

Prediction of student academic performance using Moodle data from a Further Education setting

Rory Joseph Quinn^{*a} and Dr. Geraldine Gray^a

^a Technological University Dublin

(Received February 2019; final version received May 2019)

Abstract

Increasingly, educational providers are being challenged to use their data stores to improve teaching and learning outcomes for their students. A common source of such data is learning management systems which enable providers to manage a virtual platform or space where learning materials and activities can be provided for students to engage with. This study investigated whether data from the learning management system Moodle can be used to predict academic performance of students in a blended learning further education setting. This was achieved by constructing measures of student activity from Moodle logs of further education courses. These were used to predict alphabetic student grade and whether a student would pass or fail the course. A key focus was classifiers that could predict likelihood of failure from data available early in the term. The results showed that classifiers built on all course data predicted student grade moderately well (accuracy= 60.5%, kappa = 0.43) and whether a student would pass or fail very well (accuracy= 92.2%, kappa=0.79). However, classifiers built on the first six weeks of data did not predict failing students well. In contrast, classifiers trained on the first ten weeks of data improved statistically significantly on a no-information rate ($p < 0.008$) though slightly more than half of failing students were still misclassified. The ability to detect early in the course even a minority of students at risk of failure is likely to be of use to course administrators given the economic cost involved. The evidence indicates that measures of Moodle activity on further education courses could be useful as part of an early-warning system at ten weeks.

*Corresponding author. Email: quinnroryj@hotmail.com. Irish Journal of Technology Enhanced Learning Ireland, 2020. © 2020 Quinn, R & Gray G. The Irish Journal of Technology Enhanced Learning Ireland is the journal of the Irish Learning Technology Association, an Irish-based professional and scholarly society and membership organization. (CRO# 520231) <http://www.ilta.ie/>. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0>), allowing third parties to copy and redistribute the material in any medium or format and to remix, transform, and build upon the material for any purpose, even commercially, provided the original work is properly cited and states its license.

1. Introduction

Datafication, a term introduced to popular usage by Mayer-Schonberger and Cukier (2013), can be understood as the conversion of aspects of peoples' unstructured everyday experience into a structured format that can be analysed in a formal system. Datafication is becoming pervasive in modern society, for example in health (Ruckenstein & Dow Schüll, 2017), urban planning (Tenney & Sieber, 2016), human resources (Chamorro-Premuzic, Akhtar, Winsborough & Sherman, 2017), justice (Smith, Bennett Moses & Chan, 2017). New technologies result in data being created, often as a by-product of their use. The ease and affordability of storing large amounts of data, and the development of tools and methodologies to extract knowledge from this data, fuels interest in areas like data mining, data science and big data. Organisations operating in diverse sectors increasingly view data as a resource whose intrinsic value can be realised through the creation of data products.

Advances in learning technologies mean that the process of datafication is occurring in education too. Educational organizations now have larger repositories of data than ever before and are being challenged to use this data to improve teaching and learning outcomes and to enable evidence-based decision making. The European Commission recently published a Communication on the Digital Action plan which identified three priority areas for action towards making better use of innovation and technology in education and training (European Commission, 2018). The third of which is improving education through better analysis and foresight. The Data Strategy for the Department of Education and Skills (2017) has as one of its objectives to maximize the value and use of data to improve the learning experience and the success of learners. This is one of four key objectives the Department set to realize its objective of delivering "First Class Data for Education".

This study takes place within the context of the rapid innovation in technology enhanced learning of recent years. This has led to growing volumes of data being stored by educational institutions and a consequent emphasis by educational policy makers on the need for organisations to become more data informed. One potentially useful source of data that many learning providers have access to nowadays is that produced by learning management systems (LMS), also sometimes referred to as course management systems or virtual learning environments. These systems allow for the creation of online learning spaces where students can access learning resources and activities and interact with one another and the instructor.

LMS can be used to deliver courses wholly online or as a complement to more traditional classroom teaching (blended learning) and there now exist a range of such systems both proprietary and open source (Ülker & Yilmaz, 2016). A useful feature of LMS is that they provide the ability to quantify students learning behavior and interactions with the learning space in ways that are more difficult to do in traditional 'face to face' learning. LMS automatically track and store learner's interactions with the system and this data can then be mined or analyzed for insight for example to model student behavior or predict academic performance (Papamitsiou & Economides, 2014).

In a recent review of studies published between 2012 and 2018, Viberg, Hatakka, Balter & Mavroudi (2018) identified 252 papers in English dealing with learning analytics in university populations alone. Hellas, Ihantola, Petersen, Ajanovski, Gutica, Hynninen, Knutas, Leinonen, Messom & Nam Liao (2018) found 357 papers between 2010 and 2018 dealing with the prediction of academic performance. Prediction of student academic

performance using data from LMS and/or other sources is a common task in educational data mining and learning analytics.

Tempelaar, Rienties & Giesbers (2015) compared the predictive value of learning dispositions, demographic data from student information systems, data from entry tests, formative assessment results and LMS data. They concluded that the LMS data, which was from Blackboard in their case, did not substantially predict academic performance and that results of formative assessment best predicted underperforming students. Lu, Huang, Huang, Lin, Ogata & Yang (2018) found best predictive performance was given by a dataset which combined traditional measures (e.g. homework scores, quiz scores) with measures of online resource usage (e.g. number of online activities a student engages in per week, number of times a student plays a video per week).

Zacharis (2015) found that variables extracted from Moodle LMS usage explained just over 50% of the variance in final course grade. Jo, Kim, & Yoon (2015) using just LMS data from which they extracted proxy variables measuring students time management strategies explained approximately 34% of the variance in final test scores. Gasevic, Dawson, Rogers & Gasevic (2016) using a large sample of students from 9 university courses found the variance explained by their model increased from 5%, when only student characteristics were used to predict percent mark, to 16% when trace data from Moodle was added. When analysed by course, Moodle data explained between 2% and just over 70% in percent mark depending on the course.

