





Review

Systematic Literature Review of Predictive Analysis Tools in Higher Education [†]

Martín Liz-Domínguez *, Manuel Caeiro-Rodríguez , Martín Llamas-Nistal 
and Fernando A. Mikic-Fonte 

Telematic Systems Engineering Group,atlanTTic Research Center, University of Vigo, Campus Universitario Lagoas-Marcosende, 36310 Vigo, Spain; mcaeiro@gist.uvigo.es (M.C.-R.); martin@gist.uvigo.es (M.L.-N.); mikic@gist.uvigo.es (F.A.M.-F.)

* Correspondence: mliz@gist.uvigo.es

† This paper is an extended version of our paper published in Learning Analytics Summer Institute (LASI) Spain 2019.

Received: 21 November 2019; Accepted: 13 December 2019; Published: 17 December 2019



Abstract: The topic of predictive algorithms is often regarded among the most relevant fields of study within the data analytics discipline. They have applications in multiple contexts, education being an important one of them. Focusing on higher education scenarios, most notably universities, predictive analysis techniques are present in studies that estimate academic outcomes using different kinds of student-related data. Furthermore, predictive algorithms are the basis of tools such as early warning systems (EWS): applications able to foresee future risks, such as the likelihood of students failing or dropping out of a course, and alert of such risks so that corrective measures can be taken. The purpose of this literature review is to provide an overview of the current state of research activity regarding predictive analytics in higher education, highlighting the most relevant instances of predictors and EWS that have been used in practice. The PRISMA guidelines for systematic literature reviews were followed in this study. The document search process yielded 1382 results, out of which 26 applications were selected as relevant examples of predictors and EWS, each of them defined by the contexts where they were applied and the data that they used. However, one common shortcoming is that they are usually applied in limited scenarios, such as a single course, evidencing that building a predictive application able to work well under different teaching and learning methodologies is an arduous task.

Keywords: predictive analytics; early warning systems; learning analytics; learning technologies

1. Introduction

Data analytics encompasses the collection of techniques that are used to examine data of various types to reveal hidden patterns, unknown correlations and, in general, obtain new knowledge [1]. This discipline has a very strong presence among recent trends in information and communication technologies, being of utmost relevance for researchers and industry practitioners alike. Data analytics is often coupled with the term “big data”, since analysis tasks are often performed over huge datasets. Other fields of study which are very popular nowadays, such as data mining or machine learning, are close to data analytics and share many relevant techniques.

Depending on the nature of the data that are being analyzed and the objective that the analysis task should fulfill, several sub-disciplines can be defined under data analytics. Examples of these are text analytics, audio analytics, video analytics and social media analytics. One of the most relevant, and the main focus in this paper, is *predictive analytics*, which includes the variety of techniques used to make predictions of future outcomes relying on historical and present data [2].

The ability to predict future events is essential for the proper functioning of some applications. Notable examples among these are early warning systems (EWS), which are capable of anticipating potential risks in the future thanks to information available in the present, accordingly sending alerts to the person or group of people who may be affected by these risks and/or that are capable of countering them. Their degree of reliability on information technologies greatly varies depending on the context they are applied on.

EWS are mostly known for their use to reduce the impact of natural disasters, such as earthquakes, floods and hurricanes. Upon detection of signs that such a catastrophe might happen in the near future, members of the potentially affected population are alerted and given instructions to prevent or minimize damage [3]. However, other kinds of EWS have been implemented in many different contexts. For instance, they are used in financial environments to predict economic downturns at an early stage and provide better opportunities to mitigate their negative effects [4]. In healthcare, early warning systems are used by hospital care teams to recognize the early signs of clinical deterioration, enabling the initiation of early intervention and management [5].

This document revolves around the application of predictive algorithms and EWS in educational contexts, focusing on higher education scenarios, most notably university courses. This falls under the umbrella of learning analytics (LA), a particularization of data analytics which is usually defined as “the 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” [6].

Predictors and EWS are used in higher education contexts with the general objective of supporting learning and mitigating some of the most important problems that are observed in these scenarios, such as student underperformance or dropout. Many authors have presented their own solutions in this field of study, with vast differences in terms of specific scenarios where they are applied, input data they work with and problems that they address. With such a vast collection of different approaches, it can be a difficult task to fully understand the current status of solutions in this field, as well as identifying which of them have been applied in real education scenarios with satisfactory results.

The literature review that is presented in this study has the purpose of providing an answer to the following two questions:

- *RQ1: What are the most important purposes of using predictive algorithms in higher education, and how are they implemented?*
- *RQ2: Which are the most notable examples of early warning systems applied in higher education scenarios?*

This paper is based on our publication at the conference Learning Analytics Summer Institute (LASI) Spain 2019 [7]. As an improvement upon the previous article, the review has been restructured in order to better comply with the PRISMA statement for systematic literature reviews [8]. The search and selection process have been more thoroughly explained and justified. Additionally, results are presented in a more comprehensive way, focusing on comparing key functionalities of different tools.

After the present introduction, the methodology is reported, explaining the literature search process and the criteria that were applied to assess the relevance of analyzed documents. Next, the contents of the most relevant papers are summarized, addressing the research questions proposed above. At the end of this document, some insights and discussion are presented.

2. Methodology

As aforementioned, the PRISMA model is used to properly report the search and selection processes that were performed in this systematic literature review. While initially conceived for the development of literature reviews and meta-analyses in medical areas, the general structure of PRISMA can be followed in reviews belonging to other fields of knowledge.

This section explains the document search and selection tasks step by step, providing reasons to justify all of the relevant decisions that were made in the process.

2.1. Search Procedure

The search process was performed with the purpose of finding relevant scientific papers that present applications of predictors and EWS in higher education scenarios.

The search was limited to the following types of documents: journal articles, conference proceedings, and book extracts. The election of these document types was made under the assumption that they would be likely to present a good variety of unique approaches to build and test predictive applications, thus allowing for more interesting comparisons. Additionally, to focus on relatively new applications and technologies, no documents with a publication date older than 2012 were considered, as this year marked the point when learning analytics as a whole really started to bloom [9]. Regarding language, only documents written in English were eligible.

The following online databases were used in the document retrieval process: IEEE Xplore Digital Library, ACM Digital Library, Elsevier (ScienceDirect), Wiley Online Library, Springer (SpringerLink), Emerald, and Taylor & Francis. These were selected according to the reason of them being the most important online scientific libraries that we had full access to. Additionally, queries were run on Google Scholar, Scopus, and Web of Science in order to complement the search results from the aforementioned libraries.

The document search process started in March 2019. The latest search-related tasks were performed in June of the same year, therefore, papers published after this date were not contemplated in this review. The following search procedure was applied:

1. early warning system
2. predictive analy*
3. predictive algorithm
4. 1 OR 2 OR 3
5. education
6. university
7. 4 AND 5 AND 6
8. disaster
9. medical
10. health
11. 7 AND NOT 8 AND NOT 9 AND NOT 10

The reasoning behind this strategy was to obtain documents that focus on higher education contexts while also filtering out papers that deal with the use of EWS in the medical or natural disaster prevention fields, since these are areas where predictive algorithms and EWS are very prominent, and thus have a significant amount of related literature.

Table 1 shows the number of results that were retrieved in each of the online libraries that were searched.

Table 1. Search results per database.

Source	Results
IEEE Xplore Digital Library	36
ACM Digital Library	45
Elsevier (ScienceDirect)	412
Wiley Online Library	255
Springer (SpringerLink)	91
Emerald	36
Taylor & Francis	165
Google Scholar	~13,800
Scopus	956
Web of Science	97

It is worth noting that the search on Springer was limited to the “Education” discipline. Additionally, the lists of results obtained from both Google Scholar and Scopus were ordered by relevance using the built-in sorting options that these searchers provide, and only the first 200 results in each case were considered. This was done because documents further down the list of results were not observed to be related to the educational field.

Together with the search results obtained from the online libraries, the total number of retrieved documents was 1537. However, some of the results that were obtained using Google Scholar, Scopus, and Web of Science were duplicates of documents already yielded by searches on the other library databases. Disregarding these duplicates, the final number of documents that were retrieved in the search process was 1382.

2.2. Selection Process

To narrow down the result list to only the most relevant documents, a selection strategy was defined. This strategy consisted in the application of two document filtering tasks: screening and election.

The screening process had the purpose of filtering out papers that were not related with the topic at hand. This was performed by scanning the title, abstract, and keywords of the search results. Out of the initial 1382 papers, 1315 were eliminated during the screening process because of the following reasons:

- The document is not related to the educational field (1118 papers).
- The document is related to education, but does not present the use of a predictive algorithm or EWS (111 papers).
- The document presents the use of a predictive algorithm or EWS in education, but in a context other than higher education (66 papers, most of them centered around massive open online courses (MOOCs)).

The remaining 67 articles were fully and independently analyzed during the election process, looking for these particular pieces of information:

- Date on which the document was published.
- Problem that the predictor or EWS seeks to solve.
- The prediction goal of the algorithm or application (such as student grades or dropout likelihood).
- Types of data used as input.
- Technical aspects about the predictive algorithm or algorithms that the application uses.
- User collective that received the output information (most commonly either students or teachers).
- How the output information was presented to the end user.
- Specific higher education context where the tool was applied.
- Number of students that were involved in the study.
- Reliability of predictions made by the predictor or EWS.
- Evaluation of the application’s impact over the students.
- Any other unique aspect that differentiates the particular predictor or EWS from the rest.

In the election process, documents that failed to provide important information about how the EWS was built or applied in practice were ruled out. This resulted in the elimination of 41 papers for the following reasons:

- The document was missing important information about the predictor or EWS’s implementation, and inner workings (24 papers).
- The results of applying the predictor or EWS in real higher education scenarios were insufficient, nonexistent, or poorly detailed (17 papers).

