JMETP
Journal for the Mathematics Education and Teaching Practices, 4(2), 67-76, Dec 2023
e-ISSN: 2717-8587
dergipark.org.tr/jmetp

Young Wise Publishing
youngwisepub.com
gencbilgeyayincilik.com © 2023

Research Article

Investigating factors affecting secondary school non-achievers in mathematics before and after the pandemic

Stylianos Tsakirtzis[1]*, Ioannis Georgakopoulos[2] and Christos Tsifakis[3]

University of West Attica, Greece

Article Info

Received: 12 November 2023
Accepted: 20 December 2023
Available online: 30 Dec 2023

Keywords:
Mathematics education
Risk factors
Risk model
Pandemic
Students' engagement

Abstract

Acquiring knowledge in mathematics is crucial as it serves as a fundamental component for a successful academic journey. However, numerous students encounter formidable challenges, leading to unsuccessful outcomes in their mathematical courses. Therefore, identifying secondary school non-achievers in mathematics is paramount. This necessity was accentuated during the pandemic. Any physical school operation was shut down during this period, leading to an increase in non-achievers. To identify non-achievers before and after the pandemic, we constructed two relevant risk models using a binary logistic regression analysis of student engagement data. The models were applied to a particular mathematical course taught at a Greek Gymnasium. The findings proved that participation in the prescribed written tests was the main factor that affected the performance of non-achievers before the pandemic. Similarly, the risk model developed after the pandemic indicated that the same factor continued to determine student final achievement. However, the positive effect of the same factors (after the pandemic) reducing the probability of students' failure was slightly increased

To cite this article

Tsakirtzis, S., Georgakopoulos, I., & Tsifakis, C. (2023). P Investigating factors affecting secondary school non-achievers in mathematics before and after the pandemic *Journal for the Mathematics Education and Teaching Practices, 4*(2), 67-76.

Introduction

The COVID-19 pandemic has had a significant impact on the education sector globally. Many schools and educational institutions were forced to close and switch to distance education, where students attended classes from their homes (Rapanta et al., 2021). Distance education placed additional demands on students and teachers to have appropriate equipment, internet connection, and technical skills. The lack of social interaction and physical presence at school may have affected the psychosocial student development and the educational experience. All of the above meant that some students may have had difficulties adjusting to distance learning and had reduced performance due to lack of access to resources or difficulties in self-management. In these difficulties, the effort of the teachers to adapt the educational methods they used should be highlighted. Educators have been forced to change their teaching methods to deliver effective distance learning. The upshot was that the pandemic underlined the importance of self-learning, as students needed to gain more independence in learning. The exact impact on student performance varies by region, education system, type of distance learning, and individual factors. Some students adapted better to distance learning, while others struggled. It is important to note that the approach to education during the pandemic is still evolving as situations change

[1] *Corresponding Author: Mathematician. Doctor in Philosophie of Applied Mathematics - Mechanics Division.National Technical University of Athens School of Applied Mathematics and Physical Sciences University of West Attica, Greece. E-mail: s_tsakirtzis@hotmail.com
[2] University of West Attica, Greece. E-mail: igtei@uniwa.gr
[3] Hellenic Mathematical Society, Greece. E-mail: xr.tsif@gmail.com

and educators and students adapt. The COVID-19 pandemic has had a significant impact on the education sector globally (Rapanta et al., 2021). Many schools and educational institutions were forced to close and switch to distance education, where students attended classes from their homes. Distance education placed additional demands on students and teachers to have appropriate equipment, internet connection and technical skills. The lack of social interaction and physical presence at school may have affected the psychosocial development of students and their educational experience. All of the above meant that some students may have had difficulties adjusting to distance learning and had reduced performance due to lack of access to resources or difficulties in self-management. In these difficulties, the effort of the teachers in terms of adapting the educational methods they used was important. Educators have been forced to adapt their teaching methods to deliver effective distance learning. The upshot of all was that the pandemic highlighted the importance of self-learning, as students needed to be more independent in their learning. The exact impact on student performance varies by region, education system, type of distance learning, and individual factors. Some students adapted better to distance learning, while others struggled. It is important to note that the approach to education during the pandemic is still evolving as situations change and educators and students adapt.

