is related to the M1 course, where a problem with missing data appeared. Namely, for Exam 2, different student study groups had different tasks, so the data were not comprehensive.

Moreover, in relation to the DMGT course, where the predictors were also analysed in relation to separate LOs, the analysis could not give comprehensive results as not all LOs were covered by formative assessment. Furthermore, we have a problem because we can collect only digital data about student engagement, not capturing the data from the physical environment, even though we are aware that in f2f learning there are other important incentives or obstacles for learning, or other online learning platforms (Araos et al., 2023). Finally, students at risk, besides predictors related to teaching and learning, also have strong influence coming from socio-economic status and personal challenges (Merritt & Buboltz, 2015).

5.4 | Limitations

The main limitation of our study is that it encompasses a limited number of courses (two maths courses), in a specific field of study (informatics). Widening the research to additional courses, in different study fields, and also educational contexts, may elicit different results, especially in relation to the importance of various predictors, as students may be more inclined to use different types of resources. Moreover, when it comes to the analysis of log data, it should be noted that our analysis included only student activity in the LMS, whereas potential student activity outside the LMS was not part of the analysis. Furthermore, we have unbalanced data (low numbers) about the lowest and the highest performing students and have not taken into consideration the factors outside of the LMS that we know that can influence the student-at-risk category.

6 | CONCLUSION

Assessment is part of a learning process and cannot be concentrated only on one or a few assessment points in time. Formative assessment needs to be continuous, aligned with intended learning outcomes (LOs) and summative assessment. Formative assessment should also be valid if we want it to contribute to the achievement of intended LOs, as well as if we want to use it as a predictor of student success. To make predictions of student performance based on formative assessment results, it is essential that formative assessment is constructively aligned with LOs and summative assessment.

Planned LD needs to be related to the curriculum implementation via combining LA coming from LD (design analytics) and LA capturing student traces in the learning environment. Prominent patterns of student engagement with learning material, learning environment and peers can be discovered via different machine learning methods, and here, we used the Random Forest classification algorithm to develop a predictive LA model. The model is based on the data from two university mathematical courses, delivered at different years and levels of study and incorporating 813 students in two consecutive years.

The analysis of the importance of predictors showed that formative assessment, together with previous summative assessment if

available, is an important predictor of success in summative assessment, that is, in the acquisition of LOs. A vast majority of students found formative assessments beneficial in preparation for summative exams. For a predictive model to make sense, and to be able to use the predictive power of formative assessment, the precondition is to ensure that assessment is constructively aligned with intended LOs, and formative and summative assessment are mutually linked.

Finally, successful students use a variety of resources to learn and prefer e-text books to videos.

AUTHOR CONTRIBUTIONS

Barbi Svetec: Writing – original draft; writing – review and editing; conceptualization; investigation. **Blaženka Divjak:** Conceptualization; investigation; funding acquisition; writing – original draft; methodology; writing – review and editing; supervision; validation. **Damir Horvat:** Visualization; formal analysis; data curation.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

ORCID

Blaženka Divjak https://orcid.org/0000-0003-0649-3267
Barbi Svetec https://orcid.org/0000-0003-1983-2467
Damir Horvat https://orcid.org/0000-0002-4215-9779

REFERENCES

American Educational Research Association, American Psychological Association, and the National Council on Measurement in Education. (2014). *Standards for educational & psychological testing.* https://www.aera.net/Publications/Books/Standards-for-Educational-Psychological-Testing-2014-Edition

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives (Complete ed).* Longman.

Araos, A., Damşa, C., & Gašević, D. (2023). Browsing to learn: How computer and software engineering students use online platforms in learning activities. *Journal of Computer Assisted Learning, 39*(2), 676–693. https://doi.org/10.1111/jcal.12774

Banihashem, S. K., Noroozi, O., van Ginkel, S., Macfadyen, L. P., & Biemans, H. J. A. (2022). A systematic review of the role of learning analytics in enhancing feedback practices in higher education. *Educational Research Review, 37,* 100489. https://doi.org/10.1016/j.edurev.2022.100489

Biggs, J. (1999). What the student does: Teaching for enhanced learning. *Higher Education Research & Development, 18*(1), 57–75. https://doi.org/10.1080/0729436990180105

Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Bulut, O., Gorgun, G., Yildirim-Erbasli, S. N., Wongvorachan, T., Daniels, L. M., Gao, Y., Lai, K. W., & Shin, J. (2023). Standing on the shoulders of giants: Online formative assessments as the foundation for predictive learning analytics models. *British Journal of Educational Technology, 54*(1), 19–39. https://doi.org/10.1111/bjet.13276

Divjak, B., Grabar, D., Vondra, P., & Svetec, B. (2022). Balanced learning design planning: Concept and tool. *Journal of Information and Organizational Sciences, 46*(2), 361–375. https://doi.org/10.31341/jios.46.2.6

Divjak, B., Svetec, B., Horvat, D., & Kadoić, N. (2023). Assessment validity and learning analytics as prerequisites for ensuring student-centred learning design. *British Journal of Educational Technology, 54*(1), 313–334. https://doi.org/10.1111/bjet.13290

Divjak, B., Žugec, P., & Pažur Aničić, K. (2022). E-assessment in mathematics in higher education: A student perspective. *International Journal of Mathematical Education in Science and Technology, 1–23,* 1–23. https://doi.org/10.1080/0020739X.2022.2117659

Dixson, D. D., & Worrell, F. C. (2016). Formative and summative assessment in the classroom. *Theory Into Practice, 55*(2), 153–159. https://doi.org/10.1080/00405841.2016.1148989

Ekolu, S. (2022). Practical prediction of overall performance from formative assessment results of engineering students. *International Journal of Engineering Education, 38*(4), 1106–1115.