Once the election process was completed, the final number of papers that were included in this review for qualitative synthesis was 26.

Figure 1 displays a graphical summary of the document selection process, showing the outcome of every stage.

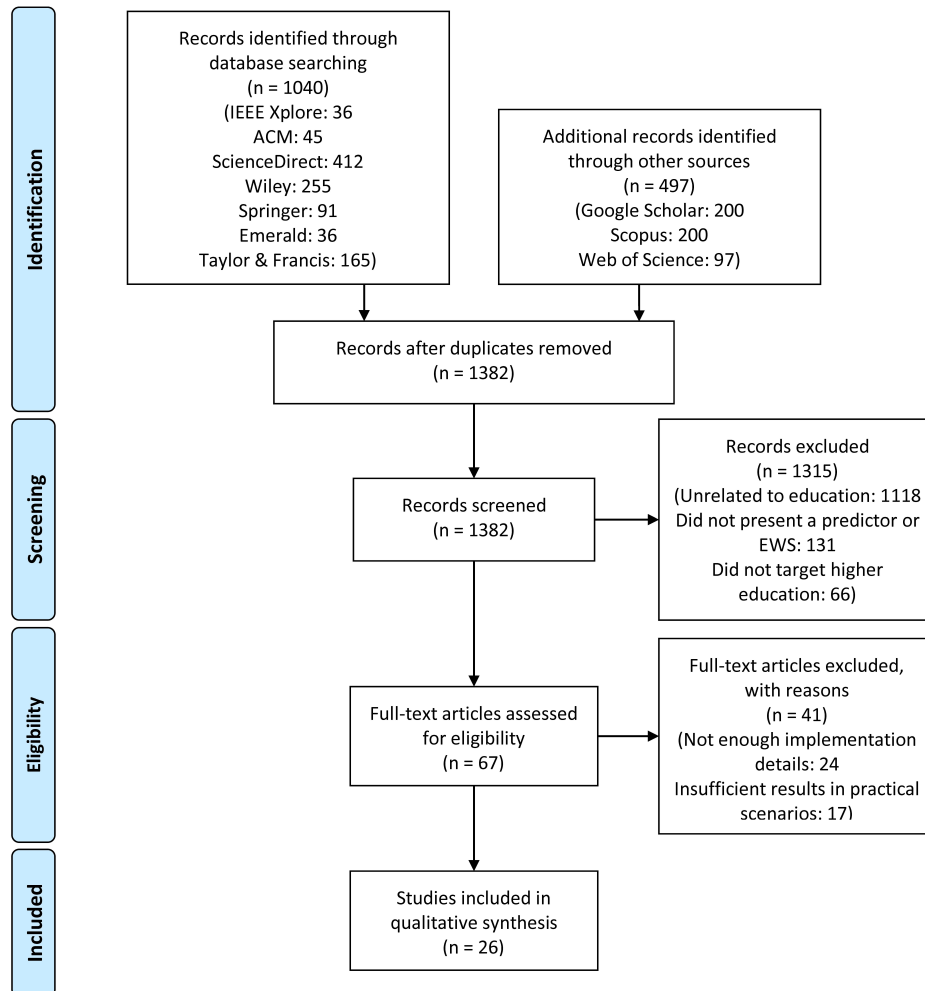


Figure 1. PRISMA flow diagram summarizing the results of the search and selection processes.

3. Results

This section covers the most relevant aspects of the selected studies, comparing them with each other: type of application developed, prediction goal, data used as input, algorithms used for analysis, and scenarios where they were applied. Appendix A contains a short summary for each paper, organized as a table.

First, the selected papers were classified depending on whether they present a *predictor* or an *early warning system*. For the sake of this review, a *predictor* in an educational context is defined as an application that, given a specific set of input data, aims to anticipate the outcome of a course or degree, normally in terms of either grades or a passing/failure classification. An *early warning system* performs the same tasks as a predictor, but, on top of that, it reports its findings to a teacher and/or students at an early enough stage so that measures can be taken to avoid or mitigate potentially negative outcomes. This means that EWS often have tighter timing requirements than predictors, as they need to perform analyses early in a course. Additionally, EWS must present analysis results in such a way that teachers and/or students are able to easily comprehend them. For this reason, EWS make use of tools for visual representation, such as dashboards, in order to deliver information to the end user.

Among the 26 selected documents, there was an even split regarding their classification as predictors or EWS: 13 studies of each type. Table A1 in Appendix A includes the category that was assigned to each one of the articles.

To make establishing comparisons easier, predictors and EWS are described in their own subsections up next.

3.1. Predictors

Predictors in education may target a specific course or an educational program or degree as a whole. In either case, two main types of prediction goals can be identified: final grade of a student, or whether the student will succeed, fail, or drop out. A few prominent examples of predictors in educational environments are described next.

Ornelas and Ordonez developed a Naïve Bayesian classifier that was applied in 13 different courses at Rio Salado Community College (Tempe, AZ, USA) [10]. The courses belonged to degree programs in the fields of science and humanities. The study used data from the institution's LMS, related to both student engagement (determined by LMS logins and participation in online activities) and student performance (meaning the points earned in course tasks). The goal of the classifier was to predict student success, defined as obtaining a grade of C or higher in a course. For most of the courses in which this tool was tested, the classifier managed to achieve an accuracy of over 90%. However, there were three scientific courses for which accuracy dropped to values between 80% and 90%. According to the authors, this could be explained by the differences in complexity compared to other courses. This experiment was conducted with a big student population, with a training sample of 5936 students and a validation sample of 2722. The dataset was also fairly balanced, with failure and success rates of 40% and 60% respectively.

Thompson et al. created a classifier based on logistic regression, targeting an introductory biology course taught during the first semester of a university major [11]. Similar to the previous example, this classifier had the goal of predicting whether the student would pass the course or not. However, instead of using data collected throughout the course, this study was based on results from tests with no direct relationship with the course itself. These were taken right at the start of the semester with the purpose of evaluating the students' scientific reasoning and mathematical abilities. More specifically, the tests were Lawson's Classroom Test of Scientific Reasoning and the ACT Mathematics Test. The predictor was tested with a population of 413 students, showing that the scores of the tests were significant predictors of student success, with a p -value smaller than 0.05 for both of them. The fact that a prediction of success can be made before the course even starts is certainly interesting, and it lays the groundwork for a possible concept of an EWS. However, it was acknowledged that academic ability, as measured by these two tests, is not necessarily what defines the odds of success, with factors such as motivation and engagement often playing a more important role.

Benablo et al. presented a classifier for student success that used data related to student procrastination as input [12]. To obtain these data, students were surveyed regarding the time they spent using social networks and playing online games. Three different classification algorithms were tested: support vector machines (SVM), k-nearest neighbors (KNN), and random forest (RF), using 10-fold validation to evaluate the performance of each of them. The predictor was tested with a cohort of 100 computer science students from a university in the Philippines. The SVM classifier performed better than the ones based on KNN and RF. SVM registered an F-measure of 0.984, while KNN and RF could only reach 0.806. Nevertheless, it must be taken into account that a population of 100 students can be too small to properly evaluate a predictor. In this case, all three classification algorithms yielded a precision of 100%, a benchmark that is not expected to be reached in a bigger scenario.

Umer et al. attempted to estimate the earliest possible time at which a reliable prediction of the students' final performance in a course could be made [13]. This study involved 99 students enrolled in a 16-week-long introductory mathematics course taught at an Australian university. Moreover,

this course implemented the continuous assessment system: on top of a final exam, students needed to complete five different assignments throughout the course, being due on Weeks 2, 4, 8, 10, and 12. The predictor collected data related to student engagement from Moodle logs—such as views of course modules and submission of tasks—as well as the grades that students obtained in the assignments. Final grades were predicted using a discrete scale (marks ranging from A to D, as well as low failure and dropout), and four different classifier algorithms were tested: KNN, RF, Naïve Bayes, and linear discriminant analysis (LDA). The performance of this predictor was evaluated in terms of how well it could classify students into high and low performers, the former corresponding to final grades of C or better and the latter including all other grades. RF was the best performing algorithm, yielding an accuracy of 70% after just one week and without any assignment grade data, and 87% after two assignments had been completed. However, as in the previous example, the small population may hurt the reliability of this performance evaluation.

Kostopoulos et al. built a co-training based student success predictor with the aim of achieving greater performance than traditional classifier algorithms [14]. Co-training is an analysis technique that consists in dividing the input data features into two independent and sufficient views, leading to the use of two predictive models and taking advantage of redundancy to improve prediction results. This study used data regarding student gender, attendance, and grades in the first view, while indicators regarding LMS activity formed the second view. The system was implemented in an introductory informatics module taught at a Greek open university, with 1073 enrolled students. Its performance was exhaustively tested, using different ratios of labeled-to-unlabeled data, as well as several combinations of classifiers for each view, including KNN, Extra Tree, RF, Gradient Boosting Classifier (GBC), and Naïve Bayes. Overall, it was observed that using Extra Tree for the first view and GBC for the second one yielded the best performance. More importantly, tests that used co-training performed better than self-trained variants—without feature splitting—as this study aimed to prove.

