Determinants of Success in Graduation: An Empirical Study

Marius GIUCLEA[1]
Alexandru AGAPIE [2]
Luiza BĂDIN[3]
Silvia DEDU[4]
Ileana MIRCIOI (VĂLIMĂRANU)[5]
Ovidiu SOLOMON[6]
Rareș BITEȘ[7]

Abstract

One of the most concerning problems that higher education institutions are facing recently is the high dropout rate of students enrolled at all levels of study. Consequently, the dropout phenomenon and the determinants of success in completing university studies are increasingly attracting the attention of both researchers and decision makers.

This paper revisits the problem of identifying the factors accountable for success in higher education and complements the aspects addressed in Agapie et al. (2020) by suggesting a basic set of measures based on Agglomerative Hierarchical Clustering (AHC).

Keywords: *university dropout, principal component analysis (PCA), agglomerative hierarchical clustering (AHC)*

JEL classification: *C38, C83, I23*
DOI: 10.24818/RMCI.2021.4.579

[1] Marius Giuclea, Bucharest University of Economic Studies, marius.giuclea@csie.ase.ro

[2] Alexandru Agapie, Bucharest University of Economic Studies, alexandru.agapie@csie.ase.ro
Gh. Mihoc – C. Iacob Institute of Mathematical Statistics and Applied Mathematics of Romanian Academy

[3] Luiza Bădin, Bucharest University of Economic Studies, luiza.badin@csie.ase.ro
Gh. Mihoc – C. Iacob Institute of Mathematical Statistics and Applied Mathematics of Romanian Academy

[4] Silvia Dedu, Bucharest University of Economic Studies, silvia.dedu@csie.ase.ro

[5] Ileana Mircioi (Vălimăranu), Bucharest University of Economic Studies, ileanamircioi@gmail.com

[6] Ovidiu Solomon, Bucharest University of Economic Studies, ovidiu.solomon@csie.ase.ro

[7] Rareș Biteș, International Computer High School of Constanța, raresbites@yahoo.com

1. Introduction

An increasing number of recent research studies provide evidence that student failure or success is influenced by an interaction between multiple factors including public higher education policies and decision makers involved in the academic process. This implies that the risk of dropout is due to a group of variables and may be regarded a result of a combined impact of factors which are external to the process of education itself.

Tinto and Cullen, who defined the dropout in their early studies in 1973, identified two main categories of dropout: dropping out of college and failing to graduate from any level of education. Both categories of abandonment are the source of a wide range of concerns for not only researchers and policy makers in the field, but also from an economic point of view. According to Kehm et al. (2019), an increasing number of recent studies are attempting to refine Tinto's original approach, taking into account pre-university and intra-university factors, as well as financial situation, family support, opportunities counseling and other external factors related to student individual characteristics, as relevant factors with potential impact on dropout. Debate on a rigorous definition of dropout is beyond the scope of our paper, however we advocate for a more flexible framework, allowing the assessment of the compounded impact of various factors acting at different levels.

Researchers and policy makers provide indications and recommend measures to reduce or prevent early school leaving, such as increasing institutional resources and designing national interventions to improve academic and social integration, motivation, study skills. In this regard, Larsen et al. (2013) consider that any activity aimed at reducing or preventing abandonment can only be successful if it takes into account the underlying motives that determined the decision of abandonment. The HEDOCE study (Vossensteyn et al., 2015) has made a well-grounded contribution to the mapping of existing policy initiatives. However, little is known so far about what would be the most functional and effective measures to stimulate student success.

Norway is one of the most concerned countries analyzing significant data from higher education. The National Statistical Institute of Norway published in 2020 a report showing that 67.5% of new students in Norway have completed a course of study within 8 years. Studies involving OECD countries emphasize that, on average, 12% of students entering a full-time undergraduate program leave the tertiary system before the second year of study with this share increasing to 20% by the end of the theoretical duration of the program.

The graduation rates can be considered an important driver for students' decision in choosing the university and in addition a determinant factor in the process of assessing the quality of teaching, further on contributing to the prestige of the higher education institution. Ivan et al. (2012) state that the school performance is a good predictor for dropout and estimate that the percentage of adolescents who have considered at least once dropping out of school lies between

13% and 20%, the main reasons being the lack of funding or various personal problems. These students are coming especially from rural areas and from families with low incomes, for which the financial problems constitute an important factor in completing university studies. Ivan et al. (2012) found a negative correlation between working hours in the household and average grades and therefore consider that policies to support access to higher education should be focused on these categories.

