European Research Studies Journal
Volume XXV, Issue 2, 2022

pp. 350-366

Developing a Risk Model to Control Attrition by Analyzing Students' Academic and Nonacademic Data

Submitted 27/04/22, 1st revision 13/05/22, 2nd revision 12/06/22, accepted 25/06/22

Vassilis Zakopoulos[1], Ioannis Georgakopoulos[2], Pelagia Kontaxaki[3]

Abstract:

Purpose: *The research objective is to address the problem of students' attrition by identifying students who are liable to fail their courses. Students' behavioral engagement data along with students' nonacademic data were analyzed in terms of a binary logistics regression with a view to developing a model to decide on the risk factors.*
Design/methodology/approach: *A binary variable was modeled to describe students at risk and students not at risk. The students' behavioral engagement data constituted the independent variables in our regression analysis whereas the variable describing students at risk was the dependent variable. The students' behavioral engagement data was collected by students' learning activities. The eLearning part was implemented by Moodle. The data was collected after the final test. The regression analysis outcome was a classification table indicating the correct classification percentage of our model. In parallel an econometric study was also carried out in order to examine liable nonacademic risk factors.*
Findings: *Factors that are related to students' engagement could be deemed to be decisive in the context of our study. The econometric study proved that governmental financial support could be viewed as a cardinal factor that could potentially deter students from dropping out of university.*
Originality/value: *The originality of our research lies in the fact that the issue of controlling students' attrition is not addressed in a fragmentary way by just carrying out a specific analysis and coming up with results, like many similar studies in the literature. Thereby, a concrete methodology was developed on the basis of an established generic risk management framework. Therefore, the control of students at risk is included in the phases of a potent framework. The added value of our research is centered on the fact that our risk model could potentially be applied to any course in order to come up with the respective risk factors.*

Keywords: *Students' engagement, risk model, risk factors, students' attrition.*

JEL codes: *A12, C53, I21, I23, I38.*

Paper Type: *A research study.*

---

[1]*Corresponding Author, Laboratory Teaching Staff, Department of Accounting and Finance, University of West Attica, Greece, v.zakopoulos@uniwa.gr;*
[2]*Associate, Department of Accounting and Finance, University of West Attica, Greece, igtei@uniwa.gr;*
[3]*PhD Candidate, Department of Economics, University of Peloponnese, Greece, p.kontaxaki@uop.gr;*

1. Introduction

A crucial problem emerges in any type of learning. This problem is called 'attrition' and it is another term to describe students' dropout, laying emphasis on the students' feeling of incompetence motivated by their liable failure in a specific course (Abu *et al.,* 2012; Beaubouef and Mason, 2005). In parallel, the use of this term could be expanded in the context of an entire curriculum, accentuating students' decision to drop out of school. Thus, students' dropout has stirred researchers to endeavor to come up with proper prediction models in order to control attrition.

These models could analyze various data sets (including academic and non-academic data). The academic data could refer to students' interaction with learning activities and the entire learning process, pointing up students' behavioral engagement (Bujang *et al.,* 2021; Macfayden and Dawson, 2010; Georgakopoulos *et al.,* 2018; Georgakopoulos and Tsakirtzis, 2021; Zakopoulos *et al.,* 2021; Zakopoulos, 2022; Anagnostopoulos *et al.,* 2020; Alyahyan and Dustegor, 2020; Davidson *et al.,* 2018; Dias *et al.,* 2020; Filippidi *et al.,* 2010). The non-academic data could include data related to students' emotional engagement along with financial and social data (Ismail *et al.,* 2018).

The previously cited risk models which are found in literature could be assessed in terms of their accuracy (Amare and Simonova, 2021). Nevertheless, the respective models are developed on the premise of a proper statistical or machine-based technique. Therefore, there is not much substantial research which deals with attrition in a non- fragmentary way. Hence, there is space for research output in that field.

This paper proposes an attempt to control students' dropout by developing a potent risk model based on a proper analysis of students' academic and non-academic data. However, the risk model development process is included in the context of a generic risk management framework. In that sense, the process of controlling attrition is included in the context of a potent framework, denoting that our study attempts to address the issue of students' dropout in a non-fragmentary way, emphasizing on the concrete methodology on which the risk model is built. The risk model could potentially lead to a potent prediction model. Additionally, the prediction model could potentially result in an efficient warning system.

