Taylor & Francis
Taylor & Francis Group

Medical Teacher

ISSN: 0142-159X (Print) 1466-187X (Online) Journal homepage: www.tandfonline.com/journals/imte20

How learning analytics can early predict under-achieving students in a blended medical education course

Mohammed Saqr, Uno Fors & Matti Tedre

MEDICAL TEACHER, 2017
VOL. 39, NO. 7, 757–767
http://dx.doi.org/10.1080/0142159X.2017.1309376

How learning analytics can early predict under-achieving students in a blended medical education course

Mohammed Saqr[a,b], Uno Fors[b] and Matti Tedre[b]*

[a]College of Medicine, Qassim University, Qassim, Kingdom of Saudi Arabia; [b]Department of Computer and System Sciences (DSV), Stockholm University, Kista, Sweden

ABSTRACT

Aim: Learning analytics (LA) is an emerging discipline that aims at analyzing students' online data in order to improve the learning process and optimize learning environments. It has yet un-explored potential in the field of medical education, which can be particularly helpful in the early prediction and identification of under-achieving students. The aim of this study was to identify quantitative markers collected from students' online activities that may correlate with students' final performance and to investigate the possibility of predicting the potential risk of a student failing or dropping out of a course.
Methods: This study included 133 students enrolled in a blended medical course where they were free to use the learning management system at their will. We extracted their online activity data using database queries and Moodle plugins. Data included logins, views, forums, time, formative assessment, and communications at different points of time. Five engagement indicators were also calculated which would reflect self-regulation and engagement. Students who scored below 5% over the passing mark were considered to be potentially at risk of under-achieving.
Results: At the end of the course, we were able to predict the final grade with 63.5% accuracy, and identify 53.9% of at-risk students. Using a binary logistic model improved prediction to 80.8%. Using data recorded until the mid-course, prediction accuracy was 42.3%. The most important predictors were factors reflecting engagement of the students and the consistency of using the online resources.
Conclusions: The analysis of students' online activities in a blended medical education course by means of LA techniques can help early predict underachieving students, and can be used as an early warning sign for timely intervention.

Introduction

E-learning has become an essential part of current healthcare education, and courses delivered or supported by technology are on the rise in size, number, and scale of adoption. (Dahlstrom et al. 2014; Liu et al. 2016). Using technology in education extends through a wide range of applications like e-logistics, e-administration, e-assessment, digital course content, multimedia, simulation, collaboration, communication, and e-support, to name a few. Most of these technologies can be bundled in comprehensive learning management systems (LMSs) (Ellaway & Masters 2008; Liu et al. 2016).

While traditional teaching methods leave little behind to track, modern LMSs generate vast amounts of information about students, their use of the material, and the learning contexts in terms of records, logs, interactions, and other digital footprints (Ferguson 2012; Siemens 2013). The availability of these datasets, increased computer power, skills learnt in business analytics, the pressure toward better teaching and learning, personalization of the content, and improving LMSs has led to increased interest in learning analytics (LA) development and research (Brown 2011; Ferguson 2012; Dahlstrom et al. 2014; Papamitsiou & Economides 2014).

LA is an emerging, relatively new, and rapidly developing discipline (Conde & Hernández-García 2015; Rienties et al. 2016) that aims at "measurement, collection, analysis and reporting of these data and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs" (Siemens 2013). Analytics have two main functions: to provide information about the current status of the learners and their learning process or provide an insight about what is yet to happen on the individual or group level in the future (Ellaway et al. 2014).

Analyzing data collected from learners' interactions within LMSs and information systems has been the common approach to LA and the one that has proven most promising (Ramos & Yudko 2008; Macfadyen & Dawson 2012; Lockyer et al. 2013; Agudo-Peregrina et al. 2014; Gaševi et al. 2015). "Course Signals" by Purdue University is

Practice points

- Learning analytics (LA) is an emerging field that uses students' online activities to learn about their online behavior for the sake of improving their outcome and optimizing learning.
- Analyzing online activity can highlight active and inactive students, which can be used as an alert to educators and academic supervisors.
- LA can be used to early predict grades and performance.
- The application of LA might help early intervention that has the potential to decrease dropout rate.

CONTACT Mohammed Saqr saqr@qumed.edu.sa College of Medicine, Qassim University, PO Box: 6655, 51452 Qassim, Kingdom of Saudi Arabia
*School of Computing, University of Eastern Finland, PO Box 111, Joensuu, Finland.

an early example of building a warning and feedback system for students and teachers using LA principles. It has been reported to have a positive effect on student retention and teachers' awareness of their students (Pistilli & Arnold 2010).

LA has been shown to enable effective, automatic tracking of students' engagement along the course (Macfadyen & Dawson 2010, 2012; Wolff et al. 2013; Cruz-Benito et al. 2015; Tempelaar et al. 2015; Gašević et al. 2016). The insights generated by LA can be shared by course teachers, academic supervisors, and administrators (Arnold & Pistilli 2012; Howard et al. 2016; Rienties et al. 2016). Those insights could help identify students at risk of under-achievement where an early intervention can lead to a meaningful change (Macfadyen & Dawson 2010; Lockyer et al. 2013; Tempelaar et al. 2015; Gašević et al. 2016; Howard et al. 2016; Rienties et al. 2016). Although traditional assessment methods offer this kind of feedback, their results most often come too late for a possible action or a significant intervention (Macfadyen & Dawson 2010; Cruz-Benito et al. 2015).

Education in the healthcare sector is under a lot of pressure to respond efficiently and timely to the rapidly changing scientific, societal, and social environment, as well as to keep programs modern and connected to the communities it serves (Ellaway & Masters 2008; Ellaway et al. 2014; Vaitsis et al. 2014; Liu et al. 2016). Another challenge facing medical schools is underachieving and potential attrition; an issue that may be a symptom of preventable problems in the medical education (in selection of students, curricula, teaching methods, assessment or policies) (O'Neill et al. 2011). While the problem of attrition in medical education seems to incur a substantial cost, it is still poorly studied and most of the published studies have focused on students' attributes at the point of admission (O'Neill et al. 2011; Stegers-Jager et al. 2015), which only explained 30% of variance in performance, recent research indicates that LA can significantly improve the predictability of academic performance and hence can help solve the problem (Macfadyen & Dawson 2010; Agudo-Peregrina et al. 2014; Papamitsiou & Economides 2014; Wolff et al. 2014; Tempelaar et al. 2015).

Although the benefits of using LA in education have been conceptually justified (Richards 2011; Ellaway et al. 2014; Gaševi et al. 2015; Tempelaar et al. 2015; Rienties et al. 2016) and the need was recently recognized in the medical education literature (Doherty et al. 2014; Ellaway et al. 2014), research in medical setting is rare. Based on results from other fields, we expect that LA can have a considerable a considerable, yet unexplored potential for healthcare education. Ellaway et al. (2014) summarized this need as follows "health professional educators will need to be ready to deal with the complex and compelling dynamics of analytics and Big Data. We therefore need to explore, discuss, and critique these emerging techniques to develop a robust understanding of their strengths and limitations in the service of health professional education".