Our research interest is directed at identifying factors that affect non-achievers in mathematics after and before the COVID-19 pandemic. These factors are drawn from student engagement. Since student engagement reflects effort (Hopf et al., 2003), our research questions are:

- Did student engagement critically affect the achievement before the pandemic?
- Were factors that critically affected the student outcome before the pandemic identical to those that affected student final achievement after the pandemic?

It is crucial to emphasize that "critical achievement" implies the numeric threshold below which non-achievers are identified. To address our research questions, we formulated two risk models: one aimed at identifying factors influencing students' performance before the COVID-19 pandemic and another to identify factors affecting performance after the pandemic. We employed a binary logistic regression analysis of students' engagement data to construct these respective risk models. These data serve as potential risk factors for students' academic challenges. Each risk model discerns the data with a genuine impact on the occurrence of students' failure, highlighting statistically significant factors. Moreover, the risk models elucidate the contribution of each factor in mitigating the probability of risk occurrence.

To illustrate the development of the risk models, we present a case study centered on a specific mathematics course taught at a Greek Secondary School (Gymnasium). The following sections provide detailed insight into the construction of the risk models and the outcomes of our research.

Literature Review

Factors related to secondary school student achievement in mathematics

In the territory of secondary school student academic accomplishment, many studies associate the learning outcome with engagement (Casillas et al., 2012; Frederick et al., 2004; Marks, 2000; Willms, 2003). Simultaneously, research has established a correlation between secondary school students' achievement and their in-class effort (Hopf et al., 2003). Another critical factor is self-efficacy (McConney & Perry, 2010; Yurt, 2014). Additionally, a separate study highlights the dependence of high school students' achievement on psychological, behavioral, and demographic factors (Casillas et al., 2012). The behavioral factors mentioned are linked to students' engagement in learning activities and the overall learning process, including completed homework and study time. Lastly, the attitude of secondary school students toward mathematics (encompassing both middle school and high school students) has also been identified as a pivotal factor with a significant impact on their performance (Hemmings et al., 2011).

Predicting non-achievers in mathematics

A study has indicated that curriculum-based data can be used to predict non-achievers (Flores & Kaylor, 2007). Other studies have underlined that the teaching approach affects secondary school non-achievers in a mathematical course (Kajander et al., 2008; Xin et al., 2005), accentuating the need for early intervention. Additionally, a multiple regression analysis of students' engagement data has been used in another study to prove that cognitive and behavioral engagement affect secondary school students' failure in mathematics to a greater extent than emotional engagement (Sciarra & Seirup, 2018).

Factors affecting secondary school students' performance during the pandemic

The impact of the COVID-19 pandemic on education has attracted intense interest in the scientific community, with many publications examining the relationship between various factors and student performance during the pandemic. A key factor is the impact of isolation and social exclusion on student performance. Studies, such as that of Smith et al. (2020), report that isolation and lack of social interaction can dramatically affect students' psychological well-being and, by extension, their academic performance. In addition, technological skills and access to devices and online resources have also been examined as essential drivers. Research papers, such as that of Rapanta et al. (2021), have pointed out that lack of access to the necessary technology and online resources can impede student performance during the pandemic. Finally, support from school and family has also emerged as an essential factor. In summary, the research literature indicates that student performance during the COVID-19 pandemic is affected by many factors, including social isolation, access to technology, and school and family support. It is beneficial to continue researching this area to develop policies and practices to support education during similar crises.

Method

According to Vose (2008), risk models are constructed through a general risk management methodology. These models identify non-achievers and indicate the impact of the risk drivers on an unsuccessful outcome. A forecast model can be generated based on such drivers. A verified forecast model could lead to a warning system for students who fail their courses.

In our case, we have developed two risk models, one for the learning process before the COVID-19 pandemic and the other for the learning process after the pandemic. A Binary Logistics regression analysis was used to build the risk models (Georgakopoulos et al., 2018; Macfayden & Dawson, 2010).

Binary logistics regression

Binary logistic regression is a statistical model used to predict the probability of a discrete binary outcome (e.g. 0 or 1, yes or no) based on one or more independent variables. The mathematical (Hosmer et al. 2013) model of binary logistic regression is based on the logistic function (often known as the sigmoid function) which has the form:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 \cdot x_1 + \beta_2 \cdot x_2 + \ldots\ldots + \beta_\rho \cdot x_\rho$$

where:

p: the probability that the event belongs to a category.

