RESEARCH ARTICLE OPEN ACCESS

ONLINE ACTIVITY OF STUDENTS AND THE IMPACT ON THE EDUCATION

Petar Halachev *
*(Informatics, University of Chemical Technology and Metallurgy,
Sofia, Bulgaria
Email: x3m@mail.orbitel.bg)

----------**********----------

Abstract:

Increasing usage of e-learning in higher education puts the question of assessing its effectiveness. One of the key indicators for evaluation of training is the mark received on the exam. This report examines the dependence of the success of students by the time spent on online training and the number of links to additional literature on subject opened by the learner. The study is performed by a single factor regression analysis and multiple correlation analysis. As a result, a relation is established between factors "time spent on online learning" and "number of visited links" to student assessment exam.

Keywords **—e-learning, effectiveness, on-line activity.**

----------**********----------

INTRODUCTION

The increasingly widespread application of e-learning in higher education raises questions about the assessment of its effectiveness. One of the main indicators of the effectiveness of education is the grade obtained in exams. This report examines the dependence of students' success on factors such as time spent in online learning and the number of open links to additional learning materials accessed by the students. The study is conducted through single-factor regression analysis and multiple correlation analysis. As a result, a relationship between the factors "time spent in online learning" and "number of visited links" and the grade obtained by the student in the exam has been established.

IMPORTANCE OF THE STUDY

A. Online Student Activity

Of crucial importance for the success of the students trained online is the time allocated for their preparation in online learning (measured in hours and recorded by the electronic learning management system), as they mainly self-train from electronic study materials. The number of links visited to additional study materials for a particular discipline demonstrates students' interest in the study material and is a prerequisite for better assimilation and performance in exams [1].

Table 1 provides information on: the sequential number of students, the grade from the exam in "Informatics", the time spent by the respective student in online learning, and the number of links to additional study materials visited by the student.

The data in Table 1 are from the e-learning of master's students at the University of Chemical Technology and Metallurgy. The online activity of 20 students has been tracked, as well as their success at

the end of the semester in the discipline "Informatics". The course workload is 60 hours of lectures and 40 hours of practice. The electronic course on Informatics has 18 links to additional study materials, developed sample programs, etc.

Student №	Grade from the Exam	Hours spent in online learning	Number of links visited
1	4.00	145	15
2	4.25	158	11
3	5.20	170	42
4	4.40	162	20
5	4.33	165	18
6	5.65	180	48
7	4.85	176	25
8	5.55	200	41
9	5.50	190	39
10	4.89	172	34
11	5.85	198	42
12	5.95	201	40
13	6.00	198	48
14	5.20	202	44
15	4.10	139	8
16	4.85	178	19
17	5.75	187	29
18	5.80	194	37
19	5.80	195	42
20	*6.00*	*210*	*46*

Table 1. Data on Students' Grade and Online Activity

The aim is to assess the influence of the factors: time allocated by students in online learning and the number of visited links to additional literature related to the electronic course on students' grades. Through the method of least squares and linear regression trends, the dependence between the variables "Exam Grade" and "Hours spent in Online Learning", as well as between the variables "Exam Grade" and "Number of Visited Links to Additional Literature" is displayed.

B. Single-factor regression analysis

Through single-factor regression analysis, it was found that the dependence between the exam grade (y) and the time spent in online learning (x) is directly proportional, i.e., with an increase in hours allocated for online learning, the grade of the students also increases.

Similarly, the dependence between the number of additional links visited by the student and the exam grade was established.

As a result of the preliminary regression analysis and the established direct proportional dependence between the indicators in this paper, it is proposed that the dependence of the performance indicator

"Exam Grade" on the above-mentioned factors and the extent of their influence on it be evaluated through multiple correlation analysis.

C. Multiple Correlation Analysis

In the multiple correlation analysis [2], the regression coefficients indicate by how many units (according to the adopted measure) the performance indicator will change with a one-unit change in the respective factor. The values of the regression coefficients are determined by solving a system of equations. The number of equations in the system depends on the number of variables. With two factors, the equation of multiple regression takes the form:

$$Y = a + b_1 x_1 + b_2 x_2 (1)$$

As a result, the system will be composed as follows:

$$\begin{aligned} \sum y &= Na + b_1 \sum x_1 + b_2 \sum x_2 \\ \sum y x_1 &= a \sum x_1 + b_1 \sum x_1^2 + b_2 \sum x_2 x_1 \\ \sum y x_2 &= a \sum x_2 + b_1 \sum x_1 x_2 + b_2 \sum x_2^2 \end{aligned} \qquad (2)$$

The strength of the correlation in multi-factor analysis is measured through the partial correlation coefficients and the multiple coefficient.