In addition to predicting numerical measures of academic performance researchers have also tried to predict categorical outcome measures such as alphabetic grade or pass/fail outcomes. Macfadyen & Dawson (2010) used logistic regression to correctly classify almost 74% of students in their study into pass and fail categories. Accuracy at classifying failing students was 81%. Zacharis (2015) achieved accuracy of almost 70% at identifying failing students and overall accuracy of 81.3%. Conijn, Kleingeld, Matzat, Snijders & van Zaanen (2016) got 68.7% overall accuracy on pass/fail prediction. Raga Jr. & Raga (2017) achieved accuracy of over 87% with their best performing algorithm on a three-class prediction task. Azcona & Casey (2015) achieved 91% accuracy at predicting whether students would pass or fail using data from a bespoke learning platform. These studies with university cohorts are using all course data from relatively short courses (less than 15 weeks). The usefulness of being able to predict academic performance using data from the full duration of a course is an open question, since predictions are only available once the course has finished.

It is apparent that wide differences exist in predictive utility of LMS data from study to study. In general research indicates that the addition of LMS data to models predicting academic performance does improve predictive accuracy but by how much is rather more difficult to quantify. Given that studies use different predictor variables, derived from activity on different types of LMS, from different courses, taught by different instructors; it should not be surprising that there are large differences in predictive accuracy from study to study. However, there can be wide differences in predictive accuracy of models trained on data from different courses even within the same institution (Gasevic et al, 2016; Conijn, Snijders, Kleingeld & Matzat, 2017). One of the main reasons for this is likely due to difference in course design and instruction (Macfadyen & Dawson, 2010; Gasevic et al, 2016).

One of the purposes of research into the prediction of student performance is the identification of at-risk students. If students at risk of failing or drop-out can be identified

early enough in the course, it may be possible to provide interventions to prevent this outcome. In this case there is usually a trade-off between time of prediction and accuracy of prediction. Accuracy generally increases over time as more course data becomes available (Howard, Meehan & Parnell, 2018). There have been some exceptions, for example, Sandoval, Gonzalez, Alarcon, Pichara and Montenegro (2018) who found better predictive accuracy for aggregated LMS data earlier in the course. This may be because most of the activity on their LMS was passive and students with high grades showed decreasing or no activity towards the end of the semester.

Most studies in this area so far have used populations of university students. Examples from an Irish context include Azcona & Casey (2015), Gray, McGuinness & Owende (2014) and Howard et al (2018). Little, if any, research has been carried out with students in a further education context in Ireland. The aim of this study was to predict student performance using data from a LMS in a further education setting in Ireland. Further education classes are typically smaller than university ones, so this necessitated combining data from multiple classes which may lessen predictive accuracy due to increased sources of variance. On the other hand, all classes had the same tutor and were studying subjects in the same discipline which controls for instructional factors to some extent.

There were two research questions in this study:

1. Using measures created from Moodle activity from the full duration of a course, is it possible to predict student academic performance on further education courses?
2. Using the same data but only for the early weeks of a course, the first six weeks and the first ten weeks, is it possible to predict whether a student will pass or fail?

2. Methods

The Moodle installation used for this study was that of Limerick and Clare Education and Training Board (LCETB), a statutory education and training authority formed in 2013 from the amalgamation of three former Vocational Education Committees (Co Clare, Co Limerick and City of Limerick). There are more than 900 courses and 8000 users on the LCETB Moodle site though not all of these are currently active.

2.1 Dataset

The outcome variables for this study were created from the results of students attending full time post leaving certificate courses between 2011 and 2018. The results used were from 29 classes in 9 different modules. All modules were in the same discipline and had the same instructor resulting in a relatively homogenous data sample. Logs for these classes were downloaded as comma separated value files using the logs interface in the report's menu in Moodle. Of the 29 classes in the dataset, 20 showed a full academic year of student Moodle activity (33 weeks approximately), 1 showed 31 weeks activity, 4 showed 21-24 weeks activity, 1 showed 18 weeks activity and 2 showed 11-12 weeks activity.

Results were received for 607 of the 690 students enrolled on Moodle for the classes included in the study. Following consultation with the tutor it was determined that the remaining 83 enrolments were students who had left courses prior to course end date and were not submitted for certification. These students were given a mark of 0, a grade of EE (early exit)

and a label of F (fail) and added to the 25 enrolments in the results file who had received a final mark of 0. This meant there were 108 enrolments in total with the label Early Exit. Figure 1 below shows the distribution of grades across the dataset including the percentage.

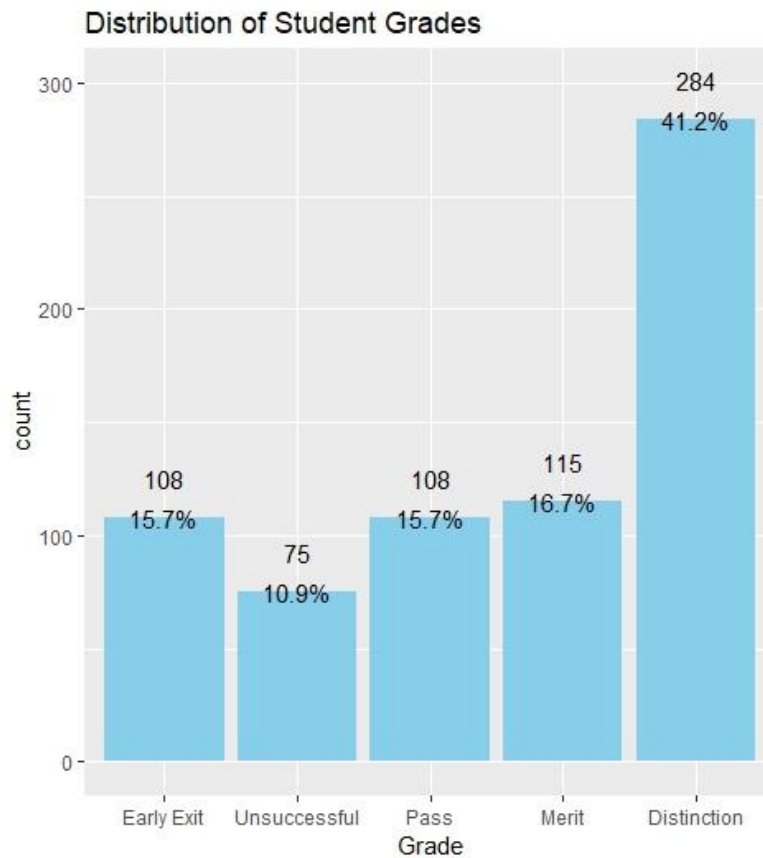


Figure 1: Distribution of Grades

The final number of enrolments ($n=690$) represented 410 individual students since some classes contained the same student cohort on more than one module. The overall early exit rate for this sample was 15.7% which is almost identical to the national dropout rate on post-leaving certificate courses estimated from a recent ESRI survey (Solas, 2017). This dataset was restricted by availability of results and represents only a small sample of all the users in the system. Thus, results need to be interpreted in the context of a small relatively homogenous student sample.