β_0, β_1 ...., β_p: adjusted factors of the model.

x_1, x_2, ..., x_p: independent variable

The odds ratio for an independent variable xi is the ratio of the probability of the event belonging to one category to the probability of not belonging (Agresti, 2015):

$$\text{Odds Ratio} = \frac{\frac{p}{1-p}\ for\ x_i = a}{\frac{p}{1-p}\ for\ x_i = b}$$

where a, b are two different values of the variable x_i.

The categorization is determined by a probability threshold, establishing when to predict Category 1 or Category 0. Typically, a threshold of 0.5 is employed, wherein a probability exceeding 0.5 results in a prediction of 1, and otherwise, a prediction of 0 is made. The classification table provides a comprehensive overview of predictions, encompassing True Positive, True Negative, False Positive, and False Negative, facilitating the assessment of model performance, including metrics such as accuracy, sensitivity, specificity, and more. More specifically, the characteristics of the above methods are presented as follows:

Nagelkerke R-squared:

Introduced by Nagelkerke in 1991, Nagelkerke R-squared is a modified variant of the traditional R-squared used in logistic regression modeling. It seeks to quantify how effectively the model elucidates the variability within the response, typically a categorical variable. The Nagelkerke R-squared scale spans from 0 to 1, with 1 denoting a flawless alignment of the model with the data.

Cox-Snell R-squared:

Proposed by Cox and Snell in 1989, Cox-Snell R-squared serves as another metric for assessing the fitness of a logistic regression model. This modified R-squared variant evaluates how well the model conforms to the dataset. Like Nagelkerke R-squared, its scale extends from 0 to 1, with a value of 1 signifying an impeccable fit.

Hosmer-Lemeshow test:

The Hosmer-Lemeshow test, formulated by Hosmer and Lemeshow in 1980, is a statistical test gauging the goodness-of-fit of a logistic regression model. This test compares the model's calculated probabilities with the actual probabilities across different data groups. A low p-value in this test suggests an inadequate model fit to the data. These metrics prove invaluable for evaluating the accuracy and suitability of a regression model, particularly in the context of logistic regression for classification problems.

The performance of a model can be evaluated by various attributes such as accuracy and efficiency. Accuracy is calculated as the ratio of the total number of correct predictions to the total number of examples. Accuracy helps to understand how well the model performs in the general population. Efficiency refers to how quickly and efficiently the model works. This can refer to training time, prediction speed, required memory, or other parameters related to running the model. It is essential to strike a balance between accuracy and efficiency. Often, higher accuracy may require more computing resources, such as computing power, memory, or execution time. The challenge is to offset these two factors to guide model selection, development, and optimization.

Data Collection

We compiled the aggregate engagement data of students from two grades, namely Grade A and Grade B, about a particular mathematics course conducted at a specific Gymnasium. The data were extracted from the official school database, encompassing all 453 students enrolled in the course during that timeframe. It is crucial to emphasize that traditional teaching methods, involving lectures, classroom activities, homework, and exercises, were integral components of the course delivery process. Notably, no part of the course was integrated into a learning management system. The data set collected is listed in Table 1.

In the case of the first risk model, the data set was drawn from 2017 to 2019, whereas in the case of the second risk model, the data set was drawn from 2021 to 2022. It should be explained that the data set during the pandemic is not included since conventional teaching was shut down and given that the research objective is to compare the risk factors before and after the pandemic. However, there is an indirect reference to the effect of the pandemic on student final achievement (deriving from the former year's and the previous year's attendance and participation rates).

Table 1. Data collected

Data	Measured (Time period)
Q1: Lifetime Attendance Rate in the former School Years.	Daily
Q2: Lifetime Attendance Rate in the Previous School Year.	Daily
Q3: Percentage of participation in the prescribed written tests in the former School Years.	Monthly
Q4: Percentage of participation in the prescribed written tests in the Previous School Year.	Monthly
Final Grade (Final Exams' Grade)	Annually

Building the Risk Models

Along with the underlined data shown in Table 1, we constructed the binary variable srisk as the variable describing non-achievers. The value "0" was given for achievers, whereas the value "1" was given for non-achievers (Anagnostopoulos et al., 2020; Georgakopoulos et al., 2018; Macfayden & Dawson, 2010). The final exam's grade defined the numeric threshold for non-achievers. All variables are listed in Table 2. The first column in Table 2 shows the data collected, and the second column indicates the variable's name.