2. Literature Review

There are reports that the acquisition of ECTS functions as a determining factor for staying or leaving the university. After the primary influence of ECTS it seemed that gender plays an important role in this decision. 82.5% of women with less than 6 ECTS left compared to 59.1% of male students. Students who obtained more than 44 credits also observed significant differences between those who chose the department as the first choice and those who did not with a greater tendency to stay or transfer university degrees to the first (98.3%). Their studies in their initial choice and in this the educational level of the mother and the gender played a role. If the mother had an academic qualification, the student was 14.8% more likely to drop. Additionally, according to the same study, there were differences even between those who could not continue out than if she had a primary or secondary education (Casanova *et al.*, 2018).

According to sociological models, the age of students along with other demographic characteristics, such as gender, ethnicity and family background, is an important factor that is often explored for the study of education nationwide. It seems that students, who enroll in university at an older age for any reason, are more likely to drop out of school. This correlation suggests that opportunity costs increase with age.

In reference to gender, men tend to drop out of college more often than women. A strong advantage of women's integration is the higher ex-post returns in higher education, the postponement of motherhood, the stronger commitment to education, as well as the deprivation of income from insurance profit (Goldin *et al.,* 2006).

Respectively, another study has proved that key elements in this decision are the reduction of gender discrimination, the change in higher education incentives and the effect of these differences on family resources (Buchmann and DiPrete, 2006). In sociological models, initial studies have found that academic inclusion is more important than social inclusion for men, while the opposite is true for women (Aina *et al.,* 2021). Other findings suggest that women are more likely to drop out university when most of their classmates are men.

Studies show that the parental background (represented by the parents' education or profession) is negatively correlated with dropping out of university. An important study points out that the parenting profession plays a key role in determining both educational attainment and voluntary dropout, providing evidence that students with unskilled parents are at greater risk of failure (Johnes and McNabb, 2004). Students from a family background with low or no participation in higher education may find it difficult to adapt to the academic culture, precisely because they cannot benefit from the support provided by parents or friends as they do not have similar experiences (Aina *et al.,* 2021). Another study indicates that the probability of graduation for students from low-income families in the US is half that of those from affluent families. One possible reason is the need for students to work on their own during their studies (Aina, 2012).

Socio-demographic factors seem to be crucial to the future of a university student (Ghignoni, 2017; Schnepf, 2014; Bradley and Migali, 2013). The marital status of the students seems to be another factor that affects their attendance or not in the department. According to another important study, students who changed their marital status during their studies (married, divorced, widowed, etc.) are more likely

(5.9 times specifically) to drop out of university than the other students (Bonaldo and Pereira, 2016). The student's self-esteem is also a factor that seems to influence his course at the university. It was observed that students with low self-esteem in their abilities were more likely to give up when faced with poor grades or financial difficulties (Bennett, 2003).

In addition, the funding from the student's family and any scholarships he may receive during his studies are very important factors for the future of the students' academic career (Ghignoni, 2017; Schnepf, 2014; Hanushek and Wößmann, 2010; Stinebrickner, 2003; Stinebrickner, 2008). Students who receive a scholarship from a university or other organization have been shown to be less likely to drop out of university, while funded students are 4.7 times more likely to drop out than those who do not. Therefore, it seems that funding works inversely on the student (Bonaldo and Pereira, 2016).

Except of the financial and social factors that affect students' attrition, the research interest is directed into identifying factors that affect students' critical achievement in an attempt to control attrition. In that field, a lot of studies have proved that students' interaction with the learning activities plays a significant role in their final learning outcome. Therefore, students at risk could be identified by a proper analysis of their engagement (Georgakopoulos *et al.*, 2018; 2020; Zakopoulos *et al.*, 2020; 2021; Anagnostopoulos *et al.*, 2020; Georgakopoulos and Tsakirtzis, 2021; Tsakirtzis and Georgakopoulos, 2020; Macfayden and Dawson, 2010; Zakopoulos, 2022). These studies have also indicated that the risk factors are course-oriented. Thereby, there is not a specific set of factors that affect students' critical achievement. On the contrary, the risk factors are heavily dependent on the course structure.

3. Research Objective

The literature review scrutiny has indicated that demographic, social, financial and academic factors assume a significant role in students' dropout. Our research objective is centered on analyzing liable risk drivers in order to come up with the academic factors which have major contribution to the risk occurrence. In detail, our research interest is directed into developing a risk model which will decide on the cardinal risk factors. The risk model will analyze academic data related to students' behavioral engagement and non-academic data related to gender, age and motivation.