Hirose used item-response theory (IRT) to make estimations of students' abilities regarding the contents of a course [15]. This study targeted introductory calculus and algebra courses at a Japanese higher education institution. These courses followed a continuous assessment system in which students needed to perform a multiple-choice type exam each week. IRT allowed assessing the question difficulty together with students' abilities, resulting in a more fair judgment. Additionally, at several points in time during the course, this study used a KNN classifier in order to predict whether students would pass or fail the course. Results from the weekly tests were used as input data, representing the trends of estimated students' abilities. The predictor was tested with a population of around 1100 students. After seven weeks—midway through the course—the classifier achieved a misclassification rate as low as 18%. However, the author pointed out that the false positive rate was noticeably high, meaning that a significant portion of well-performing students were being classified as at-risk.

Schuck tried to establish a correlation between student success and the level of violence and crime in the university campus and surrounding areas, in the context of higher education in the USA [16]. This study used data collected from a data analysis tool maintained by the US Department of Education, and included rates of violent crimes, disciplinary measures, and arrests among students. Complementary data on graduation rates and other institutional variables were pulled from a data system associated with the National Center for Education Statistics. Overall, data from 1281 higher education institutions were collected. The goal of the predictor was determining graduation rate, that is the fraction of students that would finish their degree program in the intended number of years. The underlying prediction algorithm was multivariate regression. As a result of this study, it was observed that institutions with higher rates of violent crime reported lower graduation rates. On the other hand, institutions that made more referrals to the student conduct system for minor offenses reported higher graduation rates, the same as those with a low number of arrests. This suggests that referrals are more constructive intervention methods than arrests. It is important to note that,

as opposed to the other examples of predictors listed here, this one did not target individual students, but rather university communities as a whole.

Tsiakmaki et al. investigated whether final grades of subjects in a specific semester could be predicted using results from courses in the previous one [17]. This study was performed in the Technological Educational Institution of Western Greece, targeting 592 first-year Business Administration students who started their studies between 2013 and 2017. Predictions were performed at the end of the first semester. Input data included student gender, final grades obtained by the student in first semester subjects on a 1–10 scale, and the number of unsuccessful attempts to pass each subject in previous semesters, if any. Several prediction algorithms were tested to estimate each student's final grade in second semester subjects, also on a scale from 1 to 10. These algorithms were linear regression (LR), RF, instance-based regression, M5 algorithm, SVM, Gaussian processes, and bootstrap aggregating. Ten-fold validation was used for performance assessment. In this study, RF outperformed all other algorithms, achieving a mean absolute error between 1.217 and 1.943 points, depending on the predicted subject.

A study by Adekitan and Salau had the objective of predicting the grade point average (GPA) of a student at the end of a five-year-long degree program, using as input the GPA obtained in each of the first three years [18]. This concept resembles the work done by Tsiakmaki et al.: grades from already completed courses were used to predict future performance. The study was carried out at Covenant University in Nigeria, using grade data from 1841 students belonging to seven different engineering programs. Many prediction algorithms were tested and compared in this study, divided into two categories: classifiers and regression models. For classifiers, the final GPA as a prediction goal was discretized and turned into a four-level scale. The tested algorithms were probabilistic neural network, RF, decision tree, Naïve Bayes, tree ensemble, and logistic regression. The best performing algorithm, achieving an accuracy of 89.15%, was logistic regression. On the other hand, linear and pure quadratic models were used to estimate the final GPA as a numeric value. These algorithms achieved R^2 coefficients of 0.955 and 0.957, respectively.

Jovanovic et al. designed a predictive model to be used in courses following the flipped classroom teaching method [19]. This predictor aimed to estimate each student's grade in the final exam of the course. To do this, it used data regarding interactions with pre-class learning activities. These activities consisted of videos and documents with multiple choice questions, as well as sequences of problems. The input data included indicators on the regularity of student interactions with learning material and their level of engagement with each resource type. These indicators could be either generic or course-specific. Generic indicators represented information not directly related to any course, such as the frequency of LMS logins. On the other hand, course-specific indicators were related to interactions with resources belonging to an individual course. Multiple linear regression was used as the prediction algorithm. This predictor was tested in a first-year engineering course at an Australian institution, involving 1147 students across three different academic years. The study concluded that course-specific indicators had more predictive power than generic ones. The greatest R^2 value achieved was 0.377.

Trussel and Burke-Smalley focused on studying the influence of demographic and socioeconomic factors over performance of undergraduate students [20]. The study was aimed at providing predictions of student success, which could potentially be used to enable early interventions targeting students at risk, supporting the decision making of instructors and advisors in the process. Two different prediction goals were set for this predictor, both indicative of student success. On the one hand, the students' cumulative GPA for the entire degree program was estimated using stepwise ordinary least squares (OLS) regression. On the other hand, logistic regression was applied to determine the probability that a student graduates within six years of entering university, which the authors word as the student being "retained". As input data, the predictive model received the gender and ethnicity of students as demographic indicators, and their household income and whether they are financially independent or not as socioeconomic indicators. Additionally, data regarding their high school grades and their status as full-time or part-time students were also provided. The predictive models were

tested over a population of 1919 undergraduate students enrolled in a business program at a public university in Tennessee (USA). The OLS regression model used for predicting cumulative GPA yielded an adjusted R^2 value of 0.287, and determined that high school GPA was by far the variable with the highest impact over final GPA in the degree program. As for the logistic regression model, grades were once again the most important factor, and the logistic regression model was able to classify students into “retained” and “not retained” categories with an accuracy of 82%. In both cases, some of the demographic and economic factors were statistically significant (such as gender, ethnicity, or status as financially independent), although their impact over prediction outcomes was significantly lower than the one associated to grades.

Chen studied the impact of the quality and quantity of students’ note-taking over their academic performance [21]. This study was done in the context of a first-year general psychology course at a Taiwanese university, involving 38 students. Both in-class and after-class notes were collected and copied by the professor at the end of each lecture. Note quantity was assessed in terms of number of Chinese characters, while the professor rated the quality of notes based on their accuracy and completeness regarding the actual contents of each lecture. These data were processed using hierarchical regression in order to estimate final test scores. As a result of this experiment, it was determined that the quality of notes was a significant predictor of test scores. However, quantity of notes was not related to performance in any meaningful way. An R^2 value of 0.3 was obtained in this study.

Amirkhan and Kofman investigated the effects of stress overload—defined as the destructive form of stress—over the GPA obtained by a student in one semester [22]. To assess the students’ level of stress, they were asked to fill a survey halfway through the semester, including questions regarding perceived burden of demands and insufficiencies in resources. This allowed stress to be quantified with the help of a “stress overload scale”, defined by the authors. Stress scores, alongside student demographic characteristics, were the input data fed to the system. The predictor had two main tasks: first, proving a relationship between stress overload and academic failure using structural equation modeling (SEM), and then, determining the predictive power of stress scores using path analysis. This predictor was tested with a population of 584 first-year students enrolled in mathematics and liberal-arts classes at a university in California (USA). The experiment was initially performed during the first semester and then repeated in the following one. After SEM confirmed that there was a correlation between stress overload and low performance. Path analysis revealed that stress scores predicted semester GPA better than most other traditional predictors ($p < 0.0001$). This held true for both of the studied semesters. However, dropout could only be effectively predicted using grade data.

3.2. Early Warning Systems

Among the reviewed instances of EWS, there is one in particular that stands out above the rest: Course Signals, documented by Arnold and Pistilli [23]. This is the earliest application that was considered in this study, being documented in a paper in 2012, but the EWS itself has been used in courses at Purdue University since the late 2000s. Course Signals has been highly influential to many other EWS developed after it, becoming one of the most referenced systems by researchers in the community.

Course Signals works in conjunction with the LMS used at Purdue University, Blackboard Vista. The EWS works using student-related data regarding demographics, performance in tasks and exams, effort indicators (measured by the interaction with online course materials), and prior academic history. With these data, the system is able to estimate each individual student’s risk of failing a course. Risk is represented in a three-level scale, color coded similar to a traffic light, with green meaning low or no risk of failing, yellow representing a mild risk, and red being high likelihood of failing the course. This kind of multi-level representation is easy for instructors and students to understand, and ended up becoming a staple of other EWS to come. Once the level of risk has been assessed, instructors can implement an intervention plan of their choice, including actions such as sending e-mails to the

student or scheduling a face-to-face meeting. According to the authors, there was an improvement of around 15% in student retention after Course Signals was introduced at Purdue University, and the tool garnered an overall positive reception by students and instructors.

Internally, Course Signals uses what the authors call a “Student Success Algorithm” (SSA) in order to process the input data. This algorithm assigns specific weights to each of the input categories and produces a single score, representative of the perceived level of risk [24].

A screenshot of the Course Signals user interface [25] is shown in Figure 2.

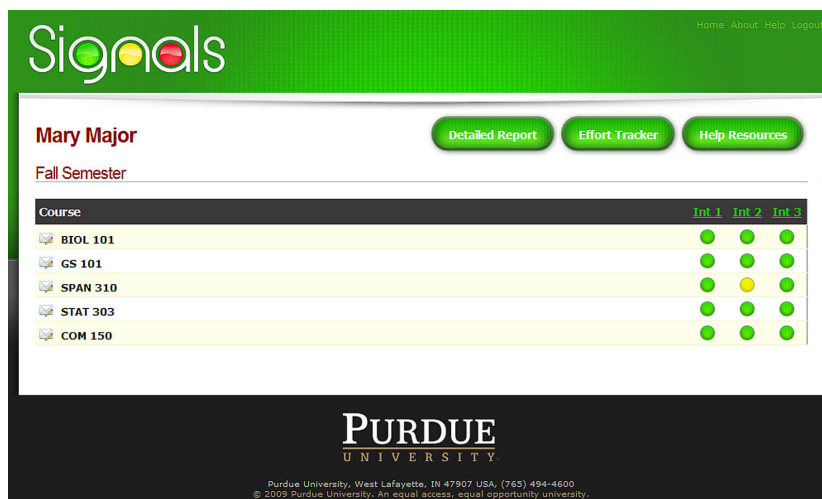


Figure 2. Student overview screen in Course Signals [25].