2.2 Classification Tasks

The first classification task was to determine whether it is possible to predict student performance using Moodle data from a blended learning course in a further education setting. This involved predicting student alphabetic grade and whether a student would pass or fail using variables created from Moodle data for the full duration of a course.

The second classification task was to determine whether it is possible to predict early in a course whether a student will pass or fail. This involved predicting the binary pass/fail label using variables created from Moodle data for the first six weeks and the first ten weeks of a

course. The time periods six and ten weeks were chosen to try and ensure enough data was available to accurately train a classifier while at the same time being early enough to allow for effective intervention.

There were five class labels used for grade prediction: Distinction (students with mark of 80%+), Merit (65-79%), Pass (50-64%), Fail (1-49%) and Early Exit (0% or no grade). Converting grades to class labels in this way loses some information since classification algorithms ignore order in the class labels.

The pass-fail outcome variable was formed by combining Early Exit and Fail students under the Fail label and the Distinction, Merit and Pass students under the Pass label. Early Exit and Fail students were combined under the one class label to ensure adequate sample size. Date of course exit was not known for students and assigning students to Fail or Early Exit categories was based on the percentage mark achieved on the course rather than the date of course exit. This provides further rationale for combining these two categories.

2.3 Data Preprocessing

To investigate the research questions in this study, a dataset was created where each instance was a student, and variables represented aggregates of various aspects of Moodle usage. Each instance was associated with the outcome variable *student performance* as measured both by final course grade and pass/fail label.

Log data generated by roles other than ‘student’ were removed from the log files. *Module name* and *type* were derived from text pattern matching on the *Event Name* and *Component* fields in the log data and aggregated on the *Course* and *User Name* fields. Aggregations were based on the variables used in Raga Jr. & Raga (2017) with the addition of login regularity. Login regularity shows how often a user logged in relative to other users on the same course and so accounts for differences in login frequency between courses. Table 1 below lists all the variables used in this study.

Table 1: Predictor Variables Used in the Study

Variable	Description
Total Activity	Number of Log Entries by User
Course Views	Number of Course Views by User
URL Views	Number of URL Views by User
Page Views	Number of Page Views by User
Assignment Views	Number of Assignment Views by User
Assignment Submissions	Number of Assignment Submissions by User
Quiz Views	Number of Quiz Views by User
Quiz Attempts	Number of Quiz Attempts by User
Lesson Views	Number of Lesson Views by User
Lesson Attempts	Number of Lessons Started/Ended/Answered by User
File Views	Number of File Views by User
Forum Posts	Number of Posts made to the Forum by User
Weekday Logins	Number of logins by User during the week (Mon-Fri)

Weekend Logins	Number of logins by User during the weekend (Sat-Sun)
Am Early	Number of logins by User between 0 and 6am
Am Late	Number of logins by User between 6am and 12pm
Pm Early	Number of logins by User between 12pm and 6pm
Pm Late	Number of logins by User between 6pm and 12am
On-Campus Login	Number of logins by User from an on-campus IP address
Off-Campus Login	Number of logins by User from an off-campus IP address
Login Regularity	Number of unique days this User logged into the course normalized by number of unique days any User logged into the course.

Once the variables were created the variable files and results files were merged. The identifier field was then removed to ensure data analysis was compliant with data protection requirements. Course names were also removed and replaced with alphabetic identifiers. All predictor variables were standardized to the range 0-1 prior to applying algorithms. This was done to avoid variables with larger ranges being given undue weight by the algorithms (Han, Kamber & Pei, 2012).

2.4 Algorithms and Software Used

The algorithms used in this study were Random Forest, Gradient Boosting, k Nearest Neighbours and Linear Discriminant Analysis. These were chosen due to having shown good predictive performance in previous studies (see for example: Sandoval et al, 2018; Howard et al, 2018; Raga Jr & Raga 2017; Kondo, Okubo & Hatanaka 2017).

The open source statistical computing software R (R Core Team, 2018) was used for analysis. The caret package (Kuhn, 2008) in R was used for the machine learning workflow. The implementation of Random Forest used in this study was that in the randomForest package (Liaw & Wiener, 2002). The gbm package (Ridgeway, 2007) was used for Gradient Boosting. Linear discriminant analysis and k Nearest Neighbour were implemented using the lda and knn methods in caret. The varImp method in caret was used to calculate variable importance. It does this by calculating the difference in prediction accuracy, averaged across all trees in the random forest, resulting from permuting each of the predictor variables (Kuhn, 2018).

2.5 Model Building and Evaluation

The dataset was split into a training and test set in the ratio 70-30, the training set used to build the models and the test set used to check accuracy (Burger, 2018). Classifiers were trained on the training set using ten-fold cross validation with ten repeats to tune the hyper-parameters. Hyper-parameters were varied using random search (Bergstra & Bengio, 2012). The hyper-parameters randomly varied were: mtry (number of variables to be selected at each split) for random forest; n.trees (number of trees), interaction.depth (tree complexity), shrinkage (learning rate), n.minobsinnode (minimum number of instances in a node before splitting should commence) for Gradient Boosting Machines; the value for k for k Nearest Neighbours. Linear Discriminant Analysis does not have hyper-parameters for tuning. The most accurate models on cross validation were then tested for accuracy on the test set to ensure the parameter tuning did not overfit the data.