This research study was performed to analyze data of students' online activity in a blended medical education course in Saudi Arabia in order to identify quantitative markers that correlate with students' performance and might be used as early warning signs for possible data driven measures.

The research questions of this study are:

1. Which tracking variables best correlate with student performance?
2. To what extent can the analysis of students' online activities be used to predict student grades, and identify the potential risk of a student failing or dropping a course?

Methodology

The analysis of student data followed a standard procedure used in data mining research and analytics (Dean 2014; Gandomi & Haider 2014; Wolff et al. 2014):

- *Acquisition and recording*: Acquiring the data from different sources.
- *Preparing the data*: Matching and cleaning mislabeled data, excluding incomplete records and appropriate annotation of data types, combining the data into one master table.
- *Performing exploratory data analysis (EDA):* Exploring data by testing the relationships between different variables to discover possible relationships, patterns, rules that could help identify the potential predictors. EDA does not require a prior hypothesis in contrast to formalistic scientific methodology that tests a previously known theory (Velleman & Hoaglin 2012).
- *Building the predictive model*: Predicting students' outcome and identifying at-risk students using appropriate predictive models. We used regression models, as they are among the most common predictive models used in education research at large (Peng et al. 2002), and in analytics research (Ramos & Yudko 2008; Macfadyen & Dawson 2010; Gašević et al. 2016; Howard et al. 2016), available in most statistical packages, and can be evaluated in several ways (Peng et al. 2002; Bewick et al. 2005). Two types of regression models according to the type of outcome to be predicted:
    - *Automatic linear modeling* (ALM) for grade prediction, ALM uses a group of predicting factors to predict a single scaled outcome. It offers improvements over traditional methods in two main areas. First, automatic variable selection, which is useful when the number of variables is high (Filippou et al. 2015). The second is automatic data preparation, data preparation is a popular concept in data science that includes re-classifying continuous variables with less than 5 unique values as ordinal and re-classifying ordinal values with more than 10 unique values as continuous variables. It normalizes *outliers* or extreme values (predictors that lie beyond 3 SDs), so that they do not to exert an exaggerated influence on the model. And finally, it does a supervised merge of similar predictors (Yang 2013).
    - *Binary logistic regression* (BLR) for prediction of at-risk students: BLR is a powerful model for the prediction of dichotomous outcomes like pass/fail or at-risk/safe. It overcomes some of the restrictive assumptions of linear regressions like linearity, normality and equal variances. The test has a large array of tests to evaluate its performance (the -2 log likelihood, Cox & Snell R^2, the Omnibus Tests of Model Coefficients,

Hosmer and Lemeshow goodness of fit) (Peng et al. 2002; Bewick et al. 2005).
- *Evaluation of predictive accuracy:* Receiver operating characteristic (ROC) plots the sensitivity (true positive rate) of each model versus "1-specificity" or (false positive rate). The area under the curve (AUC) is a quantification of model accuracy, where 0.5 means a worthless model and 1.0 represents a perfect model (Bewick et al. 2005; Gönen 2006).

Data collection (acquisition and recording)

This study was preceded by a pilot study to determine feasibility, and to identify engagement parameters and the tracking variables (Alghasham et al. 2013). This study was performed on the students of the course "Man and environment" during 2013–2014 at Qassim University, Kingdom of Saudi Arabia. This is the second course in the medical program, and the first to teach the medical subject after an introductory course. The intention of studying students in the first year was to be able to capture the full spectrum of freshly admitted students before some underachievers may drop out. The study included 133 students enrolled in a blended course where they were free to use the LMS at their will. There was no incentive or punishment of using the LMS apart from students' self-perceived benefit.

At Qassim College of Medicine, the Moodle LMS is used as the main platform for learning management, Moodle produces robust logs of students' activities; however, the available reporting tools are deficient, and Moodle does not have built-in analytic tools (Falakmasir & Habibi 2010). Therefore, we used MySQL database queries (SQL) and five add-on tools.

First, Attendance Register module was used to report total time spent by a student in a course (Moodle plugins directory: Attendance Register). Second, Configurable Reports was used to run SQL queries to generate custom reports about students' activities like number of course views, forum posts or reads, and course edits (Moodle plugins directory: Configurable Reports). Third, Analytics Graphs was used to calculate total unique days of course access, total number of course views, and total number of course views) (Moodle plugins directory: Analytics Graphs). Fourth, Mailchimp e-mail tracking was used to track students' response to e-mails, and the frequency of opening course related e-mails (About Open Tracking |MailChimp.com: KB Article). Finally, NodeXL, was used to calculate betweenness centrality (Smith et al. 2009).

Collected data

Collected data were divided into six categories and detailed in Table 1.

Engagement sub-scores

Based on the findings of the pilot study (Alghasham et al. 2013), five engagement indicators were identified. They reflect regularity of using the LMS and balance for the fact that some students would be highly active over a short time and then go into periods of inactivity (Richards 2011; Shea et al. 2013; Cruz-Benito et al. 2015; Panzarasa et al. 2016). The indicators were calculated as follows:

By login: a student was considered engaged in a certain week when he/she logs in 3 days or more in that week, that student is then given a score of one. The login engagement sub-indicator is the sum of scores of the 6 weeks. *By course views*: a score of one is given to the student when he/she views the course materials more than a -1 *Z*-score of mean course views. The views engagement sub-indicator is the sum of scores of the 6 weeks. *By forum posts*: a student is given the score of 1 when he/she posts two or more posts in a certain week. The posting engagement sub-indicator is the sum of the 6 weeks.

By time: a student is given the score of 1 when he spends more than a -1 *Z*-score of course average time of all students. The time engagement sub-indicator is the sum of the 6 weeks. *By formative assessment*: a score of 1 is given to a student if he/she tried an assessment regardless of the score. The formative engagement sub-indicator is the sum of the 6 weeks.

At-risk students

There are several standard setting methods that might be used to set the criteria for not passing a course or being a borderline student (Tekian & Norcini 2015). The choice of the method relies largely on the course and purpose of the standard setting. For this study, we followed the procedure described by Norcini (2003): a panel of expert judges were formed to set a cut point that separates students who barely pass the course (was at-risk) from students who clearly pass (Macfadyen & Dawson 2010). The panel estimated that 5% over the passing mark would define this cut point; accordingly, students were classified into two main categories:

- Potentially safe (coded as "Safe"): Have a final score of 65% or more.
- Potentially at-risk (coded as "At-risk"): Have a final score below 65%.

Research ethics

The study was approved by the Medical Research Ethics Committee of Qassim College of Medicine. All users of Qassim College of Medicine LMS sign an online privacy policy that detail possible use of data for research and user protection guarantees (Qassim College of Medicine). All data were anonymized, personal identifying information were masked, no private data or personal information were used in the analysis or published. College Privacy guidelines and policies of dealing with students' data were strictly followed (Qassim College of Medicine 2014). It is also important to mention here that using e-learning portal is neither graded nor mandatory; only depending on student self-interest, and that specific course did not contain any online-graded assignments.

