

# Data Mining for Improving Online Higher Education Amidst COVID-19 Pandemic: A Case Study in the Assessment of Engineering Students

Zoe KANETAKI, Constantinos STERGIOU, Georgios BEKAS, Christos TROUSSAS,  
Cleo SGOUROPOULOU  
*University of West Attica, Athens, Greece*

**Abstract.** Instructional materials, internet accessibility, student involvement and communication have always been integral characteristics of e-learning. During the transition from face-to-face to COVID-19 new online learning environments, the lectures and laboratories at universities have taken place either synchronously (using platforms, like MS Teams) or asynchronously (using platforms, like Moodle). In this study, a case study of a Greek university on the online assessment of learners is presented. As a testbed of this research, MS Teams was employed and tested as being a Learning Management System for evaluating a single platform use in order to avoid disruption of the educational procedure with concurrent LMS operations during the pandemic. A statistical analysis including a correlation analysis and a reliability analysis has been used to mine and filter data from online questionnaires. 37 variables were found to have a significant impact on the testing of tasks' assignment into a single platform that was used at the same time for synchronous lectures. The calculation of Cronbach's Alpha coefficient indicated that 89% of the survey questions have been found to be internally consistent and reliable variables and sampling adequacy measure (Bartlett's test) was determined to be good at 0.816. Two clusters of students have been differentiated based on the parameters of their diligence, communication abilities and level of knowledge embedding. A hierarchical cluster analysis has been performed extracting a dendrogram indicating 2 large clusters in the upper branch, three clusters in the lower branch and an ensuing lower branch containing five clusters.

**Keywords.** CAD, Data analysis, Data mining, Online Learning, Engineering Education, COVID-19.

## 1. Introduction

In the absence of face-to-face classrooms, online learning is based on the use of information technology platforms to ensure that all students have access to learning resources. Asynchronous or synchronous communication is used, with asynchronous system tools such as e-mail, discussion boards, and learning newsgroups being used [1]. Because this epidemic was unexpected, many educational institutions were unprepared for the events that followed [2]. It is necessary to arrive into vital features of internet learning that indicate the method's effectiveness. These characteristics can be interpreted in terms of network education platforms that meet the needs of instructors and learners, remote instruction on effectively completing learning activities, and whether online

education can become a useful tool for time periods where online learning is performed in a fully remote environment [3]. On the other hand, engineering needs have spawned novel learning methodologies at all levels of the educational system. Advanced technology tools, such as digital platforms and social media channels, are brought into the instructional process and serve as the primary means of information transmission [4-9]. Previous research found that a number of course features for adjustment could improve the structure, dialog, and autonomy of the student learning experience, using functionalities carefully created according to Moore's principles of transactional distance [10]. New factors have been revealed to affect online learning in [11] Emergency Remote Teaching Environments [12].

This study was inspired by the expanding use of online-based education in universities around the world, which provided a unique opportunity to collect electronic data and track students' academic progress.

The present study's novelty derives from the fact that pandemic restrictions were just recently imposed on educational institutions, and fully virtual learning environments have yet to be tested.

In the setting of a public health emergency, the current study assesses the learning approach of an online first-year engineering course from the perspective of students, evaluating tasks' assignment via a single learning platform. A statistical analysis has been performed, using data minded from online surveys, that revealed new factors having a significant correlation to the way students perceive tasks assignments during pandemic.

## **2. Module content and description**

The present study has been performed during the first semester of the academic year 2020-2021 in the University of West Attica, School of Engineering, Department of Mechanical Engineering. The related module is called Mechanical Design CAD<sup>1</sup> I. It is a first-semester module and the objectives after attending the course where to be able to create 2-Dimensional sketches and CAD drawings of objects presented in 3-Dimensional views by applying the rules of Mechanical drawing (tracing and reading orthographic projections, views layout, sectional views, dimensioning, symbolic lines). The laboratory lectures had a 3-month duration, and had been performed exclusively online, in MS Teams platform due to pandemic restrictions.