Additionally, an econometric study will be carried out to further investigate the role of the non-academic data in students'dropout. The risk model will be mainly developed through a proper analysis of students' engagement data elicited in terms of two blended courses, the e-learning part of which will be implemented by the use of Moodle LMS. The risk model will be assessed in terms of its accuracy (Amare and Simonova, 2021). The risk model will be validated and the verified model could constitute the premise on which an appropriate prediction model could be generated. The prediction model could lead to respective warning system, the enactment of which will contribute to controlling students' attrition. The risk model's development along with the prediction model's generation will be based on the stages of a generic risk management framework (Vose, 2008; Georgakopoulos *et al.*, 2018).

4. Research Methodology

Our method in terms of the risk models' development is based on a potent framework which includes stages of a generic risk management methodology (Vose, 2008; Georgakopoulos *et al.*, 2020; Zakopoulos *et al.*, 2021). In detail, our method includes the below stages:

1. Collecting the academic data in relation to students' engagement in terms of the two blended courses;
2. Collecting the non-academic data in terms of the two blended courses;
3. Analyzing the academic and non-academic data properly in order to come up with a risk model for each blended course;
4. Identifying the classification potential for the two respective risk models. The model with the highest classification percentage will constitute the second independent risk model;
5. Validating the risk model;
6. Generating a prediction model based on the risk model.

The stages 4 and 5 are not included in the context of this paper given that the risk models' validation process is in the pipeline.

The academic data could reflect students' behavioral engagement, setting out in the students' interaction with the learning activities (Bujang et al., 2021; Macfayden and Dawson, 2010; Georgakopoulos et al., 2018; Georgakopoulos and Tsakirtzis, 2021; Zakopoulos *et al.*, 2021; Zakopoulos, 2022; Anagnostopoulos *et al.*, 2020; Alyahyan and Dustegor, 2020). The non-academic data could reflect demographic data such as gender and age, along with students 'emotional engagement (Marks, 2000). In our case, the data related to emotional engagement was reflected by students 'motivation, indicating students 'positive disposition towards the lectures and the entire learning process.

The academic data is collected in terms of the two blended courses. In our case, the e-learning part is implemented by well- orchestrated activities on Moodle. The conventional part consists of a series of well-planned lectures delivered in class. The engagement data related to the conventional part that could be elicited include students' absences, exercises completed in class along with students' participation. In our case, the engagement data in regard to the e-learning part denotes the data that reflect students' interaction with Moodle. Such data include students' logins, students' completion of activities and students' study of supportive material along with the students' grades.

The demographic data is collected by the Moodle log files and other students' repositories. In regard to students' motivation, the Forum utility on Moodle could be used to provide the intended outcome. The students' motivation collection process is described in detail in the methodology application section.

A proper analysis of this data in terms of a binary logistics regression will lead to a risk model for each blended course. Each risk model will decide on the significant risk drivers. A binary variable is modeled to describe students at risk and students not at risk. The binary variable constitutes the dependent variable and the variables modeled to reflect the data collected are the independent variables in our binary logistics regression scheme (Macfayden and Dawson, 2010; Georgakopoulos *et al.*, 2018; Zakopoulos *et al.*, 2021; Anagnostopoulos *et al.*, 2020). The risk drivers for each blended course constitute the factors which are statistically significant in terms of their contribution to the risk occurrence.

The binary logistics regression outcome for each blended course is a classification table through which students are classified into students not at risk and into students who are about to fall through. The correct classification percentage for each risk model indicates the models' accuracy. The model with the highest classification percentage will be selected to constitute the potent risk model. It is important to underline that the data set in regard to students' engagement is collected after the first course-run given that most of the data refer to activities' completion and students usually complete activities at the end of the course.

The potent risk model will be validated in order to rule out the liability of emerging risk factors. Our team is currently working on the risk model's verification. The validated risk model will be used in terms of a proper statistical technique to come up with an appropriate prediction model (Georgakopoulos *et al.*, 2018).

The method used for the econometric estimates is the Least Squares Method (OLS). The least squares method is the most common method of estimating the parameters of a linear system. This method is based on minimizing the sum of the squares of random errors. If we have the linear model:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_k x_{ik} + \varepsilon_i,$$

- y_i, i=1, 2,..., n the values of the observations of the dependent variable y;
- x_{ij}, i=1, 2,..., n, j=1, 2,..., k, the values of the independent (or explanatory) variables x_j, for the i-th observation;
- $\beta_0, \beta_1, \beta_2, \ldots, \beta_k$, the unknown parameters of the model;
- ε_i, i=1, 2,..., n, the random errors, which we assume satisfy following assumptions analogous to the simple linear model;

- o $E(\varepsilon_i) = 0$, for each i;
- o $V(\varepsilon_i) = \sigma^2$, for each i, that is, the random errors satisfy hypothesis of homosexuality;
- o $cov(\varepsilon_i, \varepsilon_j) = 0$, $i \neq j$, ε_i are unrelated.