This paper represents a follow-up study to Agapie et al. (2020) who analyzed several factors with potential impact on academic performance of students. The paper is organized as follows. Section 2 describes the data and Section 3 presents the methodology and discusses the numerical results. The paper ends with a section of conclusions.

2. Data description

Our analysis is based on the responses of a sample of 44 first-year students at Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest University of Economic Studies, to a five-point Likert scale questionnaire. Along the lines of Agapie et al. (2020), the variables were grouped into three categories as follows:

Disciplines
What are the disciplines in which you encountered difficulties?

Q1 *Economics*
Q2 *Algebra*
Q3 *Basics of Statistics*
Q4 *Basics of Information Technology*
Q5 *Basics of Computer Programming*
Q6 *Basics of Operational Research*
Q7 *English / French language and Specialized Communication*
Q8 *Physical Education and Sport*

Difficulties
What are the reasons for the difficulties encountered?

Q9 *Too much information to assimilate in a short time*
Q10 *Different requirements from those in high-school*
Q11 *Difficult communication with instructors*
Q12 *Insufficient pre-university training*
Q13 *Too busy schedule*
Q14 *I had difficulties adapting to online learning*
Q15 *I had difficulties adapting to online evaluation*

Counseling and support opportunities
What type of support do you think would be appropriate to receive from the faculty in order to reduce the difficulties you face and increase your level of performance in exams?

Q16 *Additional consultations on first-year subjects*
Q17 *Group remedial activities*
Q18 *Counseling on the specifics of university life*
Q19 *Counseling on effective study methods and techniques*
Q20 *Counseling on exam stress*

We have processed the responses of the target group and performed a global analysis on a selection of variables according to the suggestion of Agapie et al. (2020), including from the category Disciplines Q2 *Algebra*, Q4 *Basics of Information Technology* and Q5 *Basics of Computer Programming*, which usually pose problems to the first-year students. The other variables selected for the global analysis are Q9, Q10, Q12, Q14, from Difficulties and Q16, Q18 from Counseling and support opportunities.

Table 1 displays the descriptive statistics of the dataset. For missing data, we used modus imputation method. The first row displays the summary statistics for the average score after the first semester, denoted by M.

Descriptive statistics

Table 1

Variable	Minimum	Maximum	Mean	Std. deviation
M	6,000	9,500	7,767	0,849
Q2	1,000	5,000	3,909	1,254
Q4	1,000	5,000	3,568	1,283
Q5	1,000	5,000	3,591	1,300
Q9	2,000	5,000	3,614	0,970
Q10	1,000	5,000	3,068	1,208
Q12	1,000	5,000	2,614	1,450
Q14	1,000	5,000	2,932	1,421
Q16	2,000	5,000	4,250	0,892
Q18	1,000	5,000	2,795	1,193

Source: Authors' own computations

In the next section we describe the methodology and present the numerical results obtained by performing Principal Component Analysis and Agglomerative Hierarchical Clustering (AHC) discussing them with respect to the three categories of variables considered for designing the analysis.

3. Methodology and numerical results

We apply in what follows Principal Component Analysis and Agglomerative Hierarchical Clustering (AHC) to analyze the interaction of the three categories of variables considered: Disciplines, Difficulties and Counseling and support opportunities, respectively.

able 2 below presents the Spearman correlations between the selected variables, showing in bold the significant values (different from 0) at a 5% level.

Correlation Matrix (Spearman)

Table 2

Variables	M	Q2	Q4	Q5	Q9	Q10	Q12	Q14	Q16	Q18
M	**1**	**-0,116**	**-0,181**	**-0,518**	**-0,100**	**-0,061**	**-0,394**	**-0,333**	**0,043**	**0,047**
Q2	**-0,116**	**1**	**0,038**	**0,056**	**0,458**	**0,402**	**0,195**	**0,181**	**0,192**	**0,065**
Q4	**-0,181**	**0,038**	**1**	**0,256**	**0,276**	**0,271**	**0,243**	**0,293**	**0,121**	**0,455**
Q5	**-0,518**	**0,056**	**0,256**	**1**	**-0,013**	**0,232**	**0,579**	**0,364**	**0,066**	**-0,072**
Q9	**-0,100**	**0,458**	**0,276**	**-0,013**	**1**	**0,369**	**0,079**	**0,264**	**0,204**	**0,167**
Q10	**-0,061**	**0,402**	**0,271**	**0,232**	**0,369**	**1**	**0,267**	**0,223**	**0,024**	**0,283**
Q12	**-0,394**	**0,195**	**0,243**	**0,579**	**0,079**	**0,267**	**1**	**0,365**	**-0,072**	**-0,064**
Q14	**-0,333**	**0,181**	**0,293**	**0,364**	**0,264**	**0,223**	**0,365**	**1**	**0,139**	**0,013**
Q16	**0,043**	**0,192**	**0,121**	**0,066**	**0,204**	**0,024**	**-0,072**	**0,139**	**1**	**0,059**
Q18	**0,047**	**0,065**	**0,455**	**-0,072**	**0,167**	**0,283**	**-0,064**	**0,013**	**0,059**	**1**