Partial correlation coefficients characterize the strength (tightness) of the relationship between the performance indicator and a given factor with the influence of other factors eliminated. When the influence of one factor is eliminated, the partial coefficients are of first order; when the influence of two factors is eliminated, the partial coefficients are of second order, and so on. First-order coefficients are calculated from the coefficients of unitary correlation, also known as zero-order coefficients. The formulas take the form:

$$\begin{aligned} r_{yx_1.x_2} &= \frac{r_{yx_1} - r_{yx_2} \times r_{x_1x_2}}{\sqrt{(1 - r_{yx_2}^2)(1 - r_{x_1x_2}^2)}} \\ r_{yx_2.x_1} &= \frac{r_{yx_2} - r_{yx_1} \times r_{x_1x_2}}{\sqrt{(1 - r_{yx_1}^2)(1 - r_{x_1x_2}^2)}} \end{aligned} \qquad (3)$$

The notation for partial correlation coefficient represents the strength of the dependence between Y and X1 with the influence of X2 eliminated.

Second-order coefficients are calculated from first-order partial correlation coefficients using the following formulas:

$$r_{yx_1.x_2x_3} = \frac{r_{yx_1.x_2} - r_{yx_3.x_2} \times r_{x_1x_3.x_2}}{\sqrt{\left(1 - r^2_{yx_3.x_2}\right)\left(1 - r^2_{x_1x_3.x_2}\right)}}$$
$$r_{yx_2.x_2x_3} = \frac{r_{yx_2.x_1} - r_{yx_3.x_1} \times r_{x_2x_3.x_1}}{\sqrt{\left(1 - r^2_{yx_3.x_1}\right)\left(1 - r^2_{x_2x_3.x_1}\right)}} \quad (4)$$
$$r_{yx_3.x_1x_2} = \frac{r_{yx_3.x_1} - r_{yx_2.x_1} \times r_{x_3x_2.x_1}}{\sqrt{\left(1 - r^2_{yx_2.x_1}\right)\left(1 - r^2_{x_3x_2.x_1}\right)}}$$

The tightness of the relationship between two or more factors and the outcome indicator is measured using the coefficient of multiple correlation.

$$R = \sqrt{1 - \frac{S^2_{yx_1x_2\ldots x_n}}{\sigma^2_y}} \quad (5)$$

Where S^2 is the square of the standard error of the estimate, calculated by the formula [3]:

$$S^2_{yx_1x_2\ldots.x_n} = \frac{\sum (y - \overline{y})}{N} \quad (6)$$

Correlation analysis was conducted using the SPSS software [4]. For this purpose, two models were constructed:

- The first model examines the change in the "Exam Grade" indicator under the influence of the variable "Hours in Online Learning".
- The second model considers the influence of the factors "Hours in Online Learning" and "Number of Visited Links to Additional Literature" on the variable "Exam Grade".

As a result of the correlation analysis, the following data were obtained (Table 2):

Information about the model built				
Model №	***R***	***R²***	**Adjusted *R²***	**Standard error in the estimate**
1	0.911 (*a*)	0.829	0.820	0.29211
2	0.937 (*b*)	0.878	0.863	0.25443
a) Predictor variables: Hours in on-line learning *b*) Predictor variables: Number of links visited Dependent variable: Exam grade				

Table 2. Information on the Constructed Correlation Models

- In the "R" column are the values of the coefficients of multiple regression between the predictor variables and the dependent variable. In the first model, the coefficient value is 0.911, indicating the correlation between the predictor variable "Hours in Online Learning" and the dependent variable "Exam Grade".
- The column R^2 shows how much of the variation in the dependent variable "Exam Grade" is accounted for by the predictor variables. For the first model, this value is 0.829, indicating that the

variable "Hours in Online Learning" accounts for 82.9% of the variation in the variable "Exam Grade". In the second model, this value increases to 0.878 or 87.8%, indicating that the variable "Number of Links Visited" accounts for an additional 4.9% (87.8% - 82.9% = 4.9%) of the variation in the variable "Exam Grade".

- The Adjusted R^2 shows the extent to which the constructed model corresponds to the dependencies in the data. The goal is for the values of R and R^2 to be as close as possible. In Table 2, this value is 0.911 - 0.820 = 0.091, or 9%, meaning that if the model were constructed based on the entire population rather than a sample, it would account for 9% less variation in the "Exam Grade" indicator.

Table 2 reflects the parameters of the constructed models. The first model examines the influence only of the first factor [5] [6], "Hours in Online Learning," on the "Exam Grade." In the second model, the influence of the independent variables "Hours in Online Learning" and "Number of Links Visited" on the dependent variable "Exam Grade" is considered. In multiple regression, the model takes the form of an equation containing a coefficient (b) for each predictor variable. Table 3 presents the calculated values of the coefficients b, which indicate the individual influence of each predictor variable on the dependent variable "Exam Grade".