The method of least squares for estimating the parameters β is based on minimizing performance:

$$S(\beta) = (y - E(y))'(y - E(y)) = (y - X\beta)'(y - X\beta)$$

4.1 Applying the Research Methodology

The potent risk model development process is demonstrated for two courses delivered at the faculty of 'Accounting, Finance and Social Sciences' at the University of West Attica. It is essential to underline that the courses shared the same instructional design. 234 students participated in the first course whereas 144 students participated in the second course. Focusing on the courses' structure, it would be beneficial to explain that both courses were designed in the context of the below common activities:

1. Specific lectures delivered in class;
2. Theoretical Material in form of slides and pdf resources along with Self-assessment quizzes were mounted on Moodle;
3. Final Test (in a form of quiz), mounted on Moodle.

It is essential to underline that a student could use Moodle chat utility during lectures and therefore students' questions could be answered in real time. A Theoretical material, including slides and other pdf resources was mounted on Moodle to help students gain knowledge on the syllabus. In parallel, self-assessment quizzes enabled students to test themselves on the comprehension of the syllabus. It is important to denote that the specific lectures were delivered in class in order to help students make practice. Additionally, students could use Forum utility on Moodle to express questions on syllabus. Students were deemed to pass the course if they achieved a final grade greater or equal to 5.

However, the Forum utility was also used to assess students' motivation. Students were asked to manifest their positive or negative disposition towards the course and the generic Informatics field through a specific question on Forum. We modeled the binomial variable 'motivation' to indicate students who initially had positive disposition and those who initially had negative disposition. The state '0'was modeled to reflect students who had negative disposition and the state '1' was modeled to reflect students who had positive disposition. The question on Forum was mounted before the first lecture in order to assess the students' emotional engagement and the students' motivation extent before assessing their behavioral engagement.

We also modeled the binomial variable student risk to describe students who were about to fail the course as it is suggested in some studies (Macfayden and Dawson, 2010; Georgakopoulos *et al.,* 2018; Anagnostopoulos *et al.,* 2020; Georgakopoulos *et al.,* 2020; Zakopoulos *et al.,* 2021; Zakopoulos, 2022; Tsakirtzis and Georgakopoulos, 2020). The state ‘0’ was modeled to indicate students not at risk whereas the state ‘1’ was modeled to indicate students at risk. We also modeled the below variables with respect to students’ interaction with the learning activities and the entire learning process.

- Number of resources viewed by students (ppt, pdf);
- Number of attended lectures by students (delivered in class);
- Number of self-assessment quizzes completed by students;
- Students’ total logins into Moodle;
- Total number of messages sent by students;
- Total number of exercises completed by students in class;
- Students’ grades on exercises;
- Students’ grades on self-assessment quizzes;
- Final Test grade.

After the final test, the previously cited variables along with the student risk variable and the variables reflecting the non-academic data were employed in terms of a binary logistics regression analysis in order to come up with the risk model. It is also important to denote that the engagement data described by the respective variables were measured two weeks before the final test given that students usually speed up the pace of their study a few weeks before the final exams. In parallel, an econometric study was carried out to further investigate the role of the non-academic data in students’ dropout.

The econometric study was conducted at country level and the data has been elicited from the World Bank for the year 2013 where there were values for all the variables selected for the sample. The countries that make up the sample are divided by income group and are presented in Table 1.

Table 1. *The countries classified by income group*

Country	**Income group**
Austria	High
Brazil	Median
Cyprus	High
Egypt	Low
Greece	High
India	Low
Iran	Median
Italy	High
Sweden	High

Source: *World Bamk, 2013.*

The analysis was performed at country level. All the variables that were modeled were analyzed in the context of the Least Squares Method (OLS) in order to determine the economic factors associated with the dropout of students from the university. The dependent variable (dropno_fem) comes from the World Bank database and is the number of women who have not dropped out higher education. The explanatory variables as well as the dependent similarly derived from the World Bank database and are described in detail in Table 2.

Table 2. *Variables Modeled*

Variable name	**Description of Variables**
grad_fem	The number of female students who graduated from Tertiary (ISCED 6 programmes). This variable in essence shows us the number of women who did not drop out of school and eventually graduated from higher education.
offage_fem	Population of the official age for tertiary education, female (number). This variable shows by country the number of female citizens who attended higher education while they were at the official age for it. It is a factor that can influence dropout.
comp_duration	Duration of compulsory education (years). This variable shows the compulsory years of study per country that according to the existing literature appear to affect the dropout of university mainly for financial reasons.
fem_stud	Percentage of female students in tertiary education enrolled in ISCED 6. It shows the pro will be examined whether or not this affects the drop-out rate of female students.
govexp_stud	Government expenditure per tertiary student. In essence, it shows the funds invested by the state per student. These funds may be a sponsorship or a scholarship, etc.