Table 2. Variables modeled

Data Description	Variable Modeled
Lifetime Attendance Rate in the former School Years.	Q1
Lifetime Attendance Rate in the Previous School Year.	Q2
Percentage of participation in the prescribed written tests in the former School Years.	Q3
Percentage of participation in the prescribed written tests in the Previous School Year.	Q4
Final Grade	finalgrade
Students at risk	srisk

We employed this data set in terms of a binary logistics regression analysis (Georgakopoulos et al., 2018; Macfayden & Dawson, 2010) after the final exam to develop the first risk model (before the pandemic). In our scheme, "srisk" was the dependent variable, and the other variables were the independent ones (coefficients). The "finalgrade" variable, describing the final grade (final exams' grade), was only used to determine non-achievers. It is also essential to explain that all independent variables were measured as Scale, whereas the dependent variable "srisk" was measured as Nominal. Additionally, we developed the second risk model (after the pandemic) using the same data set to examine the possibility of identical risk drivers

Results

The result of the binary logistic regression analysis conducted before the pandemic has given rise to Risk Model 1. Table 3 provides insights into the key performance characteristics of our model.

Table 3. Performance characteristics (Risk model 1)

Performance metrics

	Value
Accuracy	0.777
AUC	0.849
Sensitivity	0.882
Specificity	0.561
Precision	0.805

Table 3 highlights the assessment of our model as favorable, given its high scores across various performance metrics domains (Sensitivity: 88.2%; Accuracy: 77.7%; Precision: 80.5%). Notably, special attention is directed towards the precision metric, which signifies the intended classification rate. In our case, the model achieves a classification rate of 77.7% (refer to Table 4).

Table 4. Classification percentage (Risk model 1)

Performance Diagnostics

Confusion matrix

	Predicted		
Observed	0	1	% Correct
0	83	65	56.081
1	36	269	88.197
Overall % Correct			77.704

Note. The cut-off value is set to 0.5

Analyzing Table 3, it becomes evident that the intended classification rate (precision) closely aligns with the actual classification rate (sensitivity). The great specificity percentage vouches for an accurate classification of many non-achievers. However, the same precision does not hold for achievers. Consequently, our model accurately classifies 77.7% of the cases.

It is essential to underline that our model accounts for 43 % of the risk drivers (Nagelkerke R^2), implying that approximately 57 % of the liable risk drivers are not traceable (see Table 5). It is important to stress that the range for Nagelkerke R^2 is between 0 and 1. The value "1" represents a perfect model fit (Allison, 2014; Hair et al., 2006; Smith et al., 2013). Since the Nagelkerke R^2 value for our model is not too close to 1, our model fits the data to a satisfactory but not absolute extent. Therefore, the model accounts for a specific set of risk drivers, but the possibility of new risk drivers cannot be ruled out.

Table 5. Model summary (Risk model 1)

Model	Deviance	AIC	BIC	df	X^2	p	McFadden R^2	Nagelkerke R^2	Tjur R^2	Cox & Snell R^2
H_0	572.433	574.433	578.549	452						
H_1	405.431	415.431	436.011	448	167.002	< .001	0.292	0.430	0.331	0.308

Table 6 shows the coefficients that could be included in the regression model according to the p-value.

Table 6. Coefficients (Risk model 1)

Coefficients

					Wald Test			95% Confidence interval	
	Estimate	Standard Error	Odds Ratio	z	Wald Statistic	df	p	Lower bound	Upper bound
(Intercept)	0.972	0.344	2.643	2.827	7.992	1	0.005	0.298	1.646
Q1	9.471	1.375	12982.402	6.890	47.471	1	< .001	6.777	12.166
Q2	9.866	1.430	19263.576	6.897	47.572	1	< .001	7.062	12.670
Q3	-11.658	1.525	8.654×10^{-6}	-7.647	58.473	1	< .001	-14.646	-8.670
Q4	-10.112	1.507	4.060×10^{-5}	-6.711	45.037	1	< .001	-13.065	-7.159

Note. rgpav10 level '1' coded as class 1.