Participants

The study initially included 145 students (44 females and 101 males). However, 12 were excluded due to incomplete data and delayed enrollment in the LMS; thus, the final number was 133 (43 females and 90 males) over the period of 6 weeks.

Results

The study was performed using an EDA approach, testing all possible parameters to try to identify the significant ones (Macfadyen & Dawson 2010; Velleman & Hoaglin 2012). Studies in similar environments have been rare and the previous studies have not been able to find a general set of metrics that can be used as predictors of student achievement (Macfadyen & Dawson 2010; Agudo-Peregrina et al. 2014; Tempelaar et al. 2015; Wise & Shaffer 2015; Rienties et al. 2016). A previous study in a medical education course would have been helpful (Gašević et al. 2016), but unfortunately to the best of our knowledge, we could not identify a study with same scope of metrics and comparable design as this actual project.

The study identified 64 possible tracking variables in six main categories, we report here the most important and significant indicators at mid-course, second half, and end of course. First, correlation coefficient was calculated to identify the possible indicators, followed by the prediction of student grade, then we try to predict the at-risk students at the end of course and whether this is possible at mid-course or not.

Correlations

In Table 2, the findings related to correlations are displayed and the most interesting findings were the following.

Content creation/interaction: There was a positive and significant correlation between the students' final grade and interaction/content creation variables. The most important were *total edits* or *created content* $r(131) = 0.31$, $p < 0.01$, number of edits in the first half of course $r(131) = 0.3$, $p < 0.01$, total posts initiated by student $r(131) = 0.29$, $p < 0.01$ and total posts and replies $(131) = 0.29$, $p < 0.01$.

Views/hits: The correlation between views and final grade was weak in most parameters, except for the *number of resources accessed* which showed moderate correlation $r(131) = 0.32$, $p < 0.01$, followed by the *total hits on resources* $r(131) = 0.25$, $p < 0.01$.

Login frequency: The *frequency of logins* had the strongest correlation with final grade $r(131) = 0.47$, $p < 0.01$. It was moderately correlated with the number of logins in the first half of course $r(131) = 0.31$, $p < 0.01$ and the number of logins in the second half of course $r(131) = 0.36$, $p < 0.01$.

Time: The *total time* showed a weak correlation with final grade $r(131) = 0.22$, $p = 0.01$ and the other parameters of time showed similar low correlation.

Formative assessment: The *formative assessment* parameters were the most consistent and most of them showed moderate correlation with the final grade. The number of the exams attempted was the highest $r(131) = 0.46$, $p < 0.01$, followed by the grade $r(131) = 0.43$, $p < 0.01$.

Communications: There was no significant correlation between *following course communications* and final grade.

Engagement sub-indicators showed more consistent and higher correlation coefficients with final grades than simple *counts of activities*; especially, for the parameters that measured students' consistency of using the LMS resources (see Table 3). The highest was total unique days of access $r(131) = 0.46$, $p < 0.01$, followed by the formative assessment sub-indicator $r(131) = 0.45$ login sub-indicator $r(131) = 0.41$, $p < 0.01$, $r = 0.41$ and 0.46 (Table 3).

Table 1. Detailed description of collected data.

Parameter	Collected data
Logins	• Weekly, mid-course, and total course logins • Logins before and after the end of the course (pre and post-term) • Total number of days with course access
Views (hits)	• Daily number of views, weekly, mid-course • Total course views, number of unique resources accessed and types of the accessed resources
Forums	• Weekly, at mid-course and the overall total of number created posts • Reads, replies and total number of edits made in course (creation of new content) • Betweenness centrality was calculated by NodeXL to get an idea of how influential the participation of a student was in the forms, we calculated
Time	• Weekly, at mid-course and the overall total time spent online using educational materials, time navigating profile or viewing non-educational materials was excluded
Formative assessment	• Grades of each formative assessment and participation in assessment regardless of the submission of the answers

Table 2. Bivariate correlations between LMS tracking variables and final grade.

Parameter	Correlation (r)	Sig. (two-tailed)
Interaction and content creation		
Total posts initiated by a student	0.29**	<0.01
Total posts and replies by a student	0.29**	<0.01
Number of posts in the first half of course	0.27**	<0.01
Number of posts in the second half of course	0.23**	<0.01
Total edits of content created	0.31**	<0.01
Number of edits in the first half of course	0.30**	<0.01
Number of edits in the second half of course	0.22*	<0.01
Hits and views		
Total times forums were read	0.25**	<0.01
Total hits on recourses	0.28**	<0.01
Total Hits on course information	0.21*	0.02
Total course hits	0.26**	<0.01
Total views before course started	0.037	0.672
Number of hits in the first half of course	0.24**	<0.01
Number of hits in the second half of course	0.24**	<0.01
Number of resources accessed	0.32**	<0.01
Logins and course access		
Total logins	0.47**	<0.01
Number of logins in the first half of course	0.31**	<0.01
Number of logins in the second half of course	0.36**	<0.01
Time		
Time spent online	0.22*	0.01
Total time spent online first half of course	0.23**	<0.01
Time spent in the second half of course	0.18*	0.04
Formative assessment		
Formative assessment grade	0.43**	<0.01
Mid-course formative assessment grade	0.42**	<0.01
Number attempted the formative quiz	0.46**	<0.01
End of course formative assessment grade	0.25**	<0.01
Communications		
Response to communications	0.05	0.57
Total views of news forum	−0.068	0.44

*Correlation is significant at 0.05 level (2-tailed); **Correlation is significant at 0.01 level (2-tailed).

The prediction of student grade

Students' online behavior involves a number of factors and therefore a simple univariate correlation cannot simply predict the final grade; for example, spending more time online does not usually translate into higher achievement unless that time is spent learning, participating in activities and interacting (Macfadyen & Dawson 2010; Shea et al. 2013; Ballard & Butler 2016). Therefore, it would be erroneous to rely on just the previous correlations to draw conclusions about students' achievement.

Since univariate correlation and simple linear regression have limitations (inability to identify and handle outliers, prone to type I/II errors) (Yang 2013) and there was no standard tracking variables that could be used as a guide for this study (Macfadyen & Dawson 2010). We sought a predictive model that can combine all potential tracking variables of students' online activity. We found ALM – a relatively new method introduced in SPSS 19 – to be the most suitable for the analysis of this study (Yang 2013). The accuracy (adjusted R^2) of ALM for predicting the final grade was 63.5%. Figure 1 presents a scatterplot of the relationship between the actual and predicted grade.

Figure 2 shows how much each factor contributed to the prediction model. Total predictor importance is 1.0, and the most important and statistically significant predictors were login engagement sub-indicator (0.16), number of views of course information (0.11), formative assessment sub-indicator (0.10), and posting engagement sub-indicator (0.10).

Table 3. Bivariate correlations between final grade and engagement sub-indicators.