In addition to overall model accuracy, the accuracy statistics recall, specificity, precision, and balanced accuracy were calculated. Recall shows the proportion of positive examples correctly labelled by the classifier, specificity is the proportion of negative examples correctly labelled, while precision shows the proportion of examples correctly labelled as positive as a proportion of all positive predictions made by the classifier (Han et al, 2012). For the pass/fail prediction task, fail was set as the positive class while for multi class prediction (the grade prediction task) a ‘one vs all’ approach was used (Kuhn, 2018) so for example in the grade prediction task, these measures are calculated for each grade against all the examples in the other classes. Balanced accuracy was calculated as the average of recall and specificity (Kuhn, 2018).

Cohen’s kappa statistic was calculated for each model. Cohen’s kappa statistic shows how well the models’ predictions for class labels agreed with the actual class labels while controlling for the accuracy of a random classifier that guesses according to the frequency of each class (Gwet, 2014). Landis & Koch’s (1977) suggested guidelines for interpreting Cohen’s kappa were followed: Slight Agreement 0-0.2, Fair Agreement 0.21-0.40, Moderate Agreement 0.41-0.60, Substantial Agreement 0.61-0.80, Almost Perfect Agreement 0.81-1.0. Model accuracies were also compared to the no-information rate using a one-tailed binomial test to test whether the number of correct predictions significantly exceeded the no-information rate. The no-information rate is the accuracy achievable by always predicting the most common class label in the test set.

3.Results

3.1 Prediction Using All Course Data

The distribution of five class labels used for grade prediction (Figure 1) meant the no-information rate, or accuracy achievable by always predicting the most common class label (Distinction) was 41.2%. Table 2 shows the accuracy, 95% confidence interval and Cohen’s kappa statistic for each of the algorithms on the test set.

Table 2: Accuracy of each Algorithm on Test Set – Grade Prediction

	Random Forest	LDA	k-NN	GBM
Test Accuracy	60.5%	59.5%	59%	57.1%
95% CI	53-67%	52-66%	52-66%	50-64%
Kappa	0.43	0.42	0.42	0.41

Random forest was the best performing algorithm with accuracy of just over 60% but the results were almost equivalent across all classifiers. The one-tailed binomial test indicated that all the algorithms significantly outperformed the no-information rate. Although results differed slightly depending on the seed value used, the confidence interval indicates that there is a 95% likelihood that the true accuracy of this model lies in the range 53% to 67%.

Table 3: Accuracy Statistics by Class for Random Forest

Class	Recall	Specificity	Precision	Balanced Accuracy
Distinction	0.92	0.65	0.65	0.78
Merit	0.15	0.94	0.31	0.54
Pass	0.25	0.90	0.32	0.58
Unsuccessful	0.23	0.98	0.63	0.61
Early Exit	0.88	0.95	0.78	0.91

Recall is very good for students at either end of the range of grades, i.e. those with grades of distinction and early exit, but not good for the other classes. Specificity is excellent for all classes except for distinction for which it is moderate. Class statistics for the other algorithms showed a similar pattern in that sensitivity for the class labels Distinction and Early Exit was good but much lower for the other three class labels.

The distribution of the pass-fail outcome variable was somewhat imbalanced as there were almost three times as many passing students as there were failing students. This means that the no-information rate, or accuracy achievable by always predicting Pass, was 73.5%. When usage data from the whole course was used to create the predictors, all the algorithms showed good performance on the binary classification task. Random Forest (92.2% accuracy) and LDA (89.3% accuracy) performed somewhat better than Gradient Boosting (86.9%) and k-NN (85% accuracy) for this task, but all the algorithms performed significantly better than the no-information rate on this task. Best performing algorithm was Random Forest correctly predicting 148 out of 152 passing students and 42 out of 54 failing students ($\kappa=0.79$).

3.2 Prediction Using 6-week and 10-week datasets

To see if it was possible to predict students at risk of failure early enough so that intervention could take place, the same variables as in the previous task were used but these were constructed on subsets of the data, the first six weeks and the first ten weeks of Moodle data for each course.

Accuracy of classifiers trained on variables created using the six-week data set is shown in Table 4. As the courses in this dataset are blended learning courses, not all students show Moodle activity in the first six weeks of the course resulting in a reduced dataset ($n=644$) and the no-information rate for this dataset is 75.5% (Passing Students = 486). None of the classifiers significantly improved on the no-information rate when only six weeks of usage data was used to train them. Recall of all the classifiers is poor.

Table 4: Accuracy of Pass/Fail Classification on 6-week data set

	Random Forest	k-NN	GBM	LDA
Test Accuracy	74.5%	73.4%	69.8%	75.5%
Recall	0.3	0.15	0.09	0
Specificity	0.89	0.92	0.9	1

Kappa	0.21	0.09	-0.02	0
-------	------	------	-------	---

Table 5 shows accuracy of the classifiers constructed on the first ten weeks of usage data. Nearly all students are using Moodle by week ten of the courses and the no-information rate for this dataset is 74.8% (n=675, Passing Students=505). Accuracy and kappa statistic have improved on the six-week models and the algorithms are now showing predictive accuracy significantly greater than the no-information rate. For example, a binomial test comparing random forest with the no-information rate resulted in a p-value of 0.008.

Table 5: Accuracy of Pass/Fail Classification on 10-week data set

	Random Forest	GBM	LDA	k-NN
Test Accuracy	82.18%	80.2%	79.2%	77.2%
Recall	0.41	0.45	0.35	0.29
Specificity	0.96	0.92	0.94	0.93
Kappa	0.44	0.41	0.35	0.27

3.3 Variable Importance

Figure 2 below shows the correlation of predictor variables based on all course data with student's final percentage mark for the whole data set and for each module. The colour gradient on the heatmap is red-white-blue. Grey boxes indicate a frequency of zero for that activity for all students on the course (na value), in other words that activity was not used on the course.