The factors contributing to student failure are determined by coefficients with a p-value less than or equal to 0.05. Therefore, as per Table 6, in our study, these contributing drivers are the Lifetime Attendance Rate in the former School Years (Q1), the Lifetime Attendance Rate in the Previous School Year (Q2), the Percentage of participation in the prescribed written tests in the former School Years (Q3), and the Percentage of participation in the prescribed written tests in the Previous School Year (Q4). Therefore, our regression model could be given as follows:

Logit(srisk)= 9.471* Q1 + 9.866 *Q2 -11.658 * Q3 – 10.112 * Q4 +0.972

Looking at the estimates in Table 6, we can deduce that if the Lifetime Attendance Rate in the former School Years (Q1) is increased, the logarithm of the probability of student failure is also increased (9.471). The same holds for the Lifetime Attendance Rate in the Previous School Year (Q2) (9.866). However, if the Percentage of participation in the prescribed written tests in the former School Years (Q3) is increased, the probability of student failure is significantly decreased (11.658). This is also true for the Percentage of participation in the prescribed written tests in the Previous

School Year (Q4) (10.112). Therefore, it is essential to point out that although all risk drivers are entered into the regression model (Q1, Q2, Q3, Q4), only Q3 and Q4 factors lead to a decrease in the probability of student failure, constituting real risk drivers. Hence, the Percentage of participation in the prescribed written tests in the former School Years (Q3), and the Percentage of participation in the prescribed written tests in the Previous School Year (Q4) appear to affect students' critical achievement before the pandemic.

The binary logistics regression outcome (after the pandemic) has led to risk model 2. Table 7 sheds light on some cardinal performance characteristics of our model.

Table 7. Performance characteristics (Risk model *2)*

Performance metrics

	Value
Accuracy	0.790
AUC	0.854
Sensitivity	0.907
Specificity	0.462
Precision	0.826

Table 7 highlights the effectiveness of our model, as it attains high scores across nearly every performance metrics domain (Sensitivity: 90.7%; Accuracy: 79%; Precision: 82.6%). Observing Table 7, it is deduced that the intended classification rate (precision) closely aligns with the actual classification rate (sensitivity). The noteworthy specificity percentage indicates an accurate classification of many non-achievers. However, the same precision is not attainable for achievers. Consequently, our model accurately classifies 79% of the cases (see Table 8).

Table 8. Classification percentage (Risk model 2)

Performance Diagnostics

Confusion matrix

	Predicted		
Observed	0	1	% Correct
0	55	64	46.218
1	31	303	90.719
Overall % Correct			79.029

Note. The cut-off value is set to 0.5

Nevertheless, it is crucial to emphasize that our model explains 42.4% of the attributable risk drivers (Nagelkerke R^2), indicating that approximately 57.6% of the potential risk drivers remain unidentified (see Table 9). It is essential to highlight that the Nagelkerke R^2 range lies between 0 and 1, where a value of "1" signifies a perfect model fit (Allison, 2014; Hair et al., 2006; Smith et al., 2013). Given that the Nagelkerke R^2 value for our model is not near 1, it indicates that our model fits the data to a satisfactory but not absolute extent. Therefore, while the model accounts for a specific set of risk factors, the possibility of undisclosed risk drivers cannot be dismissed.

Table 9. Model summary (Risk model 2)

Model	Deviance	AIC	BIC	df	X^2	p	McFadden R^2	Nagelkerke R^2	Tjur R^2	Cox & Snell R^2
H_0	521.725	523.725	527.841	452						
H_1	366.544	376.544	397.124	448	155.180	< .001	0.297	0.424	0.322	0.290

Table 10 shows the coefficients that could be included in the regression model according to the p-value:

Table 10. Coefficients (Risk model 2)

Coefficients

	Estimate	Standard Error	Odds Ratio	z	Wald Test			95% Confidence interval	
					Wald Statistic	df	p	Lower bound	Upper bound
(Intercept)	1.303	0.360	3.681	3.618	13.087	1	< .001	0.597	2.009
Q1	9.663	1.456	15722.831	6.638	44.065	1	< .001	6.810	12.516
Q2	10.263	1.556	28657.450	6.595	43.496	1	< .001	7.213	13.313
Q3	-11.730	1.592	8.051×10^{-6}	-7.369	54.295	1	< .001	-14.850	-8.610
Q4	-10.436	1.619	2.935×10^{-5}	-6.447	41.567	1	< .001	-13.609	-7.264

Note. rgpav12 level '1' coded as class 1.