Parameters of engagement	Pearson's correlation	Sig. (two-tailed)
Login sub-indicator	0.41	<0.01**
Formative assessment sub-indicator	0.45	<0.01**
Posting sub-indicator	0.31	<0.01**
Time sub-indicator	0.32	<0.01**
Course views sub-indicator	0.28	<0.01**
Total unique days with access	0.46	<0.01**

**Correlation is significant at 0.01 level (2-tailed).

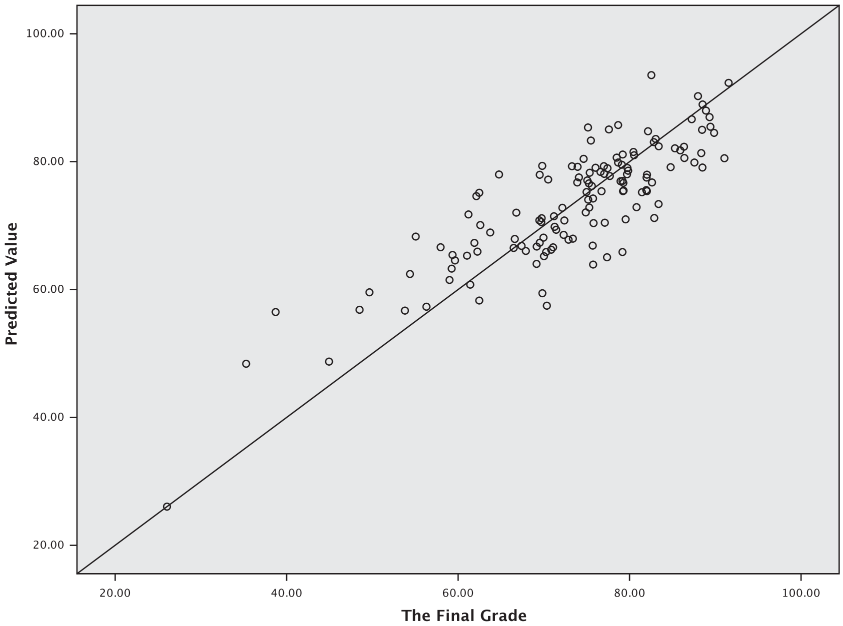

Figure 1. A scatterplot summarizing the relationship between actual and predicted grade.

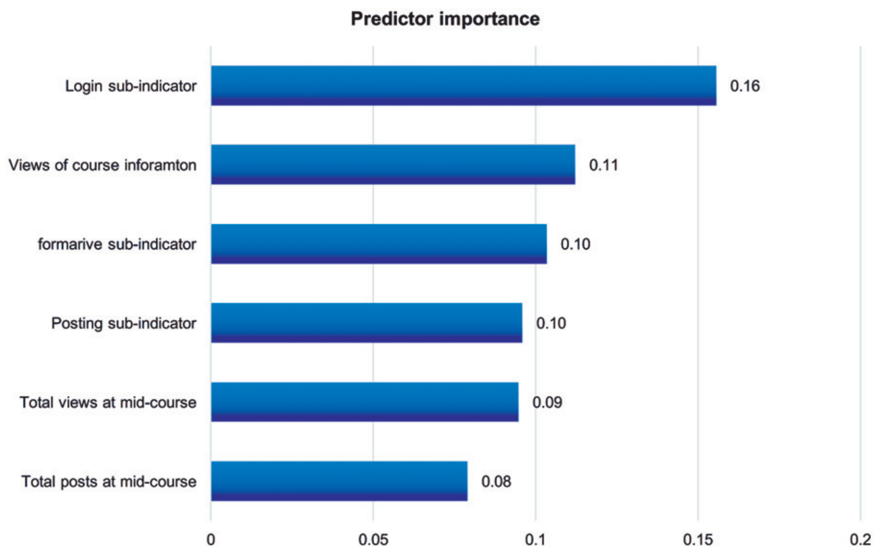

Figure 2. Visualization of predictor importance according to automatic linear modeling.

Predicting at-risk students

Using results of ALM to classify students according to safety level, the model identified 18 students as at-risk (14 true positive, 4 false positives and missed 12), the sensitivity for picking at-risk students was 53.85% (CI: 33.37%–73.41%), and the specificity was 96.26% (CI: 90.70%–98.97%)). A χ^2 test of independence was performed to examine the relationship between actual and predicted at-risk students. ALM was more likely to correctly identify at-risk students in 53.8% of cases, $\chi^2(1,\ N=133)=44.9$, $p<0.01$; detailed results are cross tabulated in Table 4.

To further illustrate the efficiency of ALM in classifying students as at-risk or safe, in Figure 3, every circle represents a student, and while circles to the left side were at-risk students, the green-colored circles are students correctly predicted to be at-risk, and on the right, side the green circles are false positive.

Binary logistic regression

BLR successfully classified 21 out of 26 as at-risk students and misclassified two students. The sensitivity of BLR was 80.8% (CI: 60.7%–93.5%) and the specificity was 96.3% (CI: 93.4%–99.8%). A χ^2 test of independence was performed to examine the relationship between actual and predicted at-risk students. BLR was more likely to correctly identify at-risk students in 80.8% of cases, $\chi^2(1,\ N=133)=91$, $p<0.01$.

The -2 log likelihood was 43.01, Cox & Snell R^2 was 0.49, and Nagelkerke R^2 was 0.77. These findings are strong indicators that a large portion of the variation of the final grade can be explained by the selected predictors. The Omnibus Tests of Model Coefficients was statistically significant ($\chi^2=88.35$, df $=39$, $p<0.01$) and Hosmer and Lemeshow goodness of fit was non-significant ($\chi^2=13.95$ df $=8$, $p=0.08$). Both results are further indications of the adequacy of fitness of the binary logistic model (Table 5).

To what extent can tracking data at mid-course predict at-risk students?

Using BLR to test the possibility of early predicting at-risk students, we were able to detect 42.3% of at-risk students correctly (CI: 23.35%–63.08%) (Table 6; Figure 4).

The -2 log likelihood was 89.05, Cox & Snell R^2 was 0.27, and Nagelkerke R^2 was 0.43 indicating that a significant part of the variation of the final grade could be explained by the mid-course predictors. The Omnibus Tests of Model Coefficients was significant ($\chi^2=42.37$, df $=19$, $p=0.002$) and Hosmer and Lemeshow goodness of fit was non-significant ($\chi^2=5.29$ df $=8$, $p=0.73$) indicating that the logistic model adequately fits the data.

Table 4. Cross-tabulation of predicted and at-risk students using ALM.

	Actual		
Predicted	At-risk	Safe	Percentage correct
At-risk	**14**	4	53.8%
Safe	12	**103**	96.3%
Total	26	107	133

Bold numbers denote correctly identified students at-risk.

Table 5. Cross-tabulation of predicted and at-risk students using the binary logistic model with the cutoff of 0.50.

	Actual		
Predicted	At-risk	Safe	Percentage correct
At-risk	**21**	2	80.8%
Safe	5	**105**	98.1%
Total	26	107	133

Bold numbers denote correctly identified.

Table 6. Cross-tabulation of predicted and at-risk students using mid-course data.

	Actual		
Predicted	At-risk	Safe	Percentage correct
At-risk	**11**	4	42.3%
Safe	15	**103**	96.3%
Total	26	107	133

Bold numbers denote correctly identified.