Source: Authors' own computations.

3.1 Principal Component Analysis

An extraction method based on Principal Component Analysis helps us detecting several relevant factors. The eigenvalues, variation and cumulative variation corresponding to each principal component are displayed in Table 3. We note that the first four factors cumulated, which correspond to eigenvalues greater than 1, explain almost 70% of the information contained in the 10 variables included in the global analysis.

Eigenvalues and Total variation explained

Table 3

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variation	Cumulative %	Total	% of Variation
F1	2.866	28.663	28.663	2.866	28.663
F2	1.776	17.761	46.423	1.776	17.761
F3	1.246	12.461	58.885	1.246	12.461

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings	
	Total	% of Variation	Cumulative %	Total	% of Variation
F4	1.020	10.201	69.085	1.020	10.201
F5	0.738	7.382	76.467		
F6	0.661	6.610	83.077		
F7	0.528	5.278	88.356		
F8	0.491	4.914	93.269		
F9	0.366	3.663	96.932		
F10	0.307	3.068	100.000		

Source: Authors' own computations.

Figure 1 illustrates the result of the scree test (Raymind B. Catell, 1966) for statistical determination of the main relevant factors, which are further retained in the analysis. The corresponding factors are those above the curve that represents the cumulative variation, respectively F1 and F2.

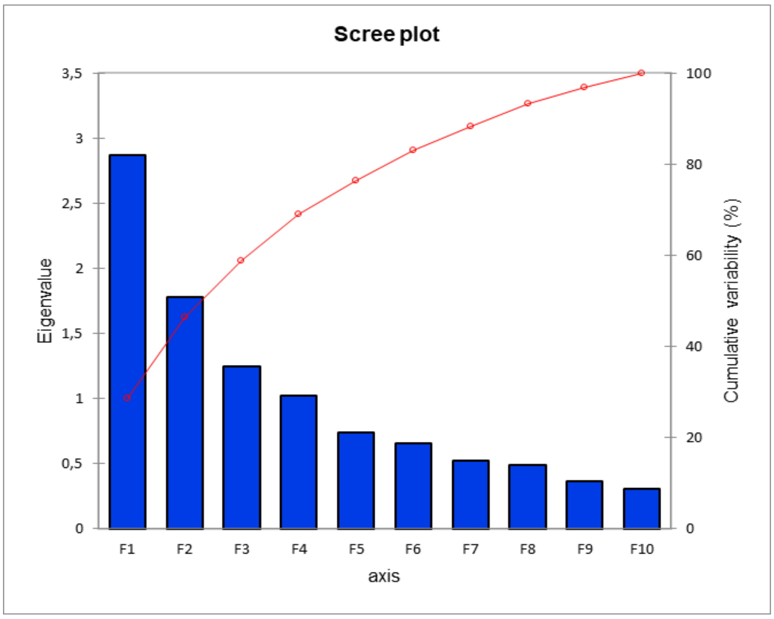

Figure 1. Scree plot

Tables 4-6 display the results obtained after performing the Rotation Method based on Varimax with Kaiser Normalization.

Rotation matrix

Table 4

	D1	D2
D1	0,784	0,620
D2	-0,620	0,784

Correlations between variables and factors after Varimax rotation

Table 5

	D1	**D2**
M	-0,739	0,035
Q2	0,136	0,602
Q4	0,308	0,537
Q5	0,843	-0,021
Q9	0,064	0,740
Q10	0,246	0,653
Q12	0,794	0,065
Q14	0,597	0,296
Q16	-0,051	0,373
Q18	-0,154	0,583

Component Score Coefficient Matrix after Varimax rotation

Table 6

	D1	D2
M	-0,322	0,094
Q2	-0,004	0,275
Q4	0,077	0,226
Q5	0,366	-0,098
Q9	-0,049	0,349
Q10	0,038	0,288
Q12	0,336	-0,052
Q14	0,226	0,080
Q16	-0,061	0,184
Q18	-0,127	0,296

Source: Authors' own computations

From Table 5 we observe that the first principal component is strongly positively correlated with the variables Q5, Q12 and moderately correlated with Q14. The second principal component is strongly positively correlated with Q9 and moderately correlated with Q2, Q4, Q10, Q18.