### *2.1. Assigning tasks*

All tasks were assigned in each lecture, including Quizzes, Practice tasks (non-existent 3-Dimensional objects) that had to be represent in 2-Dimensional drawings, as well as assignments related to real world tasks in Mechanical Engineering [13-15] (figure 1).

---

<sup>1</sup> Computer Aided Design



Figure 1. Assigning tasks

### 3. Data mining and filtering

Data have been collected by two online surveys in MS Forms, one pre-course<sup>2</sup> and one post-course<sup>3</sup>. Additional data have been mined out of MS Teams insights, and students' registration information. The total number of participants that attended CAD I during the semester where 216 (N=216) out of which 190 (subpopulation) participated in the study. 165 scores have been considered valid for the purpose of this study, representing first semester students (n=165). 87% of the participants were between 18-23 years old, 11% between 22-25 and the remaining 2% over 25 years old. 88% were male and 12% female.

A matrix of 129\*165 has been created including several constructs [11] of variables that needed to be tested during research.

#### 3.1. Correlation Analysis and discussion

A correlation analysis has been performed in SPSS 2020, among the 37 variables (survey questions) of the construct "assignments". (Table 3) The most significant relationships among all nominal variables are identified using a correlation analysis. The statistical significance is determined by identifying variables with a Spearman's rho coefficient greater than 0.20, indicating a relatively significant connection between two variables. In the following passage, correlations from the Inter-Item Correlation Matrix with a rho coefficient >0.400 are presented, and the conclusions have been schematized in the organogram shown in Figure 2:

The first variable, (Students find the CAD I module more enjoyable than other laboratory modules) is correlated with the synchronous platform familiarization (+0.432), the quality of videos on the supporting YouTube channel (+0.421). Students who find the module enjoyable have expressed lower classroom fatigue (+0.475) and consider that they are very likely to succeed in similar future tasks (+0.518). The enjoyability of the module's attendance has been proven to be highly correlated with their overall evaluation of the learning methodology, on a 10 Likert scale (+0.518). Students that highly evaluated the module find it very well organized (+0.532). Online module's task organization is also correlated with the evaluation of the class notes (+0.448). Tasks' assignment is an indicator of organization and is proven to be correlated with the 12<sup>th</sup>

<sup>2</sup> <https://forms.office.com/Pages/ResponsePage.aspx?id=7kOJDHDDS0u6UTIfQG8y7K165q2QN2ZEUqSfk2d0hHFUNjNBMEJCS1ZFVUg4WVNTOTeWVk4yWE1TSy4u>

<sup>3</sup> <https://forms.office.com/Pages/ResponsePage.aspx?id=7kOJDHDDS0u6UTIfQG8y7K165q2QN2ZEUqSfk2d0hHFUMIEySFBTVVRNOVVKRERTQ0UwTzYxRVM5Ry4u>

variable, of how well assigned tasks are assessed (+0.416). Assignments variety (14<sup>th</sup>) has proven to contribute on understanding the theoretical part (13<sup>th</sup>) and facilitated accomplishing the assignments (+0.431) and is also correlated with the quality of videos (20<sup>th</sup>) presenting the methodology of tasks (+0.426). The 20<sup>th</sup> variable can be considered as an enjoyability factor (+0.421) as well as an indicator of organization (+0.444) also correlated with the variety of tasks in the 14<sup>th</sup> variable (+0.426). Highly significant correlation can be noticed when testing the 15<sup>th</sup> assignment, which is dealing with an existing metallic structure located in the Ancient Olive Campus (of the University of West Attica) and has been shown to be relevant to future professional tasks that students consider very likely to encounter in their careers (+0.513). The variable testing the conception of 3-Dimensional planes, is correlated with the likelihood of succeeding in similar future tasks (+0.474) as well as with the overall evaluation of the module (+0.466).