As aforementioned, many EWS developed after Course Signals took heavy inspiration from it. An important example is Student Explorer, presented by Krumm et al. [26], which implements a similar three-level scale to assess students’ risk of failing a course. Student Explorer mines LMS data regarding performance, in terms of points earned in tasks and exams, and engagement, as in the number of accesses to the course site. These data are weighted to produce three categories to classify students: “encourage”, “explore”, and “engage”, in increasing order of risk. The way Student Explorer classifies students regarding their performance and engagement indicators is represented in the original paper as a table, reproduced in Figure 3.

Student Percentage Points Earned	Percentage Points behind Course Average	Site Visits Percentile Rank	Classification
$\geq 85\%$			Encourage
$75\% \leq X < 85\%$	$< 15\%$		Explore
$75\% \leq X < 85\%$	$\geq 15\%$	$< 25^{\text{th}}$ percentile	Explore
$75\% \leq X < 85\%$	$\geq 15\%$	$\geq 25^{\text{th}}$ percentile	Encourage
$65\% \leq X < 75\%$	$< 15\%$	$< 25^{\text{th}}$ percentile	Engage
$65\% \leq X < 75\%$	$< 15\%$	$\geq 25^{\text{th}}$ percentile	Explore
$65\% \leq X < 75\%$	$\geq 15\%$		Explore
$55\% \leq X < 65\%$	$\geq 10\%$		Explore
$55\% \leq X < 65\%$	$< 10\%$		Engage
$< 55\%$			Engage

Figure 3. Classification criteria in Student Explorer [26].

Student Explorer is presented as a supporting tool for student advisors, making the task of identifying struggling students much easier. The application was first implemented in STEM courses

at universities in Maryland and California, contributing to an overall improvement in student performance, as reflected by the general increase of GPA scores.

More interestingly, Student Explorer has seen improvements and further development in the years after it was first introduced. The following are add-ons and studies centered around this EWS:

- Waddington and Nam added LMS resource use to the existing input data in Student Explorer [27]. This includes information on access to lecture notes and completion of assignments. This system was tested across 10 consecutive semesters, involving a total of 8762 students in an introductory chemistry course. The authors observed the existence of a significant correlation between resource use and the final grade obtained in the course, using logistic regression as analysis method. The activities that were most influential in the final grade were those related to exam preparation.
- Brown et al. performed multiple analytics studies with the help of Student Explorer, the first of which involved determining the reasons why students fall into a medium or high risk category [28]. This was done by using event history analysis techniques to determine the probability that a student enters one of the at-risk levels. This was tested over a population of 556 first-year students belonging to different study programs. As a result of this experiment, it was determined that the main reason students are classified into the “engage” level, or high risk, is underperformance in course tasks and exams. However, there was a wider array of circumstances that increased the odds of students falling into the “explore” category, or medium risk. These circumstances included being in large classes, sophomore level courses, and courses belonging to pure scientific degrees.
- A second study by Brown et al. investigated which were the best ways to help struggling students recover [29]. This study has some similarities with the previous one: this time, the authors used event history analysis to find out which intervention methods are the best for increasing the odds of students being removed from the at-risk levels. After experimenting with a population of 2169 first-year statistics students, they concluded that students at high risk benefited the most from better exam preparation, while those at medium risk required assistance in planning their study behaviors.
- Lastly, Brown et al. analyzed the effect of co-enrollment over student performance [30]. This study used binary logistic regression as the main analysis technique. The authors classified certain courses as “difficult”, according to the criterion of them having a higher amount of students classified as at-risk compared to most other courses. This extension of Student Explorer was implemented in an introductory programming course with 987 enrolled students. The authors determined that, given a specific focal course, students had a significantly higher chance of entering the “explore” or “engage” categories if they were enrolled in a “difficult” course at the same time.

There exist many other EWS that deviate from the basic formula created by Course Signals and closely followed by Student Explorer. These are applications usually tailored for use in one specific kind of learning environment. Some notable examples are described next.

LADA (“Learning Analytics Dashboard for Advisors”) was developed by Gutiérrez et al. and, as its name implies, has the goal of supporting the decision-making process of academic advisors [31]. Among its features, there is a module that provides predictions of the students’ chances of passing a specific course. LADA uses student grades, courses booked by a student, and the number of credits per course as input data. As for the analysis technique, it uses multilevel clustering to assess risk on a five-level scale. It does so by establishing comparisons with students that had similar profiles in previous cohorts. The system was deployed in two different universities: one located in Europe and the other in Latin America. Student advisors were generally satisfied with LADA’s utility, stating that the tool allowed them to reduce the time that it took to analyze cases of individual students, enhancing the efficiency of their decision-making.

SurreyConnect, created by Akhtar et al., was presented as a tool to assist the teacher during laboratory sessions of a computer-aided design course at University of Surrey (England) [32].

This application includes features such as the ability to broadcast the computer screen of a student to the rest of the class or to remotely connect to a specific student's computer to provide assistance. The feature that turns SurreyConnect into an EWS is its analytics module, which provides predictions of students potentially at risk of being unsuccessful in the course. Since this application was specifically designed to be used in laboratory environments, it is able to use some input data that are not available in other EWS, such as where the students are seated within the lab and who their neighbors are. Additionally, class attendance and time spent doing exercises are also tracked. The significance of each type of input data was determined running an ANOVA test, and the variables correlated with student performance were identified using Pearson correlation. Through these tests, it was observed that class attendance and time spent doing tasks had a direct relationship with learning outcomes. Additionally, the location of a student in the classroom and the identity of the closest neighbors also had an impact over performance.

Howard et al. were the authors of an experiment featuring the use of an EWS in a practical statistics course at University College Dublin [33]. This course implemented a continuous assessment system: 40% of the final grade was awarded for completing a series of weekly tasks throughout the course, which had a total duration of 12 weeks. The main goal of this study was determining the best time to perform a prediction of each student's final grade: early enough so that corrective measures can be more effectively applied to low-performing students, but not so early that predictions are unreliable due to insufficient information. The input data were obtained from the university's LMS, Blackboard, and included demographic information, number of accesses to learning resources, and the results from the aforementioned weekly tasks. Eight different predictive algorithms were tested, with Bayesian Additive Regressive Trees (BART) yielding the best results. The EWS was able to predict the final grade of the students with a mean absolute error of 6.5% by Week 6, exactly halfway through the course. This performance proves that the EWS was able to make reliable enough predictions at early stages in the course, when corrective measures taken by the teacher can be most effective.

Over at Hangzhou Normal University in China, Wang et al. developed an EWS with the objective of reducing student dropout and minimizing delays in graduation [34]. This application stands out due to the types of input data that it uses. As with many similar tools, information related to student grades, attendance to classes, and use of online learning resources is used for prediction purposes. However, this EWS also includes records from the university library, as in the books that are borrowed; and the dormitory, as in the times at which a student enters and leaves the dorm. This extra information enables the possibility of more closely monitoring student habits. This EWS classifies students on a seven-level scale, regarding the nature and severity of the risks that they are exposed to: underperformance, graduation delay, or dropout. The tool was tested for three consecutive semesters using a sample of 1712 students, trying three different classification algorithms: decision tree, artificial neural network, and Naïve Bayes. Out of the three algorithms, Naïve Bayes yielded the best results, obtaining an accuracy of 86%. Additionally, a principal component analysis showed that student grades and book borrowing trends were the most important indicators for predictions.

Cohen hypothesized that students dropping out from a course will first stop actively using said course's websites and resources. Moreover, this behavior could lead to dropout from degree studies altogether. Keeping this in mind, an EWS was built with the purpose of analyzing student activity in a quantitative way in order to provide an early identification of learner dropout [35]. The tool was developed in the context of a large Israeli university. As input data, it collected student activity indicators from the institution's LMS, Moodle. These indicators included the type and number of actions performed in the LMS, as well as their timing and frequency. The system performed analyses of student activity month by month, identifying students with unexpectedly low activity traces. These students could be flagged for being completely inactive during a specific month, or for low relative activity compared to the classroom average. The EWS was tested in three different undergraduate mathematics courses, with a total sample size of 362 students, achieving an average recall of 66% when identifying dropout students. A Mann-Whitney U test confirmed that students

who failed a course received more inactivity alerts than those who passed. This was also true for students who dropped out of their degree studies the following year compared to those who did not.

Akcapinar et al. built an EWS intended to work together with BookRoll, an e-book management system used in several Asian universities that provides access to course materials [36]. This tool collected student interactions with BookRoll as input, in the form of Experience API (xAPI) statements. Tracked information included logs regarding e-book navigation, page highlighting, and note taking. Using these data, the EWS applied a predictive model in order to identify students at risk of failing the course. This study was clearly geared towards experimentation, as 13 different prediction algorithms were tested. Additionally, three ways of processing the input data were tried: raw, where numeric input data were used as is; transformed, where percentile rank transformation was used to convert raw data to values between 0 and 1; and categorical, where transformed data were discretized into “Low”, “Medium” and “High” categories. The system was tested with a cohort of 90 students in a 16-week elementary informatics course. It was observed that the best performing algorithms were Random Forest for raw data, C4.5 decision trees for transformed data, and Naïve Bayes for categorical data. Additionally, the accuracy obtained with transformed data was lower than with raw or categorical data. From the third week of the course onward, both RF with raw data and NB with categorical data were able to correctly predict over 80% of at-risk students.