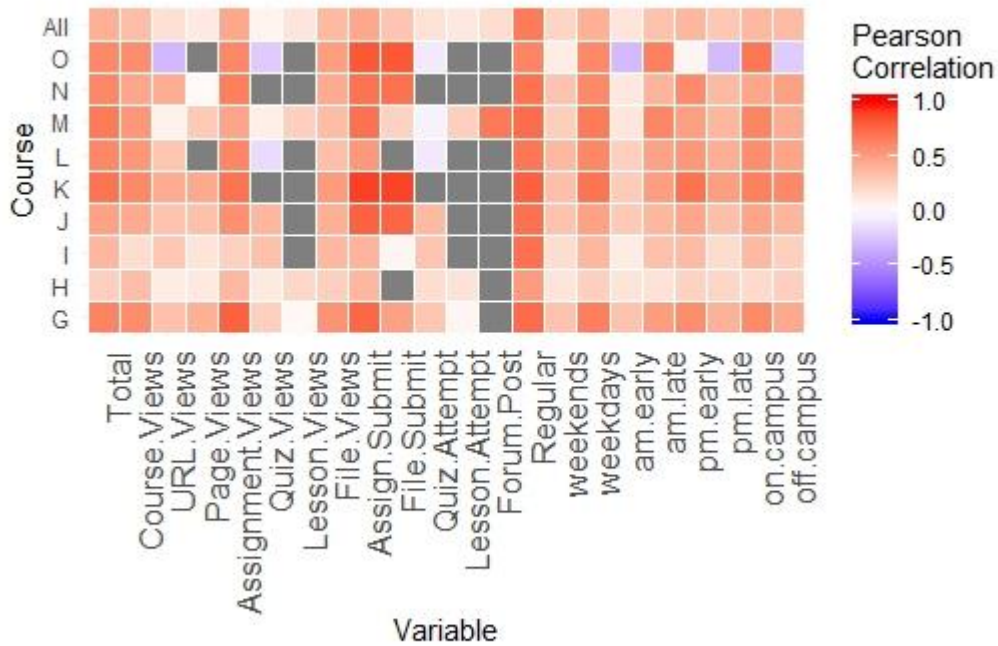


Figure 2: Correlation between percentage mark and predictor variables

It is apparent that in general there is a positive linear relationship between frequency of activity on Moodle and student mark. Variables showing highest correlations (>0.4) with percentage mark for the whole dataset are login regularity, assignment views, assignment submissions, total activity and weekday activity.

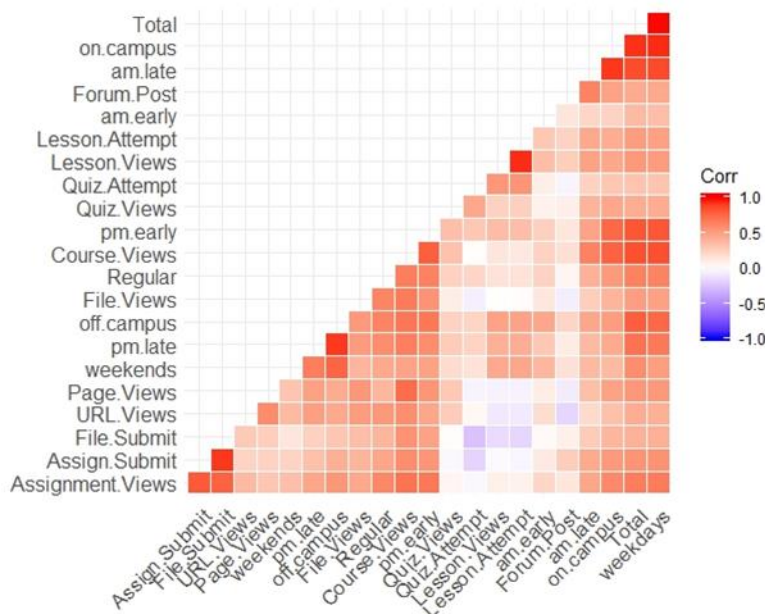


Figure 3: Correlation matrix of Predictor Variables

Figure 3 shows the correlation matrix of predictor variables. It is apparent that many of the variables are positively correlated to some extent. There is some redundancy in the dataset since all the variables except for Login Regularity are composed of frequency counts, for example Total Activity and Weekday Activity exhibit almost perfect correlation since over 95% Moodle activity in this dataset occurs on a weekday.

Figures 4 and 5 show the ten most important variables ranked in descending order of importance for random forest on the pass/fail prediction using all data (figure 4) and data from just the first ten weeks of the courses (figure 5).

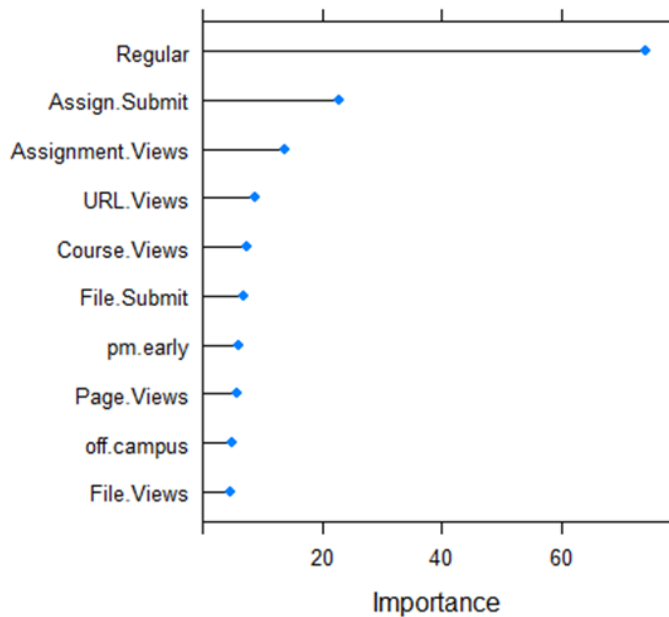


Figure 4: Variable Importance in Random Forest (Pass/Fail – All Data)