3.2 Agglomerative Hierarchical Clustering (AHC)

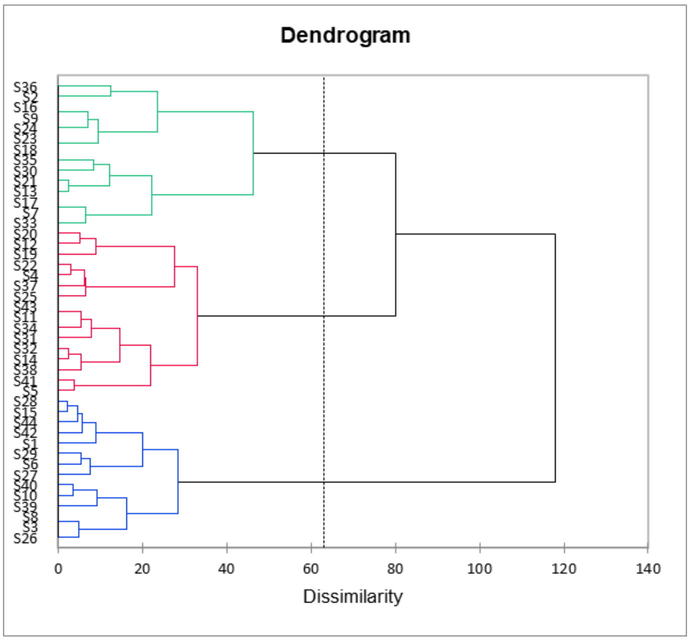

Figure 2. Dendrogram revealing dissimilarity within classes

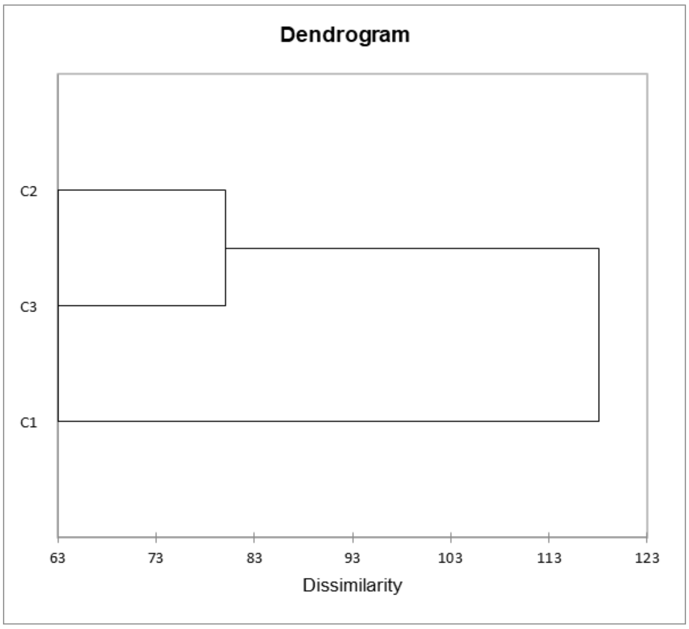

Figure 3. Dendrogram focused on dissimilarity between classes

The following table presents the classes with their corresponding centroids.

Class centroids

Table 7

Class	M	Q2	Q4	Q5	Q9	Q10	Q12	Q14	Q16	Q18
1	8,481	3,929	2,786	2,429	3,429	2,714	1,429	1,643	4,143	2,786
2	7,491	3,000	3,429	4,214	3,000	2,286	3,429	3,000	4,000	2,429
3	7,308	4,688	4,375	4,063	4,313	4,063	2,938	4,000	4,563	3,125

The results reveal a high score assigned to Q16 *Additional consultations on first-year subjects*, indicating a low variability of responses irrespective of the dissimilarity between classes displayed in the case of the rest of variables. Table 8 details the results for each class regarding the sum of weights, within-class variance, as well as minimum, average and maximum distance to centroid.