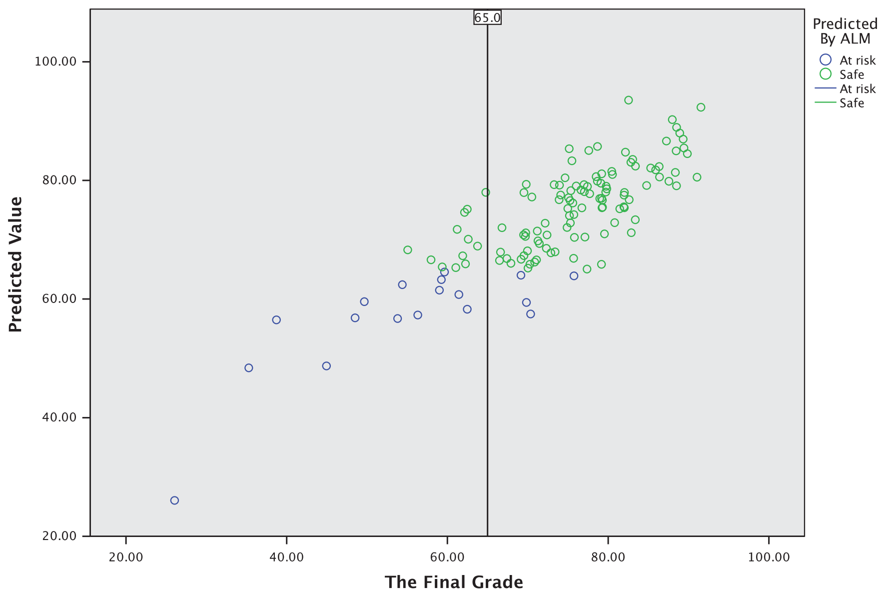

Figure 3. Scatterplot of predicted against actual grade, green circles to left of the red dotted line represent correctly identified at-risk students (true positives) using ALM.

Comparing predictive models

BLR using at the end-of-course data was the most sensitive and had excellent AUC value of 0.9, followed by linear regression, and BLR at mid-course (Figure 5). Table 7 presents a detailed comparison between the three methods and ROC area. Table 8 contains all the students who were correctly or in-correctly identified by each model along with the engagement indicators.

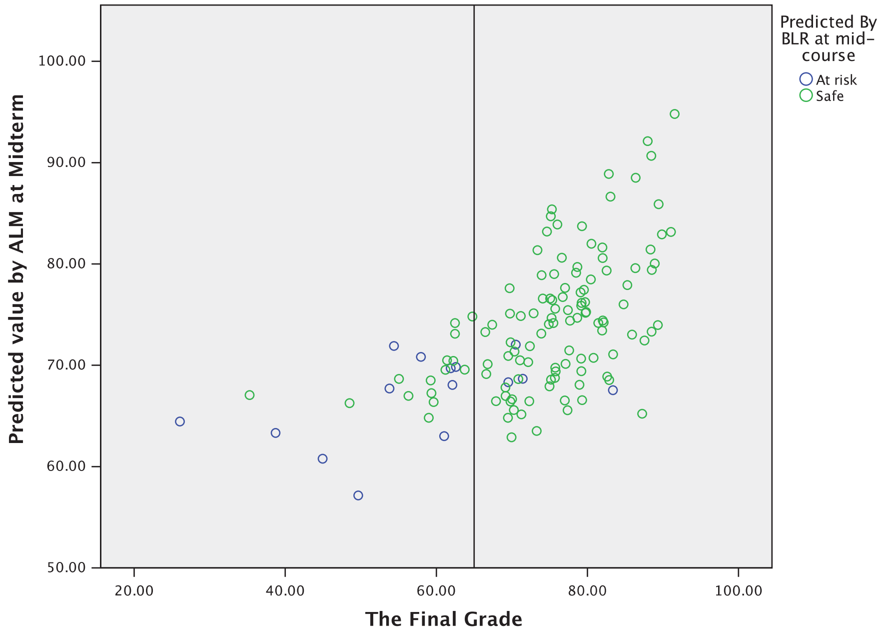

Figure 4. Scatterplot of predicted against actual grade at mid-term, blue circles to left of the vertical line represent correctly identified at-risk students (true positives).

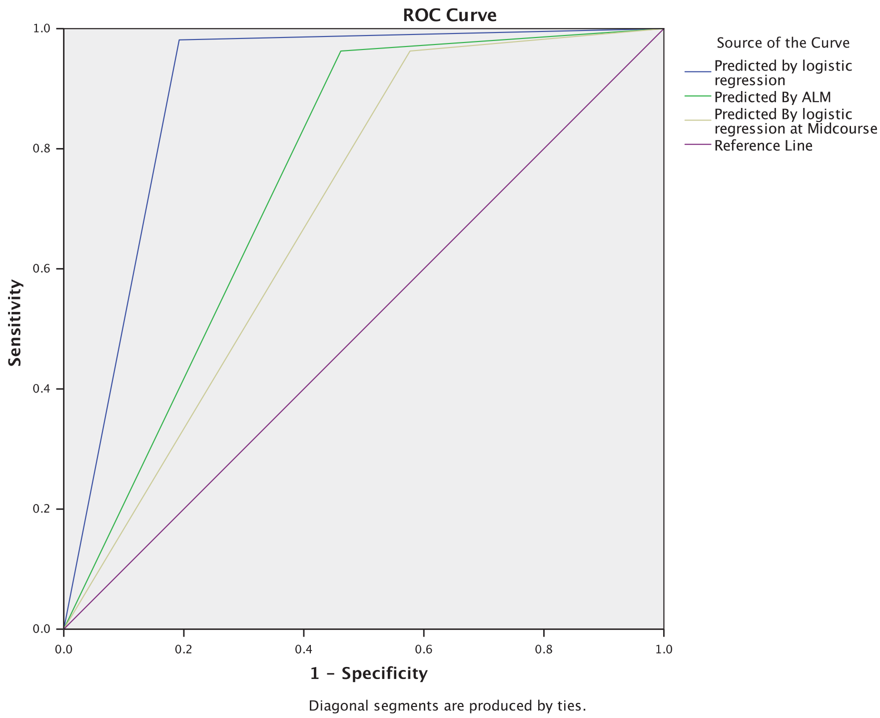

Figure 5. Receiver operating characteristic (ROC) comparing the predictive models used in the study, it shows that all three models have values more than 0.5 and that binary logistic regression was the most sensitive.

Table 7. Comparison between different predictive models AUC and sensitivities in prediction of at-risk students.

Model	Sensitivity	AUC	*p*	95% Confidence interval
Binary logistic regression	0.81	0.90	0.00	0.8–0.99
Automatic linear model	0.54	0.75	0.00	0.63–0.88
Mid-course Binary logistic regression	0.42	0.69	0.00	0.56–0.82

Table 8. All the students who were correctly or in-correctly (bold) identified by each model along with the engagement indicators.