Source: *World Bank, 2013.*

5. Results

5.1 Binary Logistics Regression Analysis Outcome (Course 1)

The risk model for the first course (Table 3) accounts for 74.3 % (Nagelkerke R Square) of the risk factors denoting that 25.7% of the liable risk factors is not identified. Thereby, there are a small number of factors that could potentially lead to the students' failure which is not identified through the use of our model. This argument is also enhanced by the fact that the Nagelkerke R square value was close to 1 (0.743), denoting a good fit to the results (Allison, 2014; Menard, 2000; Smith and McKenna, 2013).

Another metric for a good model fitness is the Hosmer and Lemeshow Test. In our case, the Sig.Value of the Hosmer and Lemeshow Test was 0.199, greater than 0.05, indicating that the model fits well to the results (Hosmer *et al.,* 2000).

Table 3. *The Regression Model Fitness' Metrics (Course 1)*

Metric	Value
Nagelkerke R Square	0.743
Hosmer & Lemeshow Test (Sig. value)	0.199

Source: *Own study.*

The classification potential of our model is indicated into Table 4.

Table 4. *Classification Percentage (Regression Model- Course 1)*

Overall Classification Percentage	89.3

Source: *Own study.*

The model achieved an 89.3% correct classification percentage (Table 4) indicating that only 10.7% of the cases were not correctly classified, meaning that a small portion of students who were not at risk were classified into students at risk.

Table 5 points out the real risk factors according to the Sig. Value. The column B on Table 5 shows the coefficients that are entered into the regression model.

Table 5. *Coefficients (Regression Model-Course 1)*

Coefficients	B (Coefficient Value)	Sig
Total Logins into Moodle	- 4.666	0.009
Number of Self-Assessment Quizzes completed	- 10.666	0.000

Source: *Own study.*

The real risk factors, having significant contribution to the reduction of the risk probability are the ones, the significance value of which is equal or less than 0.05. Thereby, according to the Sig. column on Table 5, these factors are the 'Total Logins into Moodle' and the 'Number of Self -Assessment Quizzes completed'.

In detail, a unit increase in the "Total Logins into Moodle' leads to a significant decrease (4.666 unit) in the probability of the risk occurrence; whereas a unit increase in the Number of Self- Assessment Quizzes completed, leads to a high decrease (10.666) in the respective probability.

5.2 Binary Logistics Regression Analysis Outcome (Course 2)

The model (Table 6) accounts for 53.1 % (Nagelkerke R Square) of the risk factors denoting that 46.9 % of the liable risk factors is not identified. It is also important to stress the fact that the value of Nagelkerke R square is not so close to 1, indicating a relative fit to the results (Allison, 2014). On the contrary, the 'Sig.value' of the Hosmer & Lemeshow Test is 0.113, greater than 0.05, indicating that the model appears to fit well to the results (Hosmer *et al.*, 2000).

Table 6. *The Regression Model Fitness' Metrics (Course 2)*

Metric	**Value**
Nagelkerke R Square	0.531
Hosmer &Lemeshow Test (Sig. value)	0.113

Source: *Own study.*

The classification potential of our model is indicated into Table 7.

Table 7. *Classification Percentage (Regression Model- Course 2)*

Overall Classification Percentage	81.9

Source: *Own study.*

The second model achieved an 81.9% correct classification percentage denoting that only 18.1 % of the cases were not correctly classified, meaning that a small portion of students who were not at risk were classified into students at risk.

Table 8 points out the real risk factors in terms of the second course according to the 'Sig. value'. The column B on Table 8 shows the coefficients that are entered into the regression model.

Table 8. *Coefficients (Regression Model-Course 2)*

Coefficients	**B (Coefficient Value)**	**Sig**
Number of Self-Assessment Quizzes Completed	-3.429	0.000

Source: *Own study.*

The real risk factors, which have significant contribution to the reduction of the risk probability, according to the Sig. column on Table 8 is the "Number of Self-Assessment Quizzes completed". In detail, a unit increase in the "Number of Self-Assessment Quizzes completed", according to column B on Table 8 leads to a decrease (3.429 unit) in the probability of risk occurrence.