Finally, Plak et al. implemented an EWS at Vrije Universiteit Amsterdam in order to support student counselors [37]. This study involved 758 first-year students from 12 different study programs, as well as 34 counselors. The tool estimated a dropout probability for each student based on progress indicators, as reflected by elements such as grades or number of credits obtained. The calculated dropout probability, along with extra information regarding student motivation and performance, was presented to the counselor via an analytics monitor. The outcome of this experiment showed that, while the early identification of at-risk students was useful for counselors, the EWS-assisted counseling sessions did not make an impact on student dropout in any noticeable way. This hints at the existence of an underlying problem that causes underperformance, which cannot be solved only with the identification of at-risk students.

4. Discussion

Every single study included in this literature review had the same general purpose: using data related to students and their environments to improve academic results or address problems that exist in higher education scenarios. However, the approaches presented by each one of these papers are extremely diverse in many aspects, such as the targeted context, specific analysis goal, types of input data, and used algorithms. The present discussion points out the similarities that exist among some of the studies, as well as the most distinctive particularities that were observed.

The selected papers in this review consist of 13 predictors and 13 EWS. The differences between these two types of systems are explained at the beginning of Section 3. The line separating both categories is sometimes very thin, as some of the predictors are able to offer results using only data available early in a course, which would easily allow them to serve as basis for an EWS. Examples of these situations can be observed in the works published by Thompson et al. [11], Umer et al. [13], and Hirose [15].

In terms of document types, 14 of the selected articles in this review are full papers published in journals, 11 are articles presented at conferences, and the remaining one [26] is a book chapter.

Interestingly, all 14 of the full papers were published in different journals, covering the topics of research in education, technology-enhanced learning, and educational psychology. This makes it difficult to highlight specific journals that publish a high number of works related to predictive analysis in education, but, at the same time, it proves that this is a topic of interest for data scientists and educational researchers alike. As for the relevance of the journals themselves, 10 out of the 14 papers were listed in JCR, most of them being classified in either the first or the second quartile in terms of impact. Some of the articles were found in journals that are consistently rated among the most

relevant ones in their fields, such as *Computers & Education* [19], *Internet and Higher Education* [33], and *Computers in Human Behavior* [31].

On the other hand, five of the 11 conference articles were presented in different editions of the *International Conference on Learning Analytics & Knowledge*, considered to be one of the main research forums in the learning analytics field. The remaining six were all presented in different conferences, including some long-running ones such as the *International Conference on Information, Intelligence, Systems and Applications* [17] and the *International Conference on Software and Computer Applications* [12].

It is important to note that most of the selected papers were published fairly recently as of the writing of this review. As can be observed in Figure 4, five documents were published in 2017, eleven in 2018, and five during the first half of 2019, with the remaining five studies being published between 2012 and 2016. This suggests that the topic of predictors and EWS in higher education is far from exhausted, and its popularity among researchers is still rising. It is safe to say that further developments and unique studies regarding this field of knowledge will keep appearing in the near future.

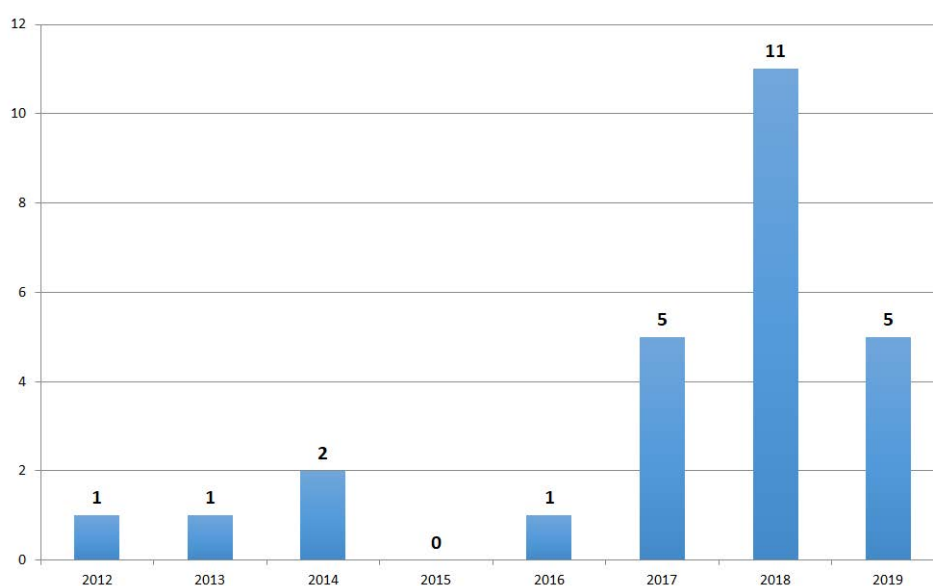


Figure 4. Number of papers by publication year.

Another interesting aspect is the diversity of studies regarding the geographical location where they were carried out. While most instances were developed in areas of North America, Europe and Asia, there were examples of predictors and EWS developed on every continent. This suggests that the attention garnered by predictive analysis in higher education is not exclusive to researchers in specific parts of the world. Instead, this topic has worldwide (albeit nonuniform) relevance, as shown in Figure 5.

Moving on to the contents of studies themselves, the selected papers could be classified regarding their general prediction goal. Table 2 shows a summary of prediction goals, as well as how frequently they appear. Implementing a classifier to predict which students are at risk of failing a course is by far the most popular goal of the predictors and EWS described in this review, appearing in 15 out of the 26 selected papers. It is worth mentioning that some of the documents define this goal with different words, such as “predicting low and high-performing students”, but, in practice, these studies are trying to achieve the same objective. The following goal in terms of popularity is the prediction of student grades, which could be of an exam, a course or the average of a term or degree program. A few other papers focused on estimating students’ risk of dropping out of a course or degree program. The study published by Schuck [16] did not fall into any of the previous three categories, as its prediction goal was the graduation rate: the fraction of students that finish their studies in the intended number of years. This was also the only predictor that analyzed data corresponding to academic institutions as a whole, rather than individual students.

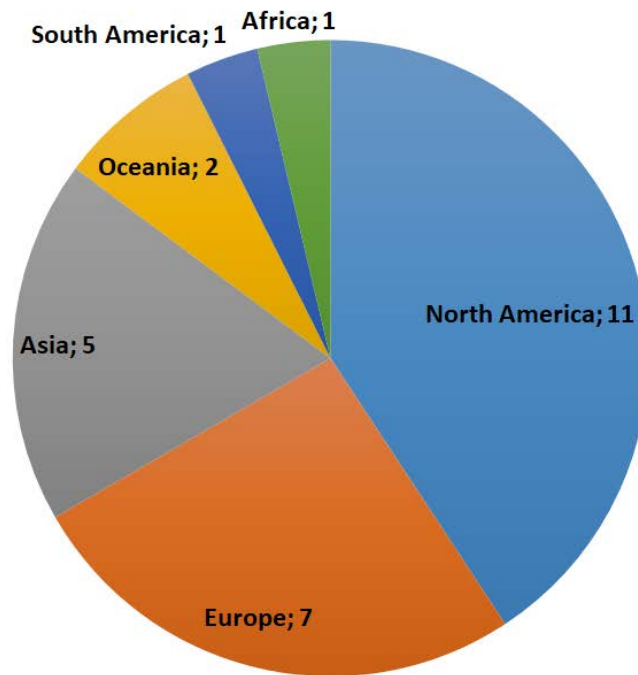


Figure 5. Number of papers by location where the study took place.

Another important detail is the general prevalence of classifiers over regression algorithms in these applications. Classifiers are typically used in the assessment of failure and dropout risks, since the prediction outputs are categorical variables. Regression algorithms were mostly used for numerical estimations of student grades. It was apparent that for most contexts, and especially in the case of EWS, the prediction of categorical outcomes such as “failing or succeeding student” provides enough information to instructors, advisors, and/or students in order to implement corrective measures if needed. Results provided by regression algorithms, on the other hand, are usually not as reliable as those obtained with classifiers, and, in many cases, estimations of numeric grades are unnecessary considering the purpose that these tools are trying to fulfill. In fact, the main research goal of some of the studies using regression algorithms was not the prediction outcome itself, but rather assessing the strength of correlations between inputs and outputs, as seen in the papers by Schuck [16] and Amirkhan and Kofman [22].

In terms of specific prediction algorithms, the most commonly used classifiers were Naïve Bayes, logistic regression, RF, KNN, SVM, and neural network. Meanwhile, regression-based predictors usually relied on some variant of the linear regression algorithm. However, many authors tried more obscure algorithms, or even self-defined ones, such as Student Explorer’s “student success algorithm” [24]. Additionally, one importance characteristic of many of these studies is their experimental nature, leading them to try out and compare the performance of many different algorithms. Examples of this trend are Akcapinar et al. [36], who tested 13 different classification algorithms, and Adekitan and Salau [18], who tried both classifiers and regression algorithms in their work.

Table 2. Number of papers by prediction goal.

Prediction Goal	Type	Number Published
Risk of failing a course	Classification	15
Dropout risk	Classification	3
Grade prediction	Regression	7
Graduation rate	Regression	1

One of the most important defining characteristics of each predictive application was the types of data that it used as input. Table 3 shows how many studies made use of each category of input data. It can be observed that the most common indicators by far are those related to student performance and engagement. Performance is measured in terms of students' grades in past exams, tasks, or courses. Most authors used this information if available, since past grades are always a strong indicator of how the student will perform in the future. Engagement and effort are most commonly measured by tracking student activity in educational online platforms, such as the institutional LMS. However, there are also examples of engagement being assessed via direct surveys to students, such as the study by Benablo et al. [12]. Engagement indicators have the advantage of being easy to access and collect in most cases, and they provide high volumes of data as well: for a given course, tens of thousands of activity records can be collected from a LMS.