	Regularity						Identified by regression models		
Serial	Course views	Formative assessment	Login	Posting	Time	Actual	ALM	BLR	BLR-mid-course
S1	0	0	6	2	3	Safe	**At-risk**	Safe	Safe
S2	3	0	6	5	6	Safe	**At-risk**	Safe	Safe
S3	3	2	6	5	6	Safe	**At-risk**	Safe	Safe
S4	2	2	3	5	5	Safe	**At-risk**	Safe	Safe
S5	0	0	6	3	5	Safe	Safe	**At-risk**	Safe
S6	3	5	6	6	6	Safe	Safe	Safe	**At-risk**
S7	0	3	6	0	2	Safe	Safe	Safe	**At-risk**
S8	3	0	6	3	6	Safe	Safe	Safe	**At-risk**
S9	2	3	5	3	6	Safe	Safe	**At-risk**	**At-risk**
S10	0	2	6	0	6	At-risk	At-risk	Safe	**Safe**
S11	2	0	5	2	6	At-risk	At-risk	At-risk	**Safe**
S12	2	0	5	2	3	At-risk	At-risk	At-risk	**Safe**
S13	2	0	5	5	6	At-risk	At-risk	At-risk	**Safe**
S14	0	0	5	0	2	At-risk	At-risk	At-risk	**Safe**
S15	2	0	6	5	6	At-risk	At-risk	At-risk	**Safe**
S16	2	3	6	2	6	At-risk	At-risk	At-risk	**Safe**
S17	0	3	3	0	2	At-risk	At-risk	At-risk	**Safe**
S18	0	0	6	2	3	At-risk	At-risk	At-risk	At-risk
S19	0	3	5	5	6	At-risk	At-risk	At-risk	At-risk
S20	0	0	5	2	2	At-risk	At-risk	At-risk	At-risk
S21	0	0	3	0	3	At-risk	At-risk	At-risk	At-risk
S22	0	0	2	0	2	At-risk	At-risk	At-risk	At-risk
S23	3	2	6	5	6	At-risk	At-risk	At-risk	At-risk
S24	3	3	6	3	6	At-risk	**Safe**	**Safe**	**Safe**
S25	2	2	6	5	3	At-risk	**Safe**	**Safe**	**Safe**
S26	0	0	6	3	6	At-risk	**Safe**	**Safe**	**Safe**
S27	5	0	6	5	6	At-risk	**Safe**	At-risk	**Safe**
S28	3	2	6	5	6	At-risk	**Safe**	At-risk	**Safe**
S29	0	5	6	0	2	At-risk	**Safe**	At-risk	**Safe**
S30	5	0	6	5	6	At-risk	**Safe**	At-risk	**Safe**
S31	2	2	6	3	6	At-risk	**Safe**	**Safe**	At-risk
S32	3	2	6	2	6	At-risk	**Safe**	At-risk	At-risk
S33	0	2	5	0	6	At-risk	**Safe**	At-risk	At-risk
S34	5	2	6	5	6	At-risk	**Safe**	At-risk	At-risk
S35	2	2	5	5	6	At-risk	**Safe**	At-risk	At-risk

Discussion

In this study, we investigated the variables that best correlate with students' performance and can be used to identify students who might be at-risk of under-achievement for possible timely intervention. We have expanded over the previous studies (Macfadyen and Dawson 2010; Wolff et al. 2014; Tempelaar et al. 2015; Gašević et al. 2016) and included multidimensional data about access, hits, time, forums, communications, and social network parameters, as well as formative assessments across different points in time.

We have calculated parameters (engagement sub-indicators) that reflect consistency of using the LMS resources and self-motivation. The concept was borrowed from marketing and brand loyalty that is long recognized by marketers (Ballard & Butler 2016; Panzarasa et al. 2016). The engagement sub-indicators also normalize the extremes of use and temporary surges by some students, that adds much to their overall counts and do not reflect a consistent practice throughout the course. The indicators also serve as an indirect measure of self-regulation in the course, since e-learning is offered in a blended scenario and other activities are going beyond the LMS (Shea et al. 2013; Cruz-Benito et al. 2015).

Our results indicated that the engagement indicators showed consistent and significantly higher correlations with the students' performance across all categories of measurement. In contrast to the simple generic metrics, which showed inconsistent and relatively weaker correlations with students' performance. Parameters such as time and hits (the most generic metrics) were the weakest, and parameters that reflected motivation and disposition such as taking the optional formative assessments, frequency of access, and content creation were the best indicators of students' performance.

The shortcomings of using generic parameters in analytics research have been previously recognized (Macfadyen & Dawson 2010) and were further emphasized by recent findings in large-scale studies (Conde & Hernández-García 2015; Rienties et al. 2016). They have produced conflicting and inconsistent results from one study to the other, and from course to course (Ramos & Yudko 2008; Tempelaar et al. 2015; Gašević et al. 2016), especially in blended scenarios (Agudo-Peregrina et al. 2014). They offer limited insights to the understanding of the complex learning environments or to the development of theory for LA (Conde & Hernández-García 2015; Gaševi et al. 2015). Reliance on generic parameters has been further criticized for having a detrimental role that has hampered the advance of LA as a field (Conde & Hernández-García 2015).

No study has established a standard set of variables that fits all online courses, and studies comparing multiple courses have actually indicated the opposite: variables and indicators differ between course learning designs (Wolff et al. 2013; Agudo-Peregrina et al. 2014; Gašević et al. 2016; Rienties et al. 2016). That highlights the need to find new sets of indicators that are contextually relevant to the course design and the learning environment (Pistilli & Arnold 2010; Gašević et al. 2016). In this study, we proposed a set of indicators (engagement sub-indicators) that reflect engagement and self-regulation (Richards 2011; Shea et al. 2013; Cruz-

Benito et al. 2015). They are also independent of the extremes of clicking behavior, and in line with the course design and previous findings of a pilot study in a similar course (Alghasham et al. 2013; Lockyer et al. 2013).

Using variables identified in the first step of the study, we were able to predict the final grade by means of automatic linear regression model with 63.5% accuracy, and identify 53.9% of at-risk students at the end of the course. The prediction improved when a binary logistic model was used, which was able to accurately classify 80.8% of the at-risk students. Using data recorded up till the mid-course, prediction accuracy was 42.3%; in the three predictive models the specificity was above 96%. The most important predictors were factors reflecting engagement of the students and the consistency of using the online resources rather than the number of clicks.

In this study, we opted for using only tracking variables collected from students' use of the LMS and the calculated engagement parameters. While adding other variables to the predictive model (variables such as previous performance or sociodemographic data) might have added to the power of prediction (Agudo-Peregrina et al. 2014; Tempelaar et al. 2015; Rienties et al. 2016), these variables are definitely not modifiable and cannot be used to design an intervention for underachieving students (Tempelaar et al. 2015), and would simply defeat the purpose of analytics as being actionable (Conde & Hernández-García 2015; Gašević et al. 2016).