Comparing the two risk models developed for each course, we can conclude that the first model would be selected to constitute the second independent risk model (based on engagement data) given that the first model fits better to the results and it has a better classification potential.

5.3 Econometric Study Outcome

Table 9 shows some descriptive statistics for each income group. It seems that the average number of female students not dropping out high-income countries is higher than in middle-income countries. Correspondingly, the average graduation rate of women in high-income countries is 30% higher than the corresponding average in low-income countries.

Table 9. *Descriptive Statistics Outcome*

HIGH INCOME					
	Observations	**Mean**	**St. Dev.**	**Minimum**	**Maximum**
dropno_fem	5	59325.63	88223.75	1974	215341
offage_fem	5	47135.60	55373.4	32953	1439606
comp_duration	5	9.6	1.34	9	12
fem_stud	5	54.35	5.54	48.08	63.15
MEDIUM INCOME					
	Observations	**Mean**	**St. Dev.**	**Minimum**	**Maximum**
dropno_fem	2	42576.5	28312.6	225558	625945
offage_fem	2	5891092	3074271	3717254	8064930
comp_duration	2	11	4.24	8	14
fem_stud	2	55.44	2.57	53.62	57.26
LOW INCOME					
	Observations	**Mean**	**St. Dev.**	**Minimum**	**Maximum**
dropno_fem	2	1936755	2383684	251236	3622274
offage_fem	2	3.01	3.69	3997265	5.62
comp_duration	2	8.5	0.71	8	9
fem_stud	2	45.84	0.14	45.74	45.94

Source: *Own study.*

The results obtained from the econometric analysis of the sample are presented in Table 9. It seems that the decision of female students to drop out of their studies or not depends on two main factors: the duration of studies and if they are at the official age of study, i.e., 19-25 years old. Specifically, the official age of studies affects this decision at a rate of statistical significance of 1%. It is important to denote that Lassibille and Gomez (2008) report that academic readiness and age at enrollment are strong withdrawal indicators regardless of curriculum.

Other papers also demonstrate the strong link between the formal university enrollment age and the decision to drop out of university, especially for women, for a variety of family, financial and personal reasons (Casanova *et al.*, 2018; Aina *et al.*, 2021). The duration of studies is another factor that overshadows the stay at the university especially for female students. The previous literature supports the same and the reasons are of economic and socio-demographic nature (Ghignoni, 2015; Schnepf, 2014; Bradley and Migali, 2013). Many female students need to work in parallel as they do not have the financial means to make a living or others who are starting a family cannot continue their studies for different reasons each (Table 10).

It should be noted that government funding data are only available for high- and middle-income countries for this and the number of observations is reduced compared to the previous econometric analysis. In contrast to the previous model, the duration of studies does not seem to play a significant role in the second model (negative sign in the variable) in contrast to government funding per student, which has a positive effect at the level of statistical significance of 1%.

This result is in line with the existing literature as according to Bonaldo and Pereira (2016) students who benefit from scholarship or government funding are less likely to drop out than others .In addition, students receiving financial support have lower dropout rates than students receiving no support (Lassibille and Gomez, 2008).

Table 10. *Factors influencing the decision of female students to drop out of university*

dropno_fem	**OLS**
offage_fem	.06476***
	(.0005302)
fem_stud	-247.3113
	(1822.368)
comp_duration	24322.26**
	(4675.548)
R-squared	0.9997
Adj R-squared	0.9996
Prob > F	0.0000
Number of obs	9

Note: ** statistically significant 10%, ** statistically significant 5%, *** statistically significant 1%, Standard errors are described in brackets*
Source: *Own study.*

Table 11 shows the factors influencing the decision of students of both sexes to drop out of university.

Table 11. *Factors influencing the decision of students of both sexes to drop out of university*

dropno_both	**Econometric assessment with OLS**
comp_duration	-233255
	(398924.5)
govexp_stud	162319.3*
	(78463.16)
R-squared	0.5498
Adj R-squared	0.3246
Prob > F	0.2027
Number of obs	7

Note: ** statistically significant 10%, ** statistically significant 5%, *** statistically significant 1%, Standard errors are described in brackets.*
Source: *Own study.*

6. Discussion

After the analysis of students' academic data based on students' engagement, we can deduce that the real risk factors, which have significant contribution to the reduction of the risk probability in the context of both courses, is the "Number of Self-Assessment Quizzes completed". This finding is in line with many studies (Georgakopoulos *et al.*, 2018; 2020; 2021; Zakopoulos *et al.*, 2021; Zakopoulos, 2022).