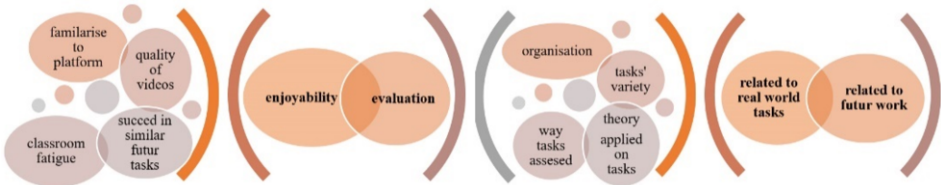


Figure 2. Organogram of correlations

### 3.2. Reliability analysis and discussion

A reliability analysis has been conducted and at the same time Cronbach’s alpha has been calculated, as an estimate of the internal consistency associated with the values derived from the analysis. Cronbach’s Alpha coefficient is indicative of the validity of the ordinal variables filtered out of the correlation analysis.

Table 1. Case Processing Summary

	N	%
Cases	Valid	164 99.4
	Excluded <sup>a</sup>	1 0.6
	Total	165 100.0

a. Listwise deletion based on all variables in the procedure.

Table 2. Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
0.891	0.894	37

As seen in Table 2, it can be estimated that 89% of the variances (37 survey questions) can be considered that concern internally reliable variables, therefore indicating a good internal consistency. The number of students participating in the study accounts to 165 (Table 1) and the valid answers are 164 (N=164). It can be noticed that one student has been excluded from the calculation, since missing values have been detected. In the column that contains Cronbach’s Alpha, if the item of the 15<sup>th</sup> question would be deleted, a coefficient of 0.894 would derive which combined with the result of 0.528 of the corrected item total correlation shows that if this variable would be entirely excluded, the overall coefficient (Table 2) will drop down by 0.004.

**Table 3.** Item-Total Statistics

	<b>Variables</b>	<b>Scale Variance if Item Deleted</b>	<b>Corrected Item- Total Correlation</b>	<b>Cronbach's Alpha if Item Deleted</b>
1	Enjoyable vs other Labs	292.45	0.577	0.886
2	CAD I helped familiarising to MS Teams VS other modules	292.59	0.484	0.887
3	Satisfied CAD I vs other modules	275.48	0.646	0.883
4	Insecure about following	293.20	0.392	0.889
5	Are you comfortable passing the finals	298.15	0.267	0.891
6	Affected by technical issues	296.62	0.320	0.890
7	Data consumption from downloading You Tube support channel videos	288.99	0.413	0.888
8	Assignments' load	299.96	0.279	0.890
9	Hours of studying during the week	295.33	0.343	0.889
10	Are you satisfied with your assignment grades	304.09	0.177	0.891
11	Is the module well organised	298.14	0.349	0.889
12	How well tasks are assessed during online lectures	280.07	0.554	0.885
13	Did the theory contribute on the assignments	300.46	0.452	0.889
14	Assignments variety (quizzes, sketches, CAD drawings)	293.02	0.487	0.887
15	Class notes, download and evaluate	288.15	0.367	0.890
16	Did quizzes help on understanding the theoretical part	295.50	0.484	0.888
17	Have you conceived the meaning of planes	293.97	0.573	0.887
18	cutting planes highlighted in 3d views	297.00	0.383	0.889
19	Have you fully perceived the object	300.67	0.319	0.890
20	Quality of videos concerning the solving methodology	302.54	0.234	0.891
21	Clearness of video, image and sound	284.40	0.508	0.886
22	Social Media applications skills	300.44	0.346	0.889
23	Classroom fatigue	296.36	0.275	0.891
24	Express out loud your questions	281.03	0.572	0.885
25	Questions being solved during synchronous lectures	295.49	0.494	0.888
26	Resent when instructor does not return graded tasks on time	291.37	0.574	0.886
27	Comments by the instruction help understand mistakes	294.57	0.365	0.889
28	Instructor helped meeting new people during synchronous lectures	293.87	0.337	0.890
29	Knowledge Weaknesses	297.81	0.291	0.890
30	Computer skills	298.81	0.308	0.890
31	Tasks assigned relevant to future work	305.38	0.126	0.892
32	Presentation and clarity of the 15th assignment	282.90	0.387	0.891
33	15th assignment related to real world tasks	292.65	0.466	0.887
34	All assignments related to real world tasks	298.70	0.293	0.890
35	Overall evaluation of the module	296.50	0.411	0.888