The factors that have statistically significant contributions to students' failure are derived from coefficients with p-values lower or equal to 0.05. Thereby, according to Table 10, in our case, these drivers are the Lifetime Attendance Rate in the former School Years (Q1), the Lifetime Attendance Rate in the Previous School Year (Q2), the Percentage of participation in the prescribed written tests in the former School Years (Q3), and the Percentage of participation in the prescribed written tests in the Previous School Year (Q4). Therefore, our regression model could be given as follows:

Logit(srisk)= 9.663* Q1 + 10.263 *Q2 -11.730 * Q3 – 10.436 * Q4 +1.303

Looking at the estimates in Table 10, we can deduce that if the Lifetime Attendance Rate in the former School Years (Q1) is increased, the logarithm of the probability of students' failure is also increased (9.663). In parallel, the same holds for the Lifetime Attendance Rate in the Previous School Year (Q2) (10.263). However, if the Percentage of participation in the prescribed written tests in the former School Years (Q3) is increased, the probability of students' failure is significantly decreased (11.730). The same holds for the Percentage of participation in the prescribed written tests in the Previous School Year (Q4) (10.436). Therefore, it is essential to clarify that although all risk drivers are entered into the regression model (Q1, Q2, Q3, Q4), only Q3 and Q4 drivers decrease the probability of students' failure, constituting real risk drivers.

Hence, the Percentage of participation in the prescribed written tests in the former School Years (Q3), and the Percentage of participation in the prescribed written tests in the Previous School Year (Q4) appear to affect students' critical achievement after the COVID-19 pandemic.

Conclusions

Both risk models excel across nearly every performance metric domain (refer to Tables 3 and 7). Additionally, both models attain a high classification rate (refer to Tables 4 and 8). Moreover, both risk models adequately explain a substantial percentage of the identified risk drivers (refer to Tables 5 and 9). The regression outcomes for both models have demonstrated that the Percentage of participation in the prescribed written tests in the former School Years (Q3) and the Percentage of participation in the prescribed written tests in the Previous School Year (Q4) appear to affect students' critical achievement before and after the COVID-19 pandemic, denoting the indirect effect of the pandemic on students' final achievement. It is also essential to underline that in another work, pre-and post-tests appear to affect student performance in mathematics (Flores & Kaylor, 2007). At this point, it is vital to point out that the period in which the prescribed written tests were performed constituted a preparatory stage before the exams. In this spirit, the prescribed written tests acted as pre- and post-tests (Georgakopoulos et al., 2020).

Although the attendance rate in the earlier years is entered into the regression models, it appears to increase the probability of student failure. The poor attendance during the pandemic reduced the effect of these risk factors. However, the participation rate in the prescribed written tests during the pandemic attenuated such drivers (see Tables 6,10). However, more research during the pandemic should be done to generalize these findings.

In an attempt to examine the research questions' validity, we can deduce that factors related to student engagement (participation rate in the prescribed written tests) critically affected student performance before and after the pandemic. The factors that affected student final achievement before the pandemic (Q1, Q2, Q3, Q4) were identical to the ones after the pandemic. It is also essential to underline that the contribution of the prescribed written tests to the reduction of the probability of student failure was slightly increased after the pandemic, denoting that the strength of such drivers was reduced by the pandemic effect.

However, since our models fit the data sufficiently (not completely), the possibility of emerging risk drivers cannot be ruled out. Additionally, more courses are needed to accentuate the similarity of these factors.

Therefore, our research could be expanded as follows:

- Apply our risk models to more courses to rule out the possibility of new drivers.
- To generate a model to forecast non-achievers before and after the COVID-19 pandemic based on the developed risk models.
- To develop warning systems for non-achievers based on the forecast models.
- To further investigate the effect of the pandemic on students' performance, analyzing e-learning data.