In parallel, the econometric study has shed light on nonacademic factors that influence the decision of students to leave the university. The econometric study has indicated that women's attendance is influenced by the official age of study and the duration of attendance. In addition, the decision of students of both sexes to leave the university or not is affected only by the government expenditure per student. These findings are also in line with some important studies (Bonaldo and Pereira, 2016; Lassibille and Gomez, 2008).

Nevertheless, it is important to highlight that demographic factor such as gender and age didn't appear to affect the students' critical performance in both courses and also according to the econometric study, these factors didn't appear to affect the students' decision to drop out of university. Finally, the students' motivation, reflecting the students' positive disposition towards the course is not included in the risk factors in the context of both courses.

7. Conclusion

The paper presents a potent framework to control attrition by analyzing academic and nonacademic factors. The paper demonstrates a robust risk model for students at risk based on a proper analysis of data related to students' behavioral and emotional engagement. In parallel, the paper lays emphasis on nonacademic factors that are related to students' dropout by carrying out a specific econometric study. It is important to underline that factors that are related to students' engagement could be deemed to be decisive in the context of our study.

However, more factors that are related to students' emotional engagement should be needed in order to rule out the liability of emerging risk factors. The econometric study proved that governmental financial support could be viewed as a cardinal factor that could potentially deter students from dropping out of university. Nevertheless, a greater sample is needed to claim that the econometrics' findings stand in any case.

Our team is currently working on further examining more nonacademic factors in order to come up with another risk model that is built on the basis of a plethora of factors. This risk model development is in the pipeline. We are expecting that the development of such a risk model will lead to the generation of a proper prediction model and the development of an early warning system which could potentially lead to a full control of attrition.

References:

Abu, S., Adera, B., Kamsani, S.R., Ametepee, L.K. 2012. Addressing the Increasing College Student Attrition Rate by creating Effective Classroom Interaction. Review of Higher Education and Self-Learning, 5(16).

Aina, C. 2013. Parental background and university dropout in Italy. Higher Education, 65(4), 437-456.

Aina, C., Baici, E., Casalone, G., Pastore, F. 2021. The determinants of university dropout: A review of the socio-economic literature. Socio-Economic Planning Sciences, 101102.

Akbarov, A., Gönen, K., Aydogan, H. 2018. Students' Attitudes toward Blended Learning in EFL Context. ActaDidactica Napocensia,11(1), 61-68.

Allison, P.D. 2014. Measures of fit for logistic regression. In: Proceedings of the SAS global forum 2014 conference, 1-13.

Alyahyan, E., Düştegör, D. 2020. Predicting academic success in higher education: literature review and best practices. International Journal of Educational Technology in Higher Education, 17(1), 1-21.

Alsalhi, N.R., Eltahir, M.E., Al-Qatawneh, S.S. 2019. The effect of blended learning on the achievement of ninth grade students in science and their attitudes towards its use. Heliyon, 5(9), e02424.

Amare, M.Y., Simonova, S. 2021. Learning analytics for higher education: proposal of big data ingestion architecture. In: SHS Web of Conferences, Vol. 92. EDP Sciences.

Anagnostopoulos, T., Kytagias, C., Xanthopoulos, T., Georgakopoulos, I., Salmon, I., Psaromiligkos, Y. 2020. Intelligent predictive analytics for identifying students at risk of failure in moodle courses. In International Conference on Intelligent Tutoring Systems, 152-162. Springer, Cham.

Appleton, J.J., Christenson, S.L., Furlong, M.J. 2008. Student engagement with school: Critical conceptual and methodological issues of the construct. Psychology in the Schools, 45(5), 369-386.

Beaubouef, T., Mason, J. 2005. Why the high attrition rate for computer science students: some thoughts and observations. ACM SIGCSE Bulletin, 37(2), 103-106.

Bennett, R. 2003. Determinants of undergraduate student dropout rates in a university business studies department. Journal of further and Higher Education, 27(2), 123-141.

Bonaldo, L., Pereira, L.N. 2016. Dropout: Demographic profile of Brazilian university students. Procedia-Social and behavioral sciences, 228, 138-143.

Bradley, S., Migali, G. 2013. University drop out: An evaluation of the effects of the tuition fee reform in the UK. Unpublished Working Paper, Department of Economics, Lancaster University.

Buchmann, C., DiPrete, T.A. 2006. The growing female advantage in college completion: The role of family background and academic achievement. American sociological review, 71(4), 515-541.

Casanova, J.R., Cervero Fernández-Castañón, A., Núñez Pérez, J.C., Almeida, L.S., Bernardo Gutiérrez, A.B. 2018. Factors that determine the persistence and dropout of university students. Psicothema, 30.