Hand, D. J., & Till, R. J. (2001). A simple generalisation of the area under the ROC curve for multiple class classification problems. *Machine Learning, 45*(2), 171–186. https://doi.org/10.1023/A:1010920819831

Hosmer, D. W., & Lemeshow, S. (2000). *Applied logistic regression* (1st ed.). Wiley. https://doi.org/10.1002/0471722146

Kabathova, J., & Drlik, M. (2021). Towards predicting student's dropout in university courses using different machine learning techniques. *Applied Sciences, 11*(7), 3130. https://doi.org/10.3390/app11073130

Kursa, M. B., & Rudnicki, W. R. (2010). Feature selection with the Boruta package. *Journal of Statistical Software, 36*(11), 1–13. https://doi.org/10.18637/jss.v036.i11

Mandrekar, J. N. (2010). Receiver operating characteristic curve in diagnostic test assessment. *Journal of Thoracic Oncology, 5*(9), 1315–1316. https://doi.org/10.1097/JTO.0b013e3181ec173d

Merritt, D. L., & Buboltz, W. (2015). Academic success in college: Socioeconomic status and parental influence as predictors of outcome. *Open Journal of Social Sciences, 3*(5), 127–135. https://doi.org/10.4236/jss.2015.35018

Nembrini, S., König, I. R., & Wright, M. N. (2018). The revival of the Gini importance? *Bioinformatics, 34*(21), 3711–3718. https://doi.org/10.1093/bioinformatics/bty373

Ramsden, P., & Ramsden, P. (2003). *Learning to teach in higher education.* Routledge. https://doi.org/10.4324/9780203507711

Rienties, B., Balaban, I., Divjak, B., Grabar, D., Svetec, B., & Vondra, P. (2023). Applying and translating learning design and analytics approaches across borders. In O. Viberg & Å. Grönlund (Eds.), *Practicable learning analytics* (pp. 35–53). Springer International Publishing. https://doi.org/10.1007/978-3-031-27646-0_3

Saqr, M., Fors, U., & Tedre, M. (2017). How learning analytics can early predict under-achieving students in a blended medical education course. *Medical Teacher, 39*(7), 757–767. https://doi.org/10.1080/0142159X.2017.1309376

Saqr, M., Jovanovic, J., Viberg, O., & Gašević, D. (2022). Is there order in the mess? A single paper meta-analysis approach to identification of predictors of success in learning analytics. *Studies in Higher Education, 47*(12), 2370–2391. https://doi.org/10.1080/03075079.2022.2061450

Sghir, N., Adadi, A., & Lahmer, M. (2023). Recent advances in predictive learning analytics: A decade systematic review (2012–2022). *Education*

and Information Technologies, *28*(7), 8299–8333. https://doi.org/10.1007/s10639-022-11536-0

Susnjak, T., Ramaswami, G. S., & Mathrani, A. (2022). Learning analytics dashboard: A tool for providing actionable insights to learners. *International Journal of Educational Technology in Higher Education*, *19*(1), 12. https://doi.org/10.1186/s41239-021-00313-7

Tempelaar, D., Rienties, B., & Giesbers, B. (2015a). Stability and sensitivity of learning analytics based prediction models: *Proceedings of the 7th International Conference on Computer Supported Education*, 156–166. https://doi.org/10.5220/0005497001560166

Tempelaar, D., Rienties, B., Mittelmeier, J., & Nguyen, Q. (2018). Student profiling in a dispositional learning analytics application using formative assessment. *Computers in Human Behavior*, *78*, 408–420. https://doi.org/10.1016/j.chb.2017.08.010

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2015b). Understanding the role of time on task in formative assessment: The case of mathematics learning. In E. Ras & D. Joosten-ten Brinke (Eds.), *Computer assisted assessment. Research into e-assessment* (Vol. 571, pp. 120–133). Springer International Publishing. https://doi.org/10.1007/978-3-319-27704-2_12

Tempelaar, D. T., Rienties, B., & Giesbers, B. (2016). Verifying the stability and sensitivity of learning analytics based prediction models: An extended case study. In S. Zvacek, M. Restivo, J. Uhomoibhi, & M. Helfert (Eds.), *Computer supported education. CSEDU 2015. Communications in computer and information science* (Vol. 583). Springer. https://doi.org/10.1007/978-3-319-29585-5_15

van der Vleuten, C. P. M., & Schuwirth, L. W. T. (2005). Assessing professional competence: From methods to programmes. *Medical Education*, *39*(3), 309–317. https://doi.org/10.1111/j.1365-2929.2005.02094.x

Vermunt, J. D., & Vermetten, Y. J. (2004). Patterns in student learning: Relationships between learning strategies, conceptions of learning, and learning orientations. *Educational Psychology Review*, *16*, 359–384. https://doi.org/10.1007/s10648-004-0005-y