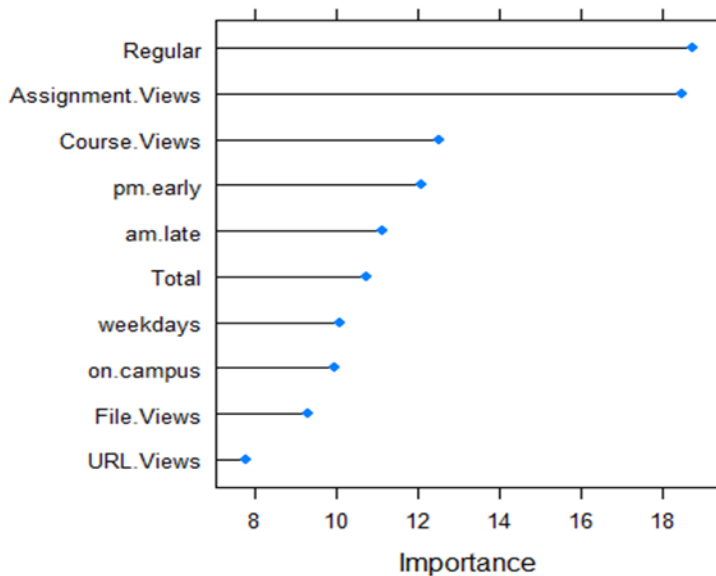


Figure 5: Variable Importance in Random Forest (Pass/Fail -10 Week dataset)

The x axis in the figures indicates the impact on accuracy inferred from permuting the variables. Variables with higher impact have higher importance scores. Following the advice of Strobl, Boulesteix, Zeileis & Hothorn (2007) the importance scores have not been normalised.

4. Discussion

The first research question in this study considered whether it is possible to predict student academic performance on further education courses using variables created using all course data. The second research question was whether it is possible to predict performance using variables created from early course data only, in this case the first six weeks and first ten weeks of courses.

The results of this study showed that when all course data is used, it is possible to predict student alphabetic grade at a level significantly above the no-information rate using only Moodle data. All four algorithms used performed significantly better than the no-information rate, and all performed within 5% accuracy of one another. Best performance on the grade prediction task was given by random forest with accuracy of 60.5% and a kappa of 0.43. This indicates we can predict student performance as measured by final course grade on further education courses moderately well using just logs of student Moodle activity. This performance is close to that found by other researchers (Raga Jr. & Raga, 2017) using similar predictors, albeit they used a different type of student population and a three-class prediction task. Most researchers on tasks such as this have not reported the kappa value of their models which can make comparison difficult particularly when the number of outcome classes differs between studies and the classes are imbalanced.

Although overall accuracy was in the moderate range on the grade prediction task, the predictive accuracy of grades at either end of the distribution was high. This may be because there were more examples of the grade Distinction with which to train a classifier and because the difference in usage patterns between students Early Exit and Distinction students was more marked than differences between students with a Fail, Pass or Merit grade.

For the Pass/Fail classification task, it was possible to predict very well whether students would pass or fail using all course data. Random Forest correctly predicted 42 out of 54 failing students and 148 out of 152 passing students. In other words, approximately 78% of failing students and almost 97% of passing students were correctly identified for an overall accuracy of more than 92%. This level of performance compares well with previous studies, (Macfayden & Dawson, 2010; Conijn et al, 2016) particularly so when one considers that this study used Moodle data only whereas other studies included additional information such as assessment scores. Several studies have shown that course assessment scores are a useful predictor of student performance (Tempelaar et al, 2015; Conijn et al, 2017). Additionally, the dataset in this study was comprised of students from different courses which can lessen predictive accuracy (Gasevic et al, 2016).

There are several possible reasons for the relatively good performance of the binary classifiers using data from the full duration of courses. Most courses in this dataset were of longer duration than those in other studies, 20 of the 29 courses were of more than 30 weeks

duration, meaning that more data was available to train the models. Studies using LMS data from university courses typically use data from relatively short courses, for example Tempelaar et al (2015) used only seven weeks of LMS data when concluding that such data did not substantially predict academic performance. It is also worth noting that although levels of Moodle usage varied a lot both between and within courses, there was proportionally less ‘passive’ activity (viewing content) on the courses in this study than some previous researchers found. Less than 83% of the activity in this dataset was passive. Sandoval et al (2018) found over 95% of the activity in their dataset was passive. In this study although there was little activity indicating interaction between students there was proportionally more activity indicating interaction with the system (e.g. taking quizzes, submitting assignments). Interestingly, on the only course (course M) which utilised the forum that allows students to interact with one another, the variable number of forum posts was highly correlated with student’s final course mark (Pearson correlation coefficient=0.67, see Figure 2).

One of the main benefits of being able to predict whether students will pass or fail a course is to provide some type of assistive intervention to “likely to fail” students so that this outcome can be avoided. Therefore, models were built on two subsets of the original Moodle activity logs. One consisting of data for the first six weeks of courses and the other consisting of data for the first ten weeks of courses. Although performance of the binary classifiers was good when LMS data for the full duration of courses was used to create the predictors, performance on these datasets from early courses was not as good.

None of the models significantly improved on the no-information rate on the six-week dataset. Random forest, the best performing model, correctly predicted only 14 out of 47 failing students. The addition of four extra weeks of data improved performance somewhat and both random forest and gradient boosting performed significantly better than chance on the ten-week dataset. Even here however the models were still only correctly able to identify a minority of failing students (23 out of 51 for Gradient Boosting). This might suggest that although LMS data can be a useful component of early warning models it is insufficient on its own to identify most failing students. However, this study did not include as predictors, assessment results, multiple fine-grained measures of LMS usage or (with the exception of one course which showed a positive correlation) measures of student interaction with one another. Given that these have been shown to be predictive of academic performance previously (Azcona & Casey, 2015; Civitas 2016; Conijn et al, 2016; Howard et al 2018; Macfadyen & Dawson, 2010; Zacharis, 2015), it is likely that the predictive accuracy observed in this study can be improved further.

Although random forest showed greater overall accuracy on the ten-week prediction task, gradient boosting classified more failing students correctly (23 vs 21 for random forest) and might be preferred if it could be shown to consistently demonstrate higher recall since the cost of misclassification on this task is not the same for each class. Failing students will likely cost both themselves and the college more than an unnecessary intervention or offer of help resulting from a false positive will. Given the relatively low cost overhead involved in implementing an early warning system based on LMS usage, the ability to detect early a minority of students at risk of failure could be very useful, even if it only prevented a small number of students from early exit or course failure.