Davidson, P., Long, E., Molnar, A., Chui, T.M. and Ting, C.Y. 2018. MS Teams and Google classroom: Preliminary qualitative comparisons & user feedback. In: 5th Pre-University Sunway Academic Conference.

Dias, S.B., Hadjileontiadou, S.J., Diniz, J., Hadjileontiadis, L.J. 2020. DeepLMS: a deep learning predictive model for supporting online learning in the Covid-19 era. Scientific reports, 10(1),1-17.

Filippidi, A., Tselios, N., Komis, V. 2010. Impact of Moodle usage practices on students performance in the context of a blended learning environment. Proceedings of Social Applications for Life Long Learning, 2-7.

Georgakopoulos, I., Kytagias, C., Psaromiligkos, Y., Voudouri, A. 2018. Identifying risks factors of students' failure in e-learning systems: towards a warning system. International Journal of Decision Support Systems, 3(3-4), 190-206.

Georgakopoulos, I., Chalikias, M., Zakopoulos, V., Kossieri, E. 2020. Identifying Factors of Students' Failure in Blended Courses by Analyzing Students'Engagement Data. Education Sciences, 10(9), 242.

Georgakopoulos, I.,Tsakirtzis, S. 2021. Generating a Model to Predict Secondary School Students at Risk in Mathematics. International Electronic Journal of Mathematics Education, 16(2), p.em0630.

Ghignoni, E. 2017. Family background and university dropouts during the crisis: the case of Italy. Higher Education, 73(1), 127-151.

Goldin, C., Katz, L.F., Kuziemko, I. 2006. The homecoming of American college women: The reversal of the college gender gap. Journal of Economic perspectives, 20(4), 133-156.

Hanushek, E.A., Woessmann, L. 2010. Education and economic growth. Economics of education, 60, 67.

Hosmer, D.W., Lemeshow, S., Cook, E.D. 2000. Applied logistic regression 2nd edition. New York: Jhon Wiley and Sons Inc.

Ismail, A.O., Mahmood, A.K., Abdelmaboud, A. 2018. Factors Influencing Academic Performance of Students in Blended and Traditional Domains. International Journal of Emerging Technologies in Learning, 13(2).

Johnes, G., McNabb, R. 2004. Never give up on the good times: Student attrition in the UK. Oxford Bulletin of Economics and Statistics, 66(1), 23-47.

Lassibille, G., Navarro Gómez, L. 2008. Why do higher education students drop out? Evidence from Spain. Education Economics, 16(1), 89-105.

Macfadyen, L.P., Dawson, S. 2010. Mining LMS data to develop an "early warning system" for educators: A proof of concept. Computers & education, 54(2), 588-599.

Marks, H.M. 2000. Student engagement in instructional activity: Patterns in the elementary, middle, and high school years. American educational research journal, 37(1),153-184.

Menard, S. 2000. Coefficients of determination for multiple logistic regression analysis. The American Statistician, 54(1), 17-24.

Schnepf, S.V. 2014. Do tertiary dropout students really not succeed in European labour markets? Available at SSRN 2409537.

Smith, T.J., McKenna, C.M. 2013. A comparison of logistic regression pseudo R2 indices. Multiple Linear Regression Viewpoints, 39(2), 17-26.

Stinebrickner, R., Stinebrickner, T.R. 2003. Working during school and academic performance. Journal of labor Economics, 21(2), 473-491.

Stinebrickner, R., Stinebrickner, T. 2008. The effect of credit constraints on the college drop-out decision: A direct approach using a new panel study. American Economic Review, 98(5), 2163-2184.

Tsakirtzis, S., Georgakopoulos, I. 2020. Developing a Risk Model to identify factors which critically affect Secondary School students' performance in Mathematics. Journal for the Mathematics Education and Teaching Practices, 1(2), 63-72.

Vose, D. 2008. Risk analysis: a quantitative guide. John Wiley & Sons.
World Bank Database. 2013. Available at: https://data.worldbank.org/.
Zakopoulos, V., Georgakopoulos, I., Kossieri, E, Kallivokas, D. 2021. Controlling Attrition in Blended Courses by Identifying Students at Risk: A Case Study on MS-Teams. International Journal of Finance, Insurance and Risk Management, 9(3), 57-68.
Zakopoulos, V. 2022. A Framework to Identify Students at Risk in Blended Business Informatics Courses: A Case Study on Moodle. International Journal of Economics & Business Administration (IJEBA), 10(1), 239-247.