Engagement has been shown to be an important factor for the adoption of learning technology (Cruz-Benito et al. 2015; Ballard & Butler 2016). Aspects of engagement, like involvement in the learning process, time spent on a task and compliance have been shown to positively correlate with effective learning and positive outcome (Shea et al. 2013; Cruz-Benito et al. 2015; Tempelaar et al. 2015; Ballard & Butler 2016), and there is a large body of research confirming that efforts at increasing student engagement would help students who need support (Cruz-Benito et al. 2015; Tempelaar et al. 2015; Ballard & Butler 2016). Since engagement has different dimensions or aspects; therefore, there are different ways to measure engagement (Cruz-Benito et al. 2015; Ballard & Butler 2016). In this study, we have explored the potential of LA to measure some of these aspects such as compliance, involvement, and quality effort in purposeful activities. The strength of this study is that it proposes a technique to spot disengaged students who can be helped. In contrast to teachers' intuition, LA is automatic, effortless, samples a large number of indicators, and offer a quantifiable risk index (Macfadyen & Dawson 2010; Brown 2011; Siemens 2013; Ellaway et al. 2014; Gaševi et al. 2015; Tempelaar et al. 2015).

Comparing students to their peers in the same course has proven to be the most effective way to build an accurate predictive model (Pistilli & Arnold 2010; Wolff et al. 2013, 2014; Gašević et al. 2016) and most predictive models were applied to individual courses (Macfadyen & Dawson 2010; Pistilli & Arnold 2010; Alghasham et al. 2013; Tempelaar et al. 2015; Howard et al. 2016), mainly, because each course is structurally different, uses distinct LMS feature and incur a different load on learners (Wolff et al. 2013; Gašević et al. 2016). Corroborating evidence came from studies investigating the use of a single predictive model across different courses, which found significant variations throughout different courses (Finnegan et al. 2008; Gašević et al. 2016). Finnegan et al. (2008) could not find a single predictor that was shared among all investigated courses and Gašević et al. (2016) found that the same predictors vary, even within the same discipline and advised against using same predictive model (one size fits all) for multipe courses.

This study is a step in a long road that will define the field of LA, we hope that our work will open the door for others to explore the possible potential of LA, build on and critique our approach. We recommend medical schools to adopt analytics capable LMSs, train staff to follow students through LA dashboards, discover relevant metrics, and prediction models tailored to their educational context.

This study is not without limitations, being exploratory in nature like most LA studies is our most important limitation, although we have tried to link our findings to theories of engagement, there is still a long road ahead to replicate these findings and build generalizable approaches. Another limitation comes from the methodology, we used different and diverse tools to collect data from various sources, since Moodle LMS does not have a built-in analytics dashboard that automates the data collection, it might need special skills to replicate these results. However, there are emerging tools and dashboards that have started to address the problem and offer these capabilities to everyone.

Conclusions

This research study was set out to identify quantitative markers that correlate with students' performance and can be used to identify potential risk of a student failing or dropping a course. We collected data from students' use of the LMS about access, hits, time, forums, communications, social network parameters, as well as formative assessments across different points in time. We also calculated engagement indicators that would reflect self-motivation and consistency of interest in using the LMS resources.

The parameters of engagement showed significant positive correlations with students' performance, especially the parameters that reflected motivation and self-regulation such as trying formative assessments, frequency of logging, and creation of new content. We were able to predict the final grade with 63.5% accuracy, and identify 53.9% of at-risk students at the end of the course. Using a binary logistic model improved prediction to 80.8%. Using data recorded up till the mid-course, prediction accuracy was 42.3%, and the most important predictors were factors reflecting engagement of the students and the consistency of using the online resources.

This study demonstrated that a significant number of at-risk medical students can be early identified, and may benefit from positive intervention that addresses modifiable factors such as engagement. The study findings represent an additional method to foresee students' performance before formal final examinations and might be a tool for early alert to underachievers.

Disclosure statement

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of this article.

Notes on contributors

Mohammed Saqr is assistant professor of medicine with interest in learning enhanced technology, learning analytics, and simulation of medicine, he is currently the head of e-learning unit, Qassim University, College of Medicine.

Matti Tedre is a professor at the University of Eastern Finland, a docent at Stockholm University, Sweden, and the author of Science of Computing: Shaping a Discipline (Taylor & Francis/CRC Press, 2014).

Uno Fors DDS, PhD, is professor of IT and learning as well as head of Department of Computer and Systems Sciences at Stockholm University, Sweden. Fors research focuses on Technology enhanced learning and especially on virtual cases for learning within the healthcare area. Fors has published more than 150 papers in scientific journals and conferences.

Glossary

Learning analytics: Measurement, collection, analysis, and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs.

Siemens G. Learning analytics: The Emergence of a Discipline.

American Behavioral Scientist. 2013;57:1380–1400.

Social network analysis: Social network analysis is the study of structure, it deals with the relational structure and patterns of relationships among social entities, which might be people, groups, or organizations.

Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health [Internet]. 2004 Dec [cited 2017 Feb 1];58(12):971–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15547054

Outlier: An outlier is an observation which deviates so much from the other observations and might exert an exaggerated effect on the results (deviate more than 3 times the SD from the mean).

Hawkins D. 1980. Identification of Outliers: Chapman and Hall.

Betweenness centrality: Is a measure of the influence a node has over the spread of information through the network.

Hawe P, Webster C, Shiell A. A glossary of terms for navigating the field of social network analysis. J Epidemiol Community Health [Internet]. 2004 Dec [cited 2017 Feb 1];58(12):971–5. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15547054