36	Dealing with weaknesses and knowledge lacks	293.77	0.496	0.887
37	Is it likely to succeed in a similar future task	271.75	0.738	0.881

### 3.3. Sample adequacy

The Kaiser-Meyer-Olkin sampling adequacy measure and the Bartlett's test were used, and the Kaiser-Meyer-Olkin measure was determined to be good at 0.816, while the Bartlett's test of sphericity indicated a significance level of 0.000 (Table 4), which is the highest significance level that can be obtained.

**Table 4.** KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		0.816
<b>Bartlett's Test of Sphericity</b>	Approx. Chi-Square	2166.018
	df	666
	Sig.	0.00

### 3.4. Cluster Analysis and Anova analysis performed on the clusters

The k-means algorithm is used to perform a cluster analysis, which divides the sample into two groups of students (Table 5). Cluster 1 seems to enjoy more and feel more satisfied from the online module than cluster 2. Students belonging to cluster 1, feel less insecure and feel more comfortable about the final exam, are less affected by technical issues and do not resent data consumption from downloading the support You Tube channel videos. Cluster 1 includes more hard-working students who do not feel a heavy load on the tasks assigned, and find the module more organized than cluster 2. A two-score difference is noticed in the way tasks are assessed, with cluster 1 evaluating higher than cluster 2, expressing questions out loud and likely to succeed in similar future tasks.

From the cluster analysis, it can be deduced that dividing the participating students in two clusters, has been based on the diligence of students, their communication skills as well as their level of knowledge embedding, permitting them to believe that they can succeed in similar future tasks.

Table 5 shows the most important criteria with Sig.0.05 for the aforementioned clustering as well as an Anova analysis. Because the Anova analysis evaluates differences between individual clusters and within individual clusters (between groups/within groups comparison), the conclusion achieved is that out of 37 variables, only two variables are considered less distinctive, students' satisfaction with their assignment grades and tasks related to future work.

**Table 5.** Cluster analysis and Anova applied on the clusters

Variables		Cluster		Anova
		1	2	Sig.
1	Enjoyable vs other Labs	5	4	0.00
2	CAD I helped familiarising to MS Teams VS other modules	4	4	0.00
3	Satisfied CAD I vs other modules	9	7	0.00
4	Insecure about following	4	3	0.00
5	Are you comfortable passing the finals	4	3	0.02
6	Affected by technical issues	4	3	0.00

7	Data consumption from downloading You Tube support channel videos	5	3	0.00
8	Assignments' load	4	3	0.00
9	Hours of studying during the week	6	5	0.00
10	Are you satisfied with your assignment grades	4	4	0.28
11	Is the module well organised	4	3	0.00
12	How well tasks are assessed during online lectures	9	7	0.00
13	Did the theory contribute on the assignments	4	4	0.00
14	Assignments variety (quizzes, sketches, CAD drawings)	4	3	0.00
15	Class notes, download and evaluate	3	2	0.00
16	Did quizzes help on understanding the theoretical part	4	4	0.00
17	Have you conceived the meaning of planes	4	4	0.00
18	cutting planes highlighted in 3d views	4	4	0.00
19	Have you fully perceived the object	5	4	0.02
20	Quality of videos concerning the solving methodology	4	4	0.01
21	Clearness of video, image and sound	9	8	0.00
22	Social Media applications skills	4	4	0.00
23	Classroom fatigue	5	4	0.01
24	Express out loud your questions	4	2	0.00
25	Questions being solved during synchronous lectures	5	4	0.00
26	Resent when instructor does not return graded tasks on time	5	4	0.00
27	Comments by the instruction help understand mistakes	4	3	0.00
28	Instructor helped meeting new people during synchronous lectures	4	3	0.00
29	Knowledge Weaknesses	3	2	0.00
30	Computer skills	4	4	0.00
31	Tasks assigned relevant to future work	5	4	0.23
32	Presentation and clarity of the 15 <sup>th</sup> assignment	8	7	0.00
33	15th assignment related to real world tasks	4	3	0.00
34	All assignments related to real world tasks	4	4	0.00
35	Overall evaluation of the module	4	3	0.00
36	Dealing with weaknesses and knowledge lacks	4	3	0.00
37	Is it likely to succeed in a similar future task	9	7	0.00