Other kinds of input data are less common, but not necessarily less significant. Some studies incorporate demographic information about the students, such as gender, age, or ethnicity. On the other hand, studies that target several different kinds of courses often include input data regarding course characteristics, such as the type of contents and their value in credits.

Some of the studies target very specific contexts, and thus are able to work with unique kinds of input data that are unavailable otherwise. For example, Chen analyzed notes taken by students during lectures [21], Akhtar et al. collected information regarding student positioning in the classroom [32], and Schuck focused on crime and violence indicators [16].

Table 3. Number of observed use instances of input data types.

Input Data Type	Appearances in Studies
Student demographics and background	6
Student engagement and effort	15
Student performance and academic history	19
Course, degree or classroom characteristics	4
Others	6

Lastly, it is important to know that testing these predictors and EWS in real higher education scenarios is essential in order to assess their utility. It is difficult to compare these studies in terms of how well predictors and EWS were tested, since optimal test scenarios vary depending on the context for which the tool was designed. However, one aspect that could help understand whether results from tests are reliable or not is the number of students who participated in the study. A higher number of students implies more volume—and usually, more variety—of input data, which can increase the credibility of measured statistics such as the accuracy of a predictive model.

Figure 6 showcases the number of students who participated in testing procedures for each predictor and EWS. This excludes papers that do not provide specific number of students: the documents presenting Course Signals [23] and Student Explorer [26] focus on describing the EWS itself rather than specific cases of application, while the study by Schuck [16] used data at an institutional level, rather than related to specific students. In the high end of the scale, there are tools tested with several thousand students. Two studies stand out above the rest in terms of population size: Ornelas and Ordonez had data of 8658 students belonging to 13 different courses [10], while Waddington and Nam collected data regarding 8762 students across 10 semesters [27]. On the other hand, the system presented by Chen was tested on a class of only 38 students [21], while Akcapinar et al. included only 90 students in their study [36], which can be regarded as insufficient in order to make a solid evaluation of a predictor's performance.

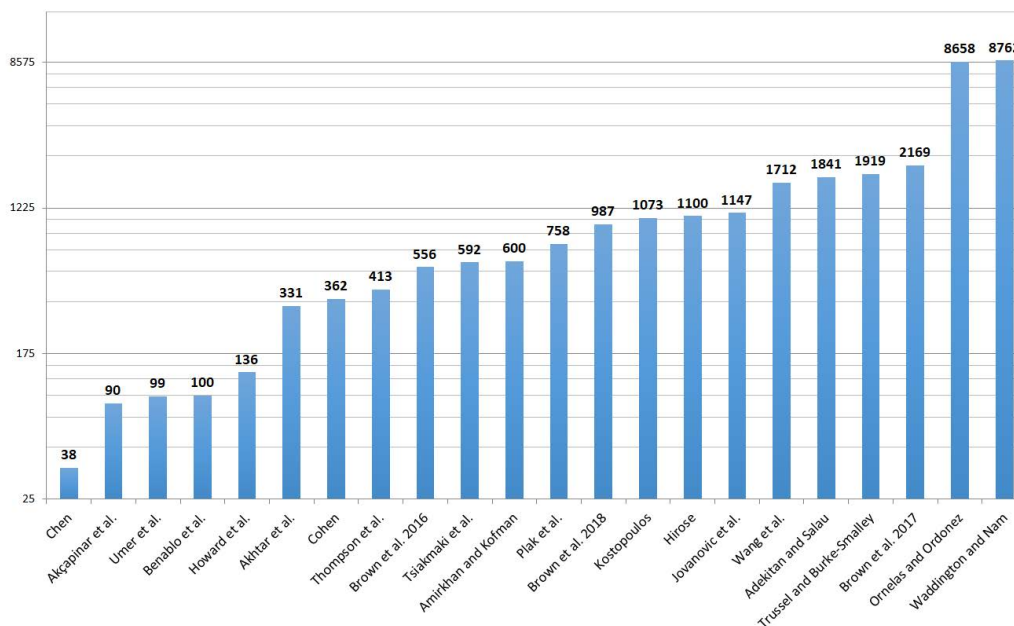


Figure 6. Size of the student population per study.

5. Conclusions

The present systematic literature review serves the purpose of providing a general overview on how predictive algorithms are being used in higher education environments, as of the time of writing this paper. After a search process that yielded 1382 results, 26 papers were selected as relevant examples of predictors and EWS applied in university contexts. Most of the selected studies were published in 2017 or later, which proves that this field of study is gathering significant research interest as of 2019.

The selected predictors and EWS present great diversity in terms of contexts where they were implemented, input data they relied on, prediction algorithms they used, and the specific prediction goal they sought. However, it is important to understand that most of these studies were performed with an experimental mindset. For the most part, these predictors and EWS are not in a mature enough state to be permanently implemented in university courses, with the notable exception of Course Signals EWS [23], a tool that served as main inspiration for many of the EWS that came after it. Increasing the level of adoption of predictors and EWS in higher education should be a priority moving forward.

The experimental nature of many of the applications included in this study is reflected by the fact that they are tailored for use in very specific learning contexts. As mentioned in Section 4, this has the advantage of allowing the use of uncommon types of input data, which are not available in environments other than the one that is targeted by the specific predictor or EWS. In addition to this, it is easier to obtain accurate predictions when focusing on a single, isolated context. However, the downside is that these applications are not useful when taken out of their intended environments. One of the keys to the success of Course Signals, besides being developed earlier than most other applications of its kind, is that it relies on activity and performance data obtained from a LMS, information that is available in most current higher education contexts. Thus, this application could be implemented in a multitude of different courses and degree programs across multiple academic years. A short-term objective in this field of research should be developing more tools able to function well in many different educational contexts, which would foster a more widespread adoption of EWS in higher education institutions.

Regarding EWS that act as teacher-assisting tools, it was also observed that the help they provide in order to carry out interventions over struggling students is rather limited. These applications perform the necessary predictions and present results in an easily understandable way, but the decision

of what corrective measures to apply is usually made by the teacher or advisor on their own. One big leap of progress for EWS would be the ability to recommend ways of helping students which could be most effective in each particular case. If these are found to be reliable, the system could even perform them automatically. To follow this line of development, however, the EWS would need to perform some sort of profile analysis for each student. Studies exist that show classifying students regarding their study habits is possible [38]. This could provide valuable information in order to perform effective interventions.

Another aspect that may hinder the development of these tools is the availability of data, or lack thereof. Openly available datasets including student-related information are scarce, causing the need for researchers to mine data themselves, usually relying on information that can be obtained from their home academic institutions. This implies that the resulting applications are typically tailored to work only in specific academic contexts. Additionally, creating a well-performing predictive application is extremely difficult for researchers who do not have access to good data sources of their own. If more student-related data became openly accessible in the future, there would be more people in the research community able to work in this field of study, be it developing new tools or proposing improvements for existing ones. This would help address the aforementioned maturity problem, and it would imply a further boost in popularity for the field of predictive analysis in education.

Author Contributions: Methodology, F.A.M.-F.; investigation, M.L.-D. and M.C.-R.; resources, M.L.-D. and M.L.-N.; writing—original draft preparation, M.L.-D.; writing—review and editing, M.L.-D., M.C.-R., M.L.-N. and F.A.M.-F.; supervision, M.C.-R., M.L.-N. and F.A.M.-F.; project administration, M.L.-N.; funding acquisition, M.L.-D., M.C.-R., M.L.-N. and F.A.M.-F.

Funding: This study was partially financed by public funds granted by the Department of Education of the Galician Regional Government (GRG), with the purpose of supporting research activities carried out by PhD students. This work was supported by the Spanish State Research Agency and the European Regional Development Fund (ERDF) under the PALLAS (TIN2016-80515-R AEI/ERDF, EU) Project, by the GRG and the ERDF through “Agrupación Estratégica Consolidada de Galicia accreditation 2016–2019”, and by the GRG under projects ED431B 2017/67 and ED431D 2017/12.

Conflicts of Interest: The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

BART	Bayesian Additive Regressive Trees
EWS	Early Warning System(s)
GBC	Gradient Boosting Classifier
GP	Gaussian Processes
GPA	Grade Point Average
KNN	K-Nearest Neighbors
LA	Learning Analytics
LADA	Learning Analytics Dashboard for Advisors
LASI	Learning Analytics Summer Institute
LDA	Linear Discriminant Analysis
MOOCs	Massive Open Online Courses
NB	Naïve Bayes
NN	Neural Network
OLS	Ordinary Least Squares
PCR	Principal Components Regression
PRISMA	Preferred Reporting Items for Systematic reviews and Meta-Analyses
RF	Random Forest
SEM	Structural Equation Modeling
STEM	Science, Technology, Engineering and Mathematics
SVM	Support Vector Machine

Appendix A. Comparative Summary of Results

Table A1. Summary of the most relevant aspects of the selected documents, sorted by publication date.