References

About Open Tracking|MailChimp.com: KB Article. Available from: http://kb.mailchimp.com/reports/about-open-tracking
Agudo-Peregrina AF, Iglesias-Pradas S, Conde-Gonzalez MA, Hernandez-Garcia A. 2014. Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning. Comput Human Behav. 31:542–550.
Alghasham A, Saqr M, Kamal H. 2013. Using learning analytics to evaluate the efficacy of blended learning in pbl based medical course. AMEE 2013 Conference at Prague Congress Centre; 2013 Aug 24–28; Prague, Czech Republic.
Arnold KE, Pistilli MD. 2012. Course signals at purdue. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge – LAK; 2012 29 April–2 May; Vancouver, British Columbia, Canada.
Ballard J, Butler PI. 2016. Learner enhanced technology. J Appl Rese HE. 8:18–43.
Bewick V, Cheek L, Ball J. 2005. Statistics review 14: logistic regression. Crit Care. 9:112–118.
Brown M. 2011. Learning analytics: the coming third wave. EDUCAUSE Learning Initiative Brief . [cited 2016 Jan 24]. Available from: https://net.educause.edu/ir/library/pdf/elib1101.pdf
Conde MÁ, Hernández-García Á. 2015. Learning analytics for educational decision making. Comput Human Behav. 47:1–3.
Cruz-Benito J, Therón R, García-Peñalvo FJ, Pizarro Lucas E. 2015. Discovering usage behaviors and engagement in an Educational Virtual World. Comput Human Behav. 47:18–25.
Dahlstrom E, Brooks DC, Bichsel J. 2014. The current ecosystem of learning management systems in higher education: student, faculty, and IT perspectives. 27.
Dean J. 2014. Big data, data mining, and machine learning: value creation for business leaders and practitioners. Hoboken (NJ): John Wiley & Sons; p. 55–70.
Doherty I, Sharma N, Harbutt D. 2014. Contemporary and future eLearning trends in medical education. Med Teach. 37:1–3.
Ellaway R, Masters K. 2008. AMEE Guide 32: e-Learning in medical education Part 1: learning, teaching and assessment. Med Teach. 30:455–473.
Ellaway RH, Pusic MV, Galbraith RM, Cameron T. 2014. Developing the role of big data and analytics in health professional education. Med Teach. 36:216–222.
Falakmasir MH, Habibi J. 2010. Using educational data mining methods to study the impact of virtual classroom in e-learning. The 3rd International Conference on Educational Data Mining; Pittsburgh, PA, USA.
Ferguson R. 2012. Learning analytics: drivers, developments and challenges. IJTEL. 4:304.
Filippou J, Cheong C, Cheong F. 2015. Designing persuasive systems to influence learning: modelling the impact of study habits on academic performance. Pacific Asia Conference on Information Systems (PACIS); Singapore.
Finnegan C, Morris LV, Lee K. 2008. Differences by course discipline on student behavior, persistence, and achievement in online courses of undergraduate general education. J Coll Stud Retent Res Theory Prac. 10:39–54.
Gandomi A, Haider M. 2014. Beyond the hype: big data concepts, methods, and analytics. Int J Inform Manage. 35:137–144.
Gaševi D, Dawson S, Siemens G. 2015. Let's not forget: learning analytics are about learning course signals: lessons learned. Tech Trends 59:64–71.
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. Internet High Educ. 28:68–84.
Gönen M. 2006. Receiver operating characteristic (ROC) curves. SAS Users Group International (SUGI). 31:210–231.
Howard E, Meehan M, Parnell A. 2016. Developing accurate early warning systems via data analytics [Journal]. Available from: http://arxiv.org/abs/1612.05735
Liu Q, Peng W, Zhang F, Hu R, Li Y, Yan W. 2016. The effectiveness of blended learning in health professions: systematic review and meta-analysis. J Med Internet Res. 18:e2.
Lockyer L, Heathcote E, Dawson S. 2013. Informing pedagogical action: aligning learning analytics with learning design. Am Behav Scient. 57:1439–1459.
Macfadyen LP, Dawson S. 2010. Mining LMS data to develop an "early warning system" for educators: a proof of concept. Comput Educ. 54:588–599.
Macfadyen LP, Dawson S. 2012. Numbers are not enough. Why e-learning analytics failed to inform an institutional strategic plan. Educat Technol Soc. 15:149–163.
Moodle plugins directory: Analytics Graphs [Internet]. [cited 2014 Feb 2]. Available from: https://moodle.org/plugins/block_analytics_graphs
Moodle plugins directory: Attendance Register [Internet]. [cited 2014 Feb 2]. Available from: https://moodle.org/plugins/mod_attendanceregister
Moodle plugins directory: Configurable Reports [Internet]. [cited 2014 Feb 2]. Available from: https://moodle.org/plugins/block_configurable_reports
Norcini JJ. 2003. Setting standards on educational tests. Med Educ. 37:464–469.
O'Neill LD, Wallstedt B, Eika B, Hartvigsen J. 2011. Factors associated with dropout in medical education: a literature review. Med Educ. 45:440–454.

Panzarasa P, Kujawski B, Hammond EJ, Roberts CM. 2016. Temporal patterns and dynamics of e-learning usage in medical education. Educ Tech Res Dev. 64:13–35.
Papamitsiou ZK, Economides AA. 2014. Learning analytics and educational data mining in practice: a systematic literature review of empirical evidence. Educat Technol Soc. 17:49–64.
Peng C-yJ, Lee KL, Ingersoll GM. 2002. An introduction to logistic regression analysis and reporting. J Educat Res. 2:3–14.
Pistilli MD, Arnold KE. 2010. In practice: purdue signals: mining real-time academic data to enhance student success. About Campus. 15:22–24.
Qassim College of Medicine. 2014. Qassim College of Medicine Privacy Policy and User Agreement. 3rd ed. [Internet]. [cited 2014 June 2]. Available from: http://qumed.org/code.htm
Ramos C, Yudko E. 2008. "Hits" (not "Discussion Posts") predict student success in online courses: a double cross-validation study. Comput Educ. 50:1174–1182.
Richards G. 2011. Measuring engagement: learning analytics in online learning. Electr Kazan 2011; 2011 Apr 19; Kazan, Russian Federation.
Rienties B, Boroowa A, Cross S, Kubiak C, Mayles K, Murphy S. 2016. Analytics4Action evaluation framework: a review of evidence-based learning analytics interventions at the Open University UK. J Interact Media Educ. 2016:11.
Shea P, Hayes S, Smith SU, Vickers J, Bidjerano T, Gozza-Cohen M, Jian S-B, Pickett A, Wilde J, Tseng C-H. 2013. Online learner self-regulation: learning presence viewed through quantitative content-and social network analysis. IRRODL. 14:427–461.
Siemens G. 2013. Learning analytics: the emergence of a discipline. Am Behav Scientist. 57:1380–1400.
Smith MA, Shneiderman B, Milic-Frayling N, Mendes Rodrigues E, Barash V, Dunne C, Capone T, Perer A, Gleave E. Analyzing (social media) networks with NodeXL. Proceedings of the fourth international conference on Communities and technologies; 2009: ACM.
Stegers-Jager KM, Themmen APN, Cohen-Schotanus J, Steyerberg EW. 2015. Predicting performance: relative importance of students' background and past performance. Med Educ. 49:933–945.
Tekian A, Norcini J. 2015. Overcome the 60% passing score and improve the quality of assessment. GMS Zeitschrift für Medizinische Ausbildung. 32:Doc43.
Tempelaar DT, Rienties B, Giesbers B. 2015. In search for the most informative data for feedback generation: learning analytics in a data-rich context. Comput Human Behav. 47:157–167.
Vaitsis C, Nilsson G, Zary N. 2014. Big data in medical informatics: improving education through visual analytics. Stud Health Technol Inform. 205:1163–1167.
Velleman PF, Hoaglin DC. 2012. Exploratory data analysis. In: Cooper H, Camic PM, Long DL, Panter AT, Rindskopf D, Sher KJ, editors. APA handbook of research methods in psychology, Vol. 3: data analysis and research publication. Washington (DC): American Psychological Association. p. 51–70.
Wise AF, Shaffer DW. 2015. Why theory matters more than ever in the age of big data. JLA. 2:5–13.
Wolff A, Zdrahal Z, Herrmannova D, Kuzilek J, Hlosta M. 2014. Developing predictive models for early detection of at-risk students on distance learning modules; 2014 Mar 24–28; The 4th International Conference on Learning Analytics and Knowledge (LAK14); Indianapolis, Indiana, USA.
Wolff A, Zdrahal Z, Nikolov A, Pantucek M. 2013. Improving retention: predicting at-risk students by analysing clicking behaviour in a virtual learning environment. Proceedings of the Third International Conference on Learning Analytics and Knowledge; Leuven, Belgium. New York: ACM Press.
Yang H. 2013. The case for being automatic: introducing the automatic linear modeling (LINEAR) procedure in SPSS statistics. Multiple Linear Regression Viewpoints. 39:27–37.