Results by class

Table 8

Class	1	2	3
Objects	14	14	16
Sum of weights	14	14	16
Within-class variance	9,016	11,584	10,200
Minimum distance to centroid	1,871	1,402	1,760
Average distance to centroid	2,833	3,174	2,987
Maximum distance to centroid	3,917	4,657	4,443

Source: Authors' own computations

The results from AHC analysis uncover and highlight three categories of students corresponding to the three classes revealed by the algorithm.

Class 1 comprises students with a good average mark (8.48 centroid), who encountered difficulties in *Algebra* to a large extent and moderate difficulties in *Basics of Information Technology (BIT)* and *Basics of Computer Programming (BCP)*, the causes being related to the large volume of information and different requirements from high school (moderately). The students in this class did not encounter difficulties in adapting to online teaching or difficulties caused by insufficient pre-university training and consider (to a large extent) that the most appropriate measures are those related to additional consultations, as well as advice on the specifics of university life (to a moderate extent).

Class 2 includes students with a lower average mark (7.49 centroid), who had difficulties in *Algebra* to a moderate extent, largely in *BIT* and at the highest level in *BCP*, the causes being the volume of information, different requirements than in high school, online teaching and insufficient pre-university training. The most appropriate measures indicated by this class of students are those related to additional consultations (to a large extent), as well as advice on the specifics of university life (to a smaller extent).

Class 3 includes students with a slightly lower average mark (7.30 centroid), who encountered difficulties in *Algebra*, to a large extent, and the same in *BIT* and *BCP*. In what concerns the causes, the main factors responsible are the volume of information, different requirements than in high school, difficulties in adapting to online teaching and difficulties caused by insufficient pre-university training. The students in this class consider that the most appropriate measures are those related to additional consultations (to a very large extent), as well as advice on the specifics of university life (to a moderate extent).

4. Conclusions

This paper revisits the problem of identifying the factors accountable for success in higher education and complements the aspects addressed in Agapie et al. (2020) by suggesting a basic set of measures based on AHC analysis. In addition to the purpose of identifying the relevant factors accountable for success in higher education and assessing their impact in increasing the retention rate and successfully completing university studies, policy makers, including institutional authorities, should conjugate their efforts for reducing university dropout rates and developing accompanying measures in this direction at national and European level.

In the long term, all factors of decision should seek to identify good practices in order to substantiate an integrated strategy for the promotion and development of the educational offer at university level. Integrated programs of educational support, tutoring, counseling, coaching, are expected to converge into an integrative mechanism of preventive and remedial interventions adapted to the profile of students, to facilitate their academic and social integration, to increase retention and at the same time to promote and stimulate the level of insertion on the labor market.

Acknowledgement

This paper is a result of research within project ROSE – CSIE *Competitiveness and Success through Learning and Education*. The target group is composed of students of the Faculty of Cybernetics, Statistics and Economic Informatics, Bucharest University of Economic Studies.

References

1. Agapie, A., Bădin, L., Dedu, S., Giuclea, M., Mircioi, I., Solomon, O. (2020), *Analyzing factors accountable for succes in completing university studies*, Review of International Comparative Management, Vol. 21 (3), 423-431.

2. Ivan L., Pricopie, R., Frunzaru, V., Cismaru, D., Corbu, N. (2012), *The Risk of Dropping Education for Romanian High School Students. Structural Factors and Educational Policies*, Procedia - Social and Behavioral Sciences, Vol. 46, 2312-2321, ISSN 1877-0428.
3. Kehm, B. M., Larsen, M. R., & Sommersel, H. B. (2019). *Student dropout from universities in Europe: A review of empirical literature*. Hungarian Educational Research Journal, 9(2), 147–164.
4. Larsen, M. S., Korbeck, K. P., Kristensen, R. M., Larsen, M. R., & Sommersel, H. B. (2013). *Dropout phenomena at universities: What is dropout? Why does dropout occur? What can be done by universities to prevent or reduce it?* Copenhagen, Denmark: Danish Clearinghouse for Educational Research. Retrieved September 26, 2018, from https://edu.au.dk/fileadmin/edu/Udgivelser/Clearinghouse/Review/Evidence_on_dropout_from_universities_technical_report_May_2013.pdf
5. National Statistical Institute of Norway, *Completion Rates of Students in Higher Education—SSB*, https://www.ssb.no/en/utdanning/statistikker/hugjen/aar.
6. Vossensteyn, H., Kottman, A., Jongbloed, B., Kaiser, F., Cremonini, L., Stensaker, B., Hovdhaugen, E., Wollscheid, S. (2015). *Dropout and Completion in Higher Education in Europe (HEDOCE)*. Luxembourg: Publications Office of the European Union.