Reference	Date	Type	Scenario/Context	Prediction Goal	Input Data	Algorithm(s)
Arnold and Pistilli [23]	April 2012	EWS	Course Signals EWS, used in courses at Purdue University (USA), with a special focus on first year students.	Students' risk of failing the course.	Obtained from the institutional LMS (Blackboard Vista): demographic information, performance and effort indicators, prior academic history.	"Student Success Algorithm", producing a single score by weighting all input parameters [24].
Chen [21]	May 2013	Predictor	A group of 38 freshmen students from a Taiwanese university.	Final grade in the course.	Quality and quantity of the notes taken by students, both during and after lectures.	Hierarchical regression analysis.
Krumm et al. [26]	March 2014	EWS	Student Explorer EWS, targeting STEM students at Maryland and California universities (USA).	Students' risk of failing a course.	Performance and effort indicators from the institutional LMS.	Weighted aggregation of input data.
Waddington and Nam [27]	March 2014	EWS	An extension of Student Explorer EWS. Tested with 8762 students enrolled in a chemistry course.	Final grade in a course.	Data used by the basic version of Student Explorer, plus use of academic resources in the LMS.	Multinomial logistic regression.
Brown et al. [28]	April 2016	EWS	An extension of Student Explorer EWS. Tested with 556 students belonging to various first-year courses.	Data used by the basic version of Student Explorer, plus contextual information such as size of cohorts and specific STEM field of the degree.	Event history analysis.	
Schuck [16]	February 2017	Predictor	Over 1000 higher education institutions in the USA.	Graduation rate, i.e. fraction of students that finish their degree within the intended number of years.	Crime and violence indicators in and around campus, provided by the US Department of Education and the National Center for Education Statistics.	Multivariate least squares regression.

Table A1. Cont.

Reference	Date	Type	Scenario/Context	Prediction Goal	Input Data	Algorithm(s)
Brown et al. [29]	March 2017	EWS	An extension of Student Explorer EWS. Tested with 2169 students in a Statistics course.	Data used by the basic version of Student Explorer, plus type of interventions performed on struggling students.	Event history analysis.	
Akhtar et al. [32]	June 2017	EWS	Laboratory sessions of computer-aided design courses at University of Surrey (England), with a sample size of 331 students.	Students' risk of failing the course.	Attendance to class, location, and neighbors within the lab, time spent doing exercises.	ANOVA, Pearson correlation, linear regression.
Cohen [35]	October 2017	EWS	A sample of 362 students of mathematics and statistics at an Israeli university.	Chance of dropping out of the course.	LMS activity: type, timing, and frequency of actions performed.	Mann–Whitney U test to prove a correlation between low student activity and a higher dropout chance.
Ornelas and Ordonez [10]	October 2017	Predictor	13 courses at Rio Salado Community College (USA), with a sample size of around 8700 students.	Students' chance of obtaining a passing grade.	Engagement and performance indicators from the institutional LMS.	Naïve Bayesian classification.
Thompson et al. [11]	February 2018	Predictor	An introductory biology course at Northern Kentucky University (USA), including 413 students.	Students' chance of passing the course.	Results from Lawson's Classroom Test of Scientific Reasoning and ACT Mathematics Test, taken before the start of the course.	Logistic regression.
Benablo et al. [12]	February 2018	Predictor	100 Information Technologies and Computer Science students in the Philippines.	Identification of underperforming students.	Student age, gender, academic standing and procrastination indicators: time spent using social networks and playing online games.	SVM, KNN, RF. SVM had the best performance.

Table A1. Cont.

Reference	Date	Type	Scenario/Context	Prediction Goal	Input Data	Algorithm(s)
Brown et al. [30]	March 2018	EWS	An extension of Student Explorer EWS. Tested with 987 students in an introductory programming course.	Students' risk of failing a course.	Data used by the basic version of Student Explorer, plus difficulty estimations of concurrent courses.	Binary logistic regression.
Howard et al. [33]	April 2018	EWS	136 students in a Practical Statistics course at University College Dublin (Ireland).	Final grade in the course.	Results of weekly tests, as well as demographic information and access to online course resources.	RF, BART, XGBoost, PCR, SVM, NN, Splines, KNN. BART had the best performance.
Hirose [15]	July 2018	Predictor	Around 1100 calculus and algebra students in Japan.	Classification of students into "successful" and "not successful" categories. Estimation of students' abilities using item response theory.	Results of weekly multiple-choice tests.	KNN classifier.
Tsiakmaki et al. [17]	July 2018	Predictor	592 Business Administration students in Greece.	Final grade in second semester courses.	Final scores of first semester subjects.	Linear regression, RF, instance-based regression, M5, SVM, GP, bootstrap aggregating. RF had the best performance.
Amirkhan and Kofman [22]	July 2018	Predictor	600 freshmen students at a major public university in the USA.	Prediction of performance and dropout probability.	Stress indicators obtained from mid-semester surveys, as well as demographic information.	Structural equation modeling, path analysis.
Trussel and Burke-Smalley [20]	November 2018	Predictor	1919 business students at a public university in Tennessee (USA).	Cumulative GPA at the end of the degree program and academic retention.	Demographic and socioeconomic attributes, performance in pre-college stage.	OLS regression, logistic regression.
Umer et al. [13]	November 2018	Predictor	99 students enrolled in an introductory mathematics module at an Australian university.	Earliest possible reliable identification of students at risk of failing the course.	Assignment results in a continuous assessment model, as well as LMS log data.	RF, Naïve Bayes, KNN, and LDA. RF had the best performance.

Table A1. Cont.

Reference	Date	Type	Scenario/Context	Prediction Goal	Input Data	Algorithm(s)
Wang et al. [34]	November 2018	EWS	1712 students from Hangzhou Normal University (China).	Risk assessment of students regarding dropout and delays in graduation.	Grades, attendance and engagement indicators, as well as records from the university library and dorm in order to monitor student habits.	Decision tree, artificial neural network, Naïve Bayes. Naïve Bayes had the best performance.
Gutiérrez et al. [31]	December 2018	EWS	Learning Analytics Dashboard for Advisors (LADA) EWS, deployed in two universities: a European one and a Latin American one.	Students' chance of passing a course.	Student grades, courses booked by a student, number of credits per course.	Multilevel clustering.
Adekitan and Salau [18]	February 2019	Predictor	1841 engineering students at a Nigerian higher education institution.	Final grade point average (GPA) over a five year program.	Cumulative GPA over the first three years of the degree.	Classifiers: NN, RF, decision tree, Naïve Bayes, tree ensemble, and logistic regression. Logistic regression had the best performance. Additionally, linear and quadratic regression models were tested.
Plak et al. [37]	March 2019	EWS	EWS deployed at Vrije Universiteit in Amsterdam (Netherlands), tested with 758 students.	Identification of low-performing students.	Progress indicators, such as grades or obtained credits.	Generalized additive model.
Kostopoulos et al. [14]	April 2019	Predictor	1073 students in an introductory informatics module at a Greek open university.	Identification of students at risk of failing a course.	Student demographics and academic achievements; and LMS activity indicators. The data were divided into two views in order to use a co-training method.	Custom co-training method, using combinations of KNN, Extra Tree, RF, GBC, and NB as underlying classifiers.

Table A1. Cont.

Reference	Date	Type	Scenario/Context	Prediction Goal	Input Data	Algorithm(s)
Akcapinar et al. [36]	May 2019	EWS	90 students in an Elementary Informatics course at an Asian university.	Identification of students at risk of failing the course.	Data from the e-book management system BookRoll: book navigation, page highlighting and note taking.	Comparison of 13 different algorithms. RF had the best performance when using raw data. However, NB outperformed the rest when using categorical data.
Jovanovic et al. [19]	June 2019	Predictor	First year engineering course at an Australian university using the flipped classroom model. Tested during three consecutive years, with a number of students ranging from 290 to 486 each year.	Final grade in the course.	Indicators of regularity and performance related to pre-class activities. These activities included videos with multiple choice questions as well as problem sequences.	Multiple linear regression.