The coefficients calculated using SPSS are presented in Table 3:

Model №	Unstandardized coefficients		Standardized and coefficients *β*	*t*	Level of significance
	b	Standard error			
1.(constant)	-0.494	0.612		-0.807	0.430
Hours	0.031	0.003	0.911	9.348	0.000
2.(constant)	0.956	0.772		1.237	0.233
Hours	0.20	0.005	0.567	3.608	0.002
Links	0.21	0.008	0.408	2.593	0.019

Table 3. Coefficients of the Dependent Variable "Exam Grade"

The coefficients b provides information about the relationship between the "Exam Grade" indicator and each predictor variable. If the value of b is positive, there is a direct proportional relationship between the predictor and the dependent variable. If it is negative, the relationship is inverse proportional. Hence, it can be inferred that increasing "Hours in Online Learning" and the "Number of Links Visited by the Student to Additional Study Materials" lead to an increase in the exam grade. The value of the coefficient b also indicates to what extent each predictor variable influences the dependent variable when the values of all other predictors are held constant.

Each value in the "Standardized Coefficients" column, β, is associated with a standard error, showing to what extent these values vary across different samples from the population [7]. These standard errors are used to determine whether the parameter b significantly differs from zero. If the value in the "Significance Level" [8] column is less than 0.05 in the t-test, this predictor significantly influences the model. The smaller the value in the "Significance Level" column and the larger the value in the t column, the more contribution that specific predictor makes [9]. For the variable "Hours in Online Learning," t = 3.608 and the significance level is 0.002, while for the variable "Number of Links," t = 2.593 and the significance level is 0.019, indicating that these two variables significantly influence the "Exam Grade" variable. From the value of the t-statistic, it can be concluded that the variable "Hours in Online Learning" has a greater influence on the dependent variable "Exam Grade" compared to the variable "Number of Links."

Standardized β values are not dependent on the measurement units of the variables. They can also be used to measure the "importance" of a predictor for the model. Their advantage is that they are directly comparable because they are measured in standard deviation units. Their values are 0.567 for the variable

"Hours in Online Learning" and 0.408 for the variable "Number of Links," again indicating that the first factor has a stronger influence on the dependent variable "Exam Grade."

At each stage of the regression analysis, the SPSS program provides information about variables that have not yet been included in the model.

Table 4 presents information about the variable "Number of Links Visited," which was excluded from the first model. The level of significance of this variable is 0.019, with t = 2.593, indicating that it plays a significant role in the model.

Model№	*β*	*t*	Level of significance	Partial correlation	Collinearity Statistics
1. Links	0.408	2.593	0.019	0.532	0.291

Table 4. Variables, excluded from the first model

CONCLUSION

As a result of the conducted multiple correlation analysis, the following conclusions can be drawn:

1. The factors "Hours spent in Online Learning" and "Number of Visited Links" exert a strong influence on the indicator "Exam Grade".
2. The factor "Hours spent in Online Learning" has a stronger influence on the "Exam Grade" compared to other factors.

In addition to studying indicators of students' grade and pass rates, other outcome indicators (such as the proportion of students working in the field, students continuing to the next academic degree) can be investigated through appropriate mathematical and statistical methods and models. The aim might be to account for the influence of factors causing changes in these indicators.

I. REFERENCES

[1] Wu D., R. Hiltz, Predicting learning from asynchronous online discussions, Journal of Asynchronous Learning Networks, 2004.
[2] Petrov V., T.Todorov, Statistics basics, ABAGAR, 2000.
[3] Kleinbaum, D., L Kupper, K. Muller, Applied regression analysis and other multivariable methods, 2007.
[4] SPSS, http://www.spss.com.
[5] Jim Frost, Introduction to Statistics: An Intuitive Guide for Analyzing Data and Unlocking Discoveries, ISBN-10 1735431109, 2020
[6] David Spiegelhalter, The Art of Statistics: How to Learn from Data, ISBN-10 1541675703, 2021, Basic Books
[7] Sandra Luna McCune, Practice Makes Perfect Statistics (Practice Makes Perfect, ISBN-10 0071638180, 2010, McGraw-Hill
[8] Albert Rutherford, Jae H. Kim PhD, The Art of Statistical Thinking: Detect Misinformation, Understand the World Deeper, and Make Better Decisions, ISBN-13 979-8358180710, 2022
[9] David Borman, Statistics 101: From Data Analysis and Predictive Modeling to Measuring Distribution and Determining Probability, Your Essential Guide to Statistics, ISBN-10 1507208170, 2018