### 3.5. hierarchical cluster analysis

A hierarchical cluster analysis is performed in SPSS to detect the number of clusters into which, the students can be divided, according to the proximity of their responses. This clustering method uses variance analysis techniques to calculate the distances between clusters. At each stage of the method, the clusters are joined, with the smallest sum of squared errors (SSE), and the sum of squares acts as a criterion of loss.

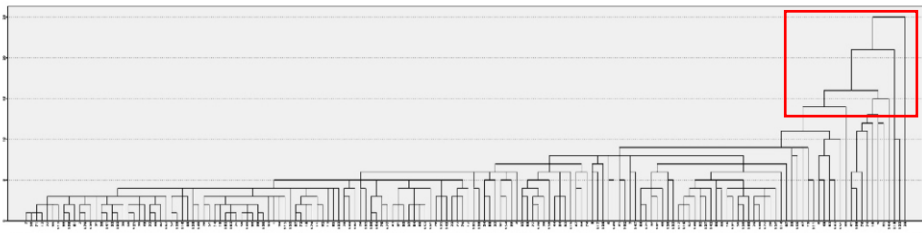


Figure 3. Dendrogram using average linkage (between groups)

The resulting tree diagram (dendrogram) is given below. In the upper branch of the hierarchical clustering tree, it is observed that there are two large clusters and this information is used to implement the k-means algorithm. The next lower branch contains three clusters, and after that there is a lower branch that contains five clusters.

#### 4. Conclusions and Future work

The e-learning paradigm has always included instructional materials, internet accessibility, student interaction, and communication [16]. Most University online courses' lectures and laboratories in universities worldwide and specifically in the presented case of the University of West Attica have been using MS Teams (Communication Platform) for synchronous lectures and Moodle or e-Class (Learning Management Systems) for asynchronous support during the transition from face-to-face to the COVID-19 emerging learning environments.

In this study, the use of MS Teams as a Learning Management System has been used and tested in order to evaluate a single platform use [17] aiming to avoid disruption for students with parallel LMSs functions during the pandemic. Data from online surveys have been mined and filtered through a statistical analysis which included a correlation analysis and a reliability analysis. 37 variables have been considered to have an important significance on testing the performance of tasks' assignment into a single platform, used at the same time for synchronous lectures. The Cronbach's Alpha coefficient has been calculated and resulted to 89% of the survey questions being internally consistent reliable variables. A cluster analysis has been performed dividing students into two clusters, based on the criteria of their diligence, their communication skills, and their level of knowledge embedding.

One of the most noteworthy findings of this research is that having a single online platform for transmitting synchronous lectures while also uploading tasks' assignments, students' evaluation, module notes, and grading, as well as asynchronous support by integrating assignments and resources links uploaded on social media channels makes task organization easier for both students and instructors.

Future work consists on performing a statistical analysis on other constructs of the survey questions, aiming to reveal variables with important significance correlated to students' academic achievements (performance on the final exam), as well as their level of satisfaction, as an indicator of whether online education can become an efficient tool for online learning, or even virtual and blended environments [18].