In any case, the contribution of our research findings to the field is valuable since our study is based on a published risk management methodology rather than simply using a statistical technique. (Georgakopoulos et al., 2018; Vose, 2008). In parallel, our research findings could be used to mitigate the negative impact of an unsuccessful outcome in mathematics, considering the effect of the pandemic.

References

Agresti, A. (2015). *Foundations of linear and generalized linear models*. John Wiley & Sons.

Allison, P. D. (2014, March). Measures of fit for logistic regression. In *Proceedings of the SAS global forum 2014 conference* (pp. 1-13). Cary, NC, USA: SAS Institute Inc.

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 *Intelligent Tutoring Systems: 16th International Conference, ITS 2020, Athens, Greece, June 8–12, 2020, Proceedings 16* (pp. 152-162). Springer International Publishing.

Casillas, A., Robbins, S., Allen, J., Kuo, YL., Hanson, M. A., & Schmeiser, C. (2012). Predicting early academic failure in high school from prior academic achievement, psychosocial characteristics, and behavior. *Journal of Educational Psychology*, *104*(2), 407.

Cox, D.R., & Snell, E.J. (1989). *Analysis of binary data* (Vol. 32). CRC press.

Flores, M.M., & Kaylor, M. (2007). The Effects of a Direct Instruction Program on the Fraction Performance of Middle School Students At-risk for Failure in Mathematics. *Journal of Instructional Psychology*, *34*(2), 84-94.

Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of educational research*, *74*(1), 59-109.

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., 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.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (1998). Multivariate data analysis. Uppersaddle River. *Multivariate Data Analysis (5th ed) Upper Saddle River*, *5*(3), 207-219.

Hemmings B., Grootenboer P., Kay R. (2011). *International Journal of Science and Mathematics Education*, 9(3), 691-705.

Hopf, D., & Xochellis, P. (2003). Gymnasium and Lyceum in Greece. *Athens: Greek Letters (in Greek).*

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (Vol. 398). John Wiley & Sons.

Hosmer, D. W., & Lemesbow, S. (1980). Goodness of fit tests for the multiple logistic regression model. *Communications in statistics-Theory and Methods*, *9*(10), 1043-1069.

Kajander, A., Zuke, C., & Walton, G. (2008). Teaching unheard voices: students at-risk in mathematics. *Canadian Journal of Education*, *31*(4), 1039-1064.

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.

McConney, A., & Perry, L. B. (2010). Socioeconomic status, self-efficacy, and mathematics achievement in Australia: A secondary analysis. *Educational Research for Policy and Practice*, *9*, 77-91.

Nagelkerke, N. J. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*(3), 691-692.

Rapanta, C., Botturi, L., Goodyear, P., Guàrdia, L., & Koole, M. (2021). Balancing technology, pedagogy and the new normal: Post-pandemic challenges for higher education. *Postdigital Science and Education*, *3*(3), 715-742.

Sciarra, D. T., & Seirup, H. J. (2008). The multidimensionality of school engagement and math achievement among racial groups. *Professional School Counseling*, *11*(4), 2156759X0801100402.

Smith, B. J., & Lim, M. H. (2020). How the COVID-19 pandemic is focusing attention on loneliness and social isolation. *Public Health Research Practices*, *30*(2), 3022008.

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

Sullivan, P., Bragg, L. A. (Ed.), Cheeseman, J., Michels, D., Mornane, A., Clarke, D., Middleton, J., & Roche, A. (2011). Challenging mathematics tasks: What they are and how to use them. 33 - 46. Mathematical Association of Victoria Annual Conference 2011, Melbourne, Victoria, Australia.

Vose, D. (2008). *Risk analysis: a quantitative guide*. John Wiley & Sons.

Willms, J. D. (2003). Student engagement at school. *A sense of belonging and participation. Paris: Organisation for Economic Co-operation and Development*, 1-84.

Xin, Y. P., Jitendra, A. K., & Deatline-Buchman, A. (2005). Effects of mathematical word Problem—Solving instruction on middle school students with learning problems. *The Journal of Special Education*, *39*(3), 181-192.

Yurt, E. (2014). The predictive power of self-efficacy sources for mathematics achievement. *Eğitim ve Bilim*, *39*(176), 159-169.