Regarding variable importance, regularity of login was the most important predictor for both the whole course dataset and the early course datasets. The magnitude of importance of this

variable was lower for the early datasets, however. Logging in regularly was associated with better final course marks and grades. This may be considered to support the view held by some researchers that proxy variables measuring students time management strategies can be useful when predicting student performance (Jo et al, 2015). It is not surprising that this variable is important as it may be considered a proxy for attendance. It shows how many days a student logged in as a proportion of all the days that any student logged into the course. Number of assignment submissions, which was the second most important variable for the grade prediction task, dropped out of the top ten most important variables for the early warning models. Most assignment submissions are likely to occur later in the course which means this variable is unavailable as a useful predictor for early warning models and may be one of the reasons why performance suffers in comparison with models built on data from the whole duration of courses.

The most predictive variable across all datasets in this study, login regularity, is a derived one rather than a frequency count. This might reinforce the suggestion that attempts to improve predictive accuracy should include derived measures of LMS usage. These could usefully include measures relating to student's attendance, fine-grained measures of engagement with course materials, and interactions with one another in the LMS. Best performance is likely to come from using datasets that combine features derived from LMS usage with more traditional indicators such as assessment results (Lu et al, 2018). Course designers should therefore give thought to building some form of early assessment into their courses and to utilising LMS features that promote student to student interaction.

5. Conclusion

Moodle use is integrated across a variety of programmes within the Further Education Division of Limerick and Clare Education and Training Board. The system saw over 100,000 logins during the calendar year 2017. There are over 900 courses on the Moodle site though not all of these are active. The analyses for this study were based on a small subset of these courses consisting of 690 enrolments across nine modules over six years all delivered by the same faculty member.

There were two related research questions in this study. Both considered whether it is possible to predict student performance using only Moodle data from a blended learning Further Education setting. The first of these looked at whether student's performance can be predicted using Moodle data across the whole duration of a course and the second at whether a student will pass or fail can be predicted using only Moodle data gathered early in a course i.e. at six and ten weeks.

The results showed that by using Moodle data from the whole course it was possible to predict student alphabetic grade moderately well and whether a student would pass or fail very well. All the algorithms performed statistically significantly better than the no-information rate on these tasks with best performance given by random forest. To be of more practical use however it is desirable to know early in the course whether a student may be at risk of failing the course.

Classifiers trained on a six-week dataset performed no better than chance but those trained on a ten-week dataset did significantly better than chance. Even the algorithm best able to

identify failing students at ten weeks only classified 23 out of 51 failing students correctly therefore it is concluded that although Moodle data may be useful as a component of early warning models at ten weeks it is unlikely to be sufficient on its own to accurately predict most failing students. Predictive accuracy might improve using other variables derived from LMS usage and combining these with offline information such as assessment results. The ability to predict early students at risk of failure may be considered useful even if only a minority of such students are identified since the benefits might well outweigh the costs.

Nearly all variables in the dataset showed a positive correlation with student mark indicating that in general higher levels of Moodle usage were associated with higher final course marks. On the best performing algorithm, random forest, regularity of login was the most important variable when predicting academic performance across all datasets.

Ideally, the results observed in this study should be replicated with larger FET datasets. The sample size in this study was relatively small. All students in this sample were attending modules in the same discipline with the same instructor, which limits generalisability of results. It remains to be seen what degree of predictive accuracy is achievable on Moodle data aggregated across a broader set of FET disciplines. More fine-grained measures of LMS usage, combined with results of formative assessment, may be useful for increasing accuracy and shortening the time scale further within which at-risk students can be identified.

References

- Azcona, D. & Casey, K. (2015). Micro-analytics for student performance prediction leveraging fine-grained learning analytics to predict performance. *International Journal of Computer Science and Software Engineering*, 4(8), 218-223.
- Bergstra, J., & Bengio, Y. (2012) Random search for hyper-parameter optimisation. *Journal of Machine Learning Research*, (13), 281-305. Retrieved from <http://dblp.uni-trier.de/db/journals/jmlr/jmlr13.html#BergstraB12>
- Burger, S.V. (2018). *Introduction to Machine Learning with R*. Sebastopol, USA: O'Reilly.
- Casey, K. & Azcona, D. (2017) Utilising student activity patterns to predict performance. *International Journal of Educational Technology in Higher Education*, 14(4), <https://doi.org/10.1186/s41239-017-0044-3>
- Chamorro-Premuzic, T., Akhtar, R., Winsborough, D., & Sherman, R.A. (2017). The datafication of talent: how technology is advancing the science of human potential at work. *Current Opinion in Behavioural Sciences*, 18, 13-16. <https://doi.org/10.1016/j.cobeha.2017.04.007>
- Civitas Learning (2016) *Community Insights: Emerging Benchmarks and Student Success Trends From Across the Civitas*, 1(2), 2016. Retrieved from <https://go.civitaslearning.com/community-insights-issue-2-pdf>
- Coinijn, R., Kleingeld, A., Matzat, U., Snijders, C. & van Zaanen, M. (2016) Influence of course characteristics, student characteristics, and behaviour in learning management systems on student performance. *30th Conference on Neural Information Processing Systems (NIPS 2016)*, Barcelona, Spain. Retrieved from <https://research.tilburguniversity.edu/en/publications/influence-of-course-characteristics-student-characteristics-and-b>
- Conijn, R., Snijders, C., Kleingeld, A., & Matzat, U. (2017). Predicting student performance from LMS data: a comparison of 17 blended courses using Moodle LMS. *IEEE Transactions on Learning Technologies*, 10(1), 17–29. <https://doi.org/10.1109/TLT.2016.2616312>
- Department of Education and Skills (2017) *Data Strategy for the Department of Education and Skills – Data Strategy and Action Plan January 2017*. Dublin: Government Publications. Retrieved from <https://www.education.ie/en/The-Department/Management-Organisation/Data-Strategy-for-The-Department-of-Education-and-Skills-2017-.pdf>
- European Union Commission. (2018). Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on the Digital Education Action Plan. *Office for Official Publications of the European Communities*. Retrieved from <https://ec.europa.eu/education/sites/education/files/digital-education-action-plan.pdf>
- Gašević, D., Dawson, S., Rogers, T., & Gasevic, D. (2016). Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28, 68–84. <https://doi.org/10.1016/j.iheduc.2015.10.002>