#### References

- [1] A. Chaudhary, Teaching During Covid-19. *esculapio*. (2021). 16. 10.51273/esc20.25164-guesteditorial
- [2] J. Luo, J. Rajendra, M. Roberts, N. Rapallo, F. Khan, C. Mardon, (2020). Education During COVID-19.
- [3] C. Whittle, S. Tiwari, S. Yan, J. Williams, "Emergency remote teaching environment: a conceptual framework for responsive online teaching in crises", in *Information and Learning Sciences*, (2020) Vol. 121 No. 5/6, pp. 311-319, published on 29/06/2020, <https://doi.org/10.1108/ILS-04-2020-0099>
- [4] C. Troussas, A. Krouska, C. Sgouropoulou, I. Voyiatzis, Ensemble Learning Using Fuzzy Weights to Improve Learning Style Identification for Adapted Instructional Routines, *Entropy* **22**(7) (2020), 735. <https://doi.org/10.3390/e22070735>.
- [5] A. Krouska, C. Troussas, C. Sgouropoulou, Fuzzy Logic for Refining the Evaluation of Learners' Performance in Online Engineering Education, *European Journal of Engineering and Technology Research* **4**(6) (2019), 50-56. DOI: <https://doi.org/10.24018/ejers.2019.4.6.1369>.



- [6] C. Troussas, F. Giannakas, C. Sgouropoulou, I. Voyiatzis, Collaborative activities recommendation based on students' collaborative learning styles using ANN and WSM, *Interactive Learning Environments* (2020). DOI: 10.1080/10494820.2020.1761835.
- [7] C. Troussas, A. Krouska and C. Sgouropoulou, A Novel Teaching Strategy Through Adaptive Learning Activities for Computer Programming, *IEEE Transactions on Education* **64**(2) (2021), 103-109. DOI: 10.1109/TE.2020.3012744.
- [8] C. Troussas, A. Krouska, C. Sgouropoulou, Collaboration and fuzzy-modeled personalization for mobile game-based learning in higher education, *Computers and Education* **144** (2019), 103698. <https://doi.org/10.1016/j.compedu.2020.103698>.
- [9] C. Troussas, A. Krouska and C. Sgouropoulou, Improving Learner-Computer Interaction through Intelligent Learning Material Delivery Using Instructional Design Modeling, *Entropy* **23**(6) (2021), 668. <https://doi.org/10.3390/e23060668>
- [10] M. Moore, *Theory of transactional distance*, Keegan, D., ed. "Theoretical Principles of Distance Education, Routledge, UK, 1997.
- [11] Z. Kanetaki, C. Stergiou, G. Bekas and E. Kanetaki, "Machine Learning and Statistical Analysis applied on Mechanical Engineering CAD course: A Case Study During ERTE Pahse in the Context of Higher Education," 2020 4th International Symposium on Multidisciplinary Studies and Innovative Technologies (ISMSIT), (2020) pp. 1-13, doi: 10.1109/ISMSIT50672.2020.9254924.
- [12] C. Hodges, S. Moore, B. Lockee, T. Trust, A. Bond, The Difference Between Emergency Remote Teaching and Online Learning, *Educausereview* March 27, (2020).
- [13] H.O. Fuchs, Outside reality inside the classroom. *Journal of Engineering Education* 60 (7) (1970): 745–47.
- [14] A. Yadav, G.M. Shaver, P. Meckl, Lessons Learned: Implementing the Case Teaching Method in a Mechanical Engineering Course. *Journal of Engineering Education*. (2010), 99: 55-69. <https://doi.org/10.1002/j.2168-9830.2010.tb01042.x>
- [15] M. Binkley, O. Erstad, J. Herman, S. Raizen, M. Ripley, M. Miller-Ricci, M. Rumble, Defining Twenty-First Century Skills. (2011). 10.1007/978-94-007-2324-5\_2.
- [16] M. Krašna, I. Pesek, "Influence of Moodle and MS Teams on teaching-learning-studying (TLS) processes," 2020 43rd International Convention on Information, Communication and Electronic Technology (MIPRO), (2020), pp. 612-616, doi: 10.23919/MIPRO48935.2020.9245356.
- [17] K. Wilbur, Evaluating the online platform of a blended-learning pharmacist continuing education degree program. *Med. Educ. Online* (2016).
- [18] K. Kabassi, I. Dragonas, A. Ntouzevits et al. Evaluating a learning management system for blended learning in Greek higher education. *SpringerPlus* 5, 101 (2016). <https://doi.org/10.1186/s40064-016-1705-8>