References

1. Kempster, S.; Mathews, T. Earth Science Data Analytics: Definitions, Techniques and Skills. *Data Sci. J.* **2017**, *16*, 6, doi:10.5334/dsj-2017-006. [CrossRef]
2. Gandomi, A.; Haider, M. Beyond the hype: Big data concepts, methods, and analytics. *Int. J. Inf. Manag.* **2015**, *35*, 137–144, doi:10.1016/j.ijinfomgt.2014.10.007. [CrossRef]
3. Reid, B. Global early warning systems for natural hazards: Systematic and people-centred. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **2006**, *364*, 2167–2182, doi:10.1098/rsta.2006.1819. [CrossRef]
4. Bussiere, M.; Fratzscher, M. Towards a new early warning system of financial crises. *J. Int. Money Financ.* **2006**, *25*, 953–973, doi:10.1016/j.jimonfin.2006.07.007. [CrossRef]
5. Smith, M.E.B.; Chiovaro, J.C.; O’Neil, M.; Kansagara, D.; Quinones, A.; Freeman, M.; Motu’apuaka, M.; Slatore, C.G. *Early Warning System Scores: A Systematic Review*; VA Evidence-Based Synthesis Program Reports; Department of Veterans Affairs: Washington, DC, USA, 2014.
6. Siemens, G. About: Learning Analytics & Knowledge: February 27–March 1, 2011 in Banff, Alberta. Available online: <https://web.archive.org/web/20190112095031/https://tekri.athabascau.ca/analytics/> (accessed on 12 January 2019).
7. Liz-Domínguez, M.; Caeiro-Rodríguez, M.; Llamas-Nistal, M.; Mikic-Fonte, F. Predictors and Early Warning Systems in Higher Education — A Systematic Literature Review. In *Learning Analytics Summer Institute Spain 2019: Learning Analytics in Higher Education*; LASI Spain 2019; CEUR: Vigo, Spain, 2019; pp. 84–99.
8. Liberati, A.; Altman, D.G.; Tetzlaff, J.; Mulrow, C.; Gøtzsche, P.C.; Ioannidis, J.P.A.; Clarke, M.; Devereaux, P.J.; Kleijnen, J.; Moher, D. The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *PLoS Med.* **2009**, *6*, e1000100, doi:10.1371/journal.pmed.1000100. [CrossRef] [PubMed]
9. Hwang, G.J.; Spikol, D.; Li, K.C. Guest Editorial: Trends and Research Issues of Learning Analytics and Educational Big Data. *J. Educ. Technol. Soc.* **2018**, *21*, 134–136.
10. Ornelas, F.; Ordonez, C. Predicting Student Success: A Naïve Bayesian Application to Community College Data. *Technol. Knowl. Learn.* **2017**, *22*, 299–315, doi:10.1007/s10758-017-9334-z. [CrossRef]
11. Thompson, E.D.; Bowling, B.V.; Markle, R.E. Predicting Student Success in a Major’s Introductory Biology Course via Logistic Regression Analysis of Scientific Reasoning Ability and Mathematics Scores. *Res. Sci. Educ.* **2018**, *48*, 151–163, doi:10.1007/s11165-016-9563-5. [CrossRef]
12. Benablo, C.I.P.; Sarte, E.T.; Dormido, J.M.D.; Palaoag, T. Higher Education Student’s Academic Performance Analysis Through Predictive Analytics. In Proceedings of the 2018 7th International Conference on Software and Computer Applications, Kuantan, Malaysia, 8–10 February 2018; ACM: New York, NY, USA, 2018; pp. 238–242, doi:10.1145/3185089.3185102. [CrossRef]
13. Umer, R.; Susnjak, T.; Mathrani, A.; Suriadi, S. A learning analytics approach: Using online weekly student engagement data to make predictions on student performance. In Proceedings of the 2018 International Conference on Computing, Electronic and Electrical Engineering (ICE Cube), Quetta, Pakistan, 12–13 November 2018; pp. 1–5, doi:10.1109/ICECUBE.2018.8610959. [CrossRef]
14. Kostopoulos, G.; Karlos, S.; Kotsiantis, S.B. Multi-view Learning for Early Prognosis of Academic Performance: A Case Study. *IEEE Trans. Learn. Technol.* **2019**, *12*, 212–224, doi:10.1109/TLT.2019.2911581. [CrossRef]
15. Hirose, H. Success/Failure Prediction for Final Examination Using the Trend of Weekly Online Testing. In Proceedings of the 2018 7th International Congress on Advanced Applied Informatics (IIAI-AAI), Yonago, Japan, 8–13 July 2018; pp. 139–145, doi:10.1109/IIAI-AAI.2018.00036. [CrossRef]
16. Schuck, A.M. Evaluating the Impact of Crime and Discipline on Student Success in Postsecondary Education. *Res. Higher Educ.* **2017**, *58*, 77–97, doi:10.1007/s11162-016-9419-x. [CrossRef]
17. Tsiakmaki, M.; Kostopoulos, G.; Koutsonikos, G.; Pierrakeas, C.; Kotsiantis, S.; Ragos, O. Predicting University Students’ Grades Based on Previous Academic Achievements. In Proceedings of the 2018 9th International Conference on Information, Intelligence, Systems and Applications (IISA), Zakynthos, Greece, 23–25 July 2018; pp. 1–6, doi:10.1109/IISA.2018.8633618. [CrossRef]
18. Adekitan, A.I.; Salau, O. The impact of engineering students’ performance in the first three years on their graduation result using educational data mining. *Heliyon* **2019**, *5*, e01250, doi:10.1016/j.heliyon.2019.e01250. [CrossRef] [PubMed]

19. Jovanovic, J.; Mirriahi, N.; Gašević, D.; Dawson, S.; Pardo, A. Predictive power of regularity of pre-class activities in a flipped classroom. *Comput. Educ.* **2019**, *134*, 156–168, doi:10.1016/j.compedu.2019.02.011. [CrossRef]
20. Trussel, J.M.; Burke-Smalley, L. Demography and student success: Early warning tools to drive intervention. *J. Educ. Bus.* **2018**, *93*, 363–372, doi:10.1080/08832323.2018.1496893. [CrossRef]
21. Chen, P.H. The Effects of College Students' In-Class and After-Class Lecture Note-Taking on Academic Performance. *Asia-Pacific Educ. Res.* **2013**, *22*, 173–180, doi:10.1007/s40299-012-0010-8. [CrossRef]
22. Amirkhan, J.H.; Kofman, Y.B. Stress overload as a red flag for freshman failure and attrition. *Contempor. Educ. Psychol.* **2018**, *54*, 297–308, doi:10.1016/j.cedpsych.2018.07.004. [CrossRef]
23. Arnold, K.E.; Pistilli, M.D. Course Signals at Purdue: Using Learning Analytics to Increase Student Success. In Proceedings of the 2nd International Conference on Learning Analytics and Knowledge (LAK '12), Vancouver, BC, Canada, 29 April–2 May 2012; ACM: New York, NY, USA, 2012; pp. 267–270, doi:10.1145/2330601.2330666. [CrossRef]
24. Arnold, K.E. Signals: Applying Academic Analytics. *Educ. Q.* **2010**, *33*, n1.
25. Signals Tells Students How They're Doing Even Before the Test. Available online: <https://web.archive.org/web/20100717072052/http://news.uns.purdue.edu/x/2009b/090827ArnoldSignals.html> (accessed on 17 July 2010).
26. Krumm, A.E.; Waddington, R.J.; Teasley, S.D.; Lonn, S. A Learning Management System-Based Early Warning System for Academic Advising in Undergraduate Engineering. In *Learning Analytics: From Research to Practice*; Larusson, J.A., White, B., Eds.; Springer: New York, NY, USA, 2014; pp. 103–119, doi:10.1007/978-1-4614-3305-7_6. [CrossRef]
27. Waddington, R.J.; Nam, S. Practice Exams Make Perfect: Incorporating Course Resource Use into an Early Warning System. In Proceedings of the Fourth International Conference on Learning Analytics and Knowledge (LAK '14), Indianapolis, IN, USA, 24–28 March 2014; ACM: New York, NY, USA, 2014; pp. 188–192, doi:10.1145/2567574.2567623. [CrossRef]
28. Brown, M.G.; DeMonbrun, R.M.; Lonn, S.; Aguilar, S.J.; Teasley, S.D. What and when: The Role of Course Type and Timing in Students' Academic Performance. In Proceedings of the Sixth International Conference on Learning Analytics & Knowledge (LAK '16), Edinburgh, UK, 25–29 April 2016; ACM: New York, NY, USA, 2016; pp. 459–468, doi:10.1145/2883851.2883907. [CrossRef]
29. Brown, M.G.; DeMonbrun, R.M.; Teasley, S.D. Don't Call It a Comeback: Academic Recovery and the Timing of Educational Technology Adoption. In Proceedings of the Seventh International Learning Analytics & Knowledge Conference (LAK '17), Vancouver, BC, Canada, 13–17 March 2017; ACM: New York, NY, USA, 2017; pp. 489–493, doi:10.1145/3027385.3027393. [CrossRef]
30. Brown, M.G.; DeMonbrun, R.M.; Teasley, S.D. Conceptualizing Co-enrollment: Accounting for Student Experiences Across the Curriculum. In Proceedings of the 8th International Conference on Learning Analytics and Knowledge (LAK '18), Sydney, New South Wales, Australia, 7–9 March 2018; ACM: New York, NY, USA, 2018; pp. 305–309, doi:10.1145/3170358.3170366. [CrossRef]
31. Gutiérrez, F.; Seipp, K.; Ochoa, X.; Chiluiza, K.; De Laet, T.; Verbert, K. LADA: A learning analytics dashboard for academic advising. *Comput. Hum. Behav.* **2018**, doi:10.1016/j.chb.2018.12.004. [CrossRef]
32. Akhtar, S.; Warburton, S.; Xu, W. The use of an online learning and teaching system for monitoring computer aided design student participation and predicting student success. *Int. J. Technol. Des. Educ.* **2017**, *27*, 251–270, doi:10.1007/s10798-015-9346-8. [CrossRef]
33. Howard, E.; Meehan, M.; Parnell, A. Contrasting prediction methods for early warning systems at undergraduate level. *Internet Higher Educ.* **2018**, *37*, 66–75, doi:10.1016/j.iheduc.2018.02.001. [CrossRef]
34. Wang, Z.; Zhu, C.; Ying, Z.; Zhang, Y.; Wang, B.; Jin, X.; Yang, H. Design and Implementation of Early Warning System Based on Educational Big Data. In Proceedings of the 2018 5th International Conference on Systems and Informatics (ICSAI), Nanjing, China, 10–12 November 2018; pp. 549–553, doi:10.1109/ICSAI.2018.8599357. [CrossRef]
35. Cohen, A. Analysis of student activity in web-supported courses as a tool for predicting dropout. *Educ. Technol. Res. Dev.* **2017**, *65*, 1285–1304. doi:10.1007/s11423-017-9524-3. [CrossRef]
36. Akçapınar, G.; Hasnine, M.N.; Majumdar, R.; Flanagan, B.; Ogata, H. Developing an early-warning system for spotting at-risk students by using eBook interaction logs. *Smart Learn. Environ.* **2019**, *6*, 4, doi:10.1186/s40561-019-0083-4. [CrossRef]

37. Plak, S.; Cornelisz, I.; Meeter, M.; van Klaveren, C. Early Warning Systems for More Effective Student Counseling in Higher Education—Evidence from a Dutch Field Experiment. In Proceedings of the SREE Spring 2019 Conference, Washington, DC, USA, 6–9 March 2019; p. 4.
38. Jovanovic, J.; Gasevic, D.; Dawson, S.; Pardo, A.; Mirriahi, N. Learning analytics to unveil learning strategies in a flipped classroom. *Internet Higher Educ.* **2017**, *33*, doi:10.1016/j.iheduc.2017.02.001. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).