- Gray, G., McGuinness, C. & Owende, P. (2014). An application of classification models to predict learner progression in tertiary education. 2014 IEEE International Advance Computing Conference. <https://doi.org/10.1109/IAdCC.2014.6779384>
- Gwet, K.L. (2014) *Handbook of Inter-Rater Reliability 4th Edition*, Gaithersburg: Advanced Analytics, LLC.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques*. Waltham, MA: Kaufman Publishers.
- Hellas, A., Ithantola, P., Petersen, A. Ajanovski, A.A., Gutica, M., Hynninen, T., Knutas, A., Leinonen, J., Messom, C. & Nam Liao, S. (2018). Predicting academic performance: a systematic literature review. In *Proceedings Companion of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE '18 companion)* July 2-4, 2018 Larnaca, Cyprus <https://doi.org/10.1145/3293881.3295783>
- Howard, E., Meehan, M., & Parnell, A. (2018). Contrasting prediction methods for early warning systems at undergraduate level. *The Internet and Higher Education*, 37, 66–75. <https://doi.org/10.1016/j.iheduc.2018.02.001>
- Jo, I.H., Kim, D. & Yoon, M. (2015). Constructing proxy variables to measure adult learners time management strategies in LMS. *Educational Technology and Society*, 18(3), 214-225. Retrieved from <http://www.jstor.org/stable/jeductechsoci.18.3.214>
- Kondo, N., Okubo, M., & Hatanaka, T. (2017). Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data (pp. 198–201). *Presented at the Advanced Applied Informatics (IIAI-AAI), 2017 6th IIAI International Congress on, IEEE*. <https://doi.org/10.1109/IIAI-AAI.2017.51>
- Kuhn, M. (2008). Building Predictive Models in R Using the caret Package. *Journal of Statistical Software*, 28(5). <https://doi.org/10.18637/jss.v028.i05>
- Kuhn, M. (2018). *The caret Package*. Retrieved from <http://topepo.github.io/caret/index.html>
- Landis, J. R., & Koch, G. G. (1977). The Measurement of Observer Agreement for Categorical Data. *Biometrics*, 33(1), 159. <https://doi.org/10.2307/2529310>
- Liaw, A. & Wiener, M. (2002) Classification and regression by randomForest. *R news*, 2(3), 18-22.
- Lu, O.H.T., Huang, A.Y.Q., Huang, J.C., Lin, A.J.Q., Ogata, H. & Yang, S.J.H. (2018) Applying learning analytics for the early prediction of students' academic performance in blended learning. *Educational Technology and Society*, 21(2), 220-232. Retrieved from <http://www.jstor.org/stable/26388400>
- Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, 54(2), 588–599. <https://dx.doi.org/10.1016/j.compedu.2009.09.008>
- Mayer-Schönberger, V., & Cukier, K. (2013). *Big data: a revolution that will transform how we live, work, and think*. Boston: Houghton Mifflin Harcourt.

- Papamitsiou, Z., & Economides, A. A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Journal of Educational Technology & Society*, 17(4), 49.
- R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. <https://www.R-project.org>.
- Raga Jr, R. C., & Raga, J. (2017). Monitoring class activity and predicting student performance using Moodle action log data. *International Journal of Computing Sciences Research*, 1(3), 1–16. <https://doi.org/10.25147/ijcsr.2017.001.1.09>
- Ridgeway, G. (2007) *Generalised Boosted Models: a guide to the GBM package*. Available at <http://www.saedsayad.com/docs/gbm2.pdf>
- Ruckenstein, M. & Dow Schüll, N. (2017). The datafication of health. *Annual Review of Anthropology*, 46, 261-278. <https://doi.org/10.1146/annurev-anthro-102116-041244>
- Sandoval, A., Gonzalez, C., Alarcon, R., Pichara, K., & Montenegro, M. (2018). Centralised student performance prediction in large courses based on low-cost variables in an institutional context. *The Internet and Higher Education*, 37, 76–89. <https://doi.org/10.1016/j.iheduc.2018.02.002>
- Smith, G.J.D., Bennett Moses, L. & Chan, J. (2017). The challenges of doing criminology in the big data era: towards a digital and data-driven approach. *The British Journal of Criminology*, 57(2), 259-274. <https://doi.org/10.1093/bjc/azw096>
- Solas. (2017). *Solas Response to the Findings of the Evaluation of the National Post Leaving Certificate (PLC) Programme*. Retrieved from http://www.solas.ie/SolasPdfLibrary/PLC/SOLAS_response_PLC_evaluation.pdf
- Strobl, C., Boulesteix, A.L., Zeileis, A. & Hothorn, T. (2007). Bias in random forest variable importance measures: illustrations, sources and a solution. *BMC BioInformatics*, 8(1), 25. <https://doi.org/10.1186/1471-2105-8-25>
- Tempelaar, D.T., Rienties, B., & Giesbers, B. (2015). In search for the most informative data for feedback generation: learning analytics in a data rich context. *Computers in Human Behaviour*, 47, 157-167. <https://doi.org/10.1016/j.chb.2014.05.038>
- Tenney, M. & Sieber, R. (2016). Data-driven participation: algorithms, cities, citizens, and corporate control. *Urban Planning*, 1(2), 101-113. <http://dx.doi.org/10.17645/up.v1i2.645>
- Ülker, D., & Yılmaz, Y. (2016). Learning management systems and comparison of open source learning management systems and proprietary learning management systems. *Journal of Systems Integration*, 18–24. <https://doi.org/10.20470/jsi.v7i2.255>
- Viberg, O., Hatakka, M., Balter, O. & Mavroudi, A. (2018). The Current Landscape of Learning Analytics in Higher Education. *Computers in Human Behaviour*, 89, 98-110. <https://doi.org/10.1016/j.chb.2018.07.027>
- Zacharis, N.Z. (2015). A multivariate approach to predicting student outcomes in web enabled blended learning courses. *The Internet and Higher Education*, 27, 44-53. <https://doi.org/10.1016/j.iheduc.2015.05.002>