

ICININFO

The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules

Petros Belsis, Ioannis Chalaris, Manolis Chalaris, Christos Skourlas, Anastasios Tsolakidis*

Technological Educational Institute of Athens, Ag. Spyridonos, 12210 Aigaleo, Athens, Greece

Abstract

The length of studies of the students who “linger” in Higher Education has not been justified in many countries, and the Higher Education Institutes try to solve the problem using various methods. The problem of students who “linger” in their Departments beyond the six or seven years is seen as complex one, in the Greek Higher Education. Two main alternative methods have been discussed: Giving the students who “linger” a low priority for registration in the laboratory classes, and limiting the number of times of attending laboratory based courses. Eventually, according to the new legislation the Greek Higher Education Institutes must cut off access to the students who “linger” too long. This study focuses on this hard problem. Clustering techniques and the mining of Association rules are used. The results of clustering and the generation of the association rules are based on students’ questionnaires collected in the laboratory classes. Various interesting results and rules are extracted and discussed.

© 2014 Elsevier Ltd. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/3.0/>).

Selection and peer-review under responsibility of the 3rd International Conference on Integrated Information.

Keywords: Association rules, Data Mining, Higher Education, Student Retention, Student Intervention, Rules’ extraction

The problem of students who “linger” in their Departments beyond the six or seven years is seen as complex one, in the Greek Higher Education. Chalaris et al. focused on the problems of the students that take courses for longer period of time in the framework of the Greek Higher Education (Chalaris et al., 2012). They mentioned that the years required for graduation are more than the normal period of the four years and there are many students that are still “lingering” in their Departments beyond the six years. They also mentioned the related problem of the extremely low percentage of graduates scoring about 8.5/10. The application of classification techniques and the extraction of association rules were based on students’ questionnaires. More specifically, questionnaires filled by 10.000 students in the theoretical classes of the TEI of Athens were collected and some

* Corresponding author. Petros Belsis Tel.: +30-2105910974; fax: +30-2105910975.

E-mail address: pbelsis@cs.teiath.gr

association rules were extracted related to learning goals of the courses, the teaching practices, etc. It was claimed that “these rules and clustering techniques could be used for solving the problem of the students that are still “lingering”, and the problem of the low “scoring” of the graduates”.

In this paper our aim is to describe more precisely the problem of the Length of Studies and investigate solutions that could help the students that are “lingering” in the Greek Higher Education, and especially to increase the chances of completing their studies for the students at the TEI of Athens. In the past two main alternative methods were discussed: Giving the students who “linger” a low priority for registration in the laboratory classes, and limiting the number of times of attending laboratory based courses. Eventually, according to the new legislation the Greek Higher Education Institutes must cut off access to the students who “linger” too long.

The remainder of the paper is organized as follows: In Section 2, we present the methodology and the data used. In Section 3, a new overview of the problems is given, and some new experiments are described, and briefly discussed. The questionnaires used are collected in the laboratory classes of the TEI of Athens. We mainly investigate the lingering problem based on the correlation of various factors to the participation of students in the classes. In Section 4, we discuss our results, present some conclusions and discuss future activities.

1. Methodology

Technological Educational Institution of Athens (TEI of Athens) was founded in 1983 and comprises five Faculties. In the period 2011-12 (two semesters) 50.000 questionnaires were filled by 10.000 students in the TEI of Athens. Almost all the courses of TEI of Athens have two parts: the theoretical part of the course and the laboratory part of the course. Attending the laboratory class is obligatory at the TEI of Athens. There are two types of questionnaires for the evaluation of the two parts of each course. These two types of the questionnaires conform to the proposed ones by the Hellenic Quality Assurance and Accreditation Agency. Chalaris et al. focused on the theoretical part of the courses collected questionnaires that include more than forty questions, and are organized in five sections: course-centred items, lecturer and teaching effectiveness, teaching assistant, laboratory/practical work, and student-centred items. More specifically, they focused on the five different items depicted in Table I (q0-q4), and they tried “to determine the correlations among responses to the related questions” and also “to measure some components of effective teaching” (Chalaris et al., 2012). Table I depicts these five items of the past analysis and the six items of our new one (Q1-Q6). The Q0 item was not considered in the present analysis because it is not included in the questionnaires filled in the laboratory classes. The two new items Q2, Q6 are chosen to measure the influence of the “Use of New Technologies - in the classroom and the labs-”, and the appropriateness of the “Educational material”. More specifically, the usefulness of the educational material is related to the appropriateness of the lecturers’ notes and the recommended literature provided, and the study’s guides. The students evaluated the related questions using the scale 1: agree strongly to 5: disagree strongly. Our analysis, in order to establish the correlations between the six items, is conducting towards two main complementary research directions: Extraction of Association Rules, and use of clustering techniques.

Table I Seven items for conducting analysis

Item	Explanation – Comments
Q0: Use of library	Does the student use the library for further study?
Q1: Aim and objectives of the course	With respect to the future profession were the learning goals of the course clear?
Q2: Educational material	Is the educational material –lecturers’ notes, recommended literature, study’s guides- appropriate and useful?
Q3: Teaching effectiveness of the lecturer	Does the lecturer explain the content of the course in a clear and distinct way?
Q4: Participation in the class of the theoretical part of the course	Do your fellow students actively participate in the class?
Q5: Study	What is the total expenditure of time for the course?
Q6: Use of New Technology	Does the lecturer use new technology in the classroom and the labs?

2. The problem of the Length of studies and some experiments

There is a new dimension of the problem of the Length of Studies in the Greek Higher Education: There are many students that are still “lingering” beyond the six years and according to the new legislation they will not be able to attend lessons after completing 6 years of studies. It means that these students will not have the chance to graduate. In Table II we can see students “lingering” from their 4th to 5th years of studies and students “lingering” for more than 5 years in each Faculty. As we can see the problem is harder in the case of the Faculty of Technological Applications than in the case of the Faculty of Health and Caring Professions and the Faculty of Management and Economics.

Table II The percentage of students that are still “lingering”

Faculty	More than 5 years	Percentage of graduates – Length of study between 4 and 5 years
Faculty of Management and Economics	28.96	52,86
Faculty of Technological Applications	38.1	29.5
Faculty of Fine Arts and Design	30.1	39.62
Faculty of Health and Caring Professions	16.15	68,3
Faculty of Food Technology and Nutrition	34.24	29.15

In the following tables III-V the six items and the calculation of the clusters for three Faculties are depicted. In the case of the Faculty of Health and Caring Professions (see table III) we can see that 7677 students evaluate using 1: agree strongly to 5: disagree strongly. Cluster 0 (C0) includes 574 students (7.47%), Cluster 1 (C1) includes 4207 (54.8%), Cluster 2 (C2) includes 2896 (37.7%).

Table III Three clusters for the Faculty of Health and Caring Professions.

Item	Full	C0	C1	C2
Questionnaires	(7677)	(574)	(4207)	(2896)
Q1: Aim and objectives of the course	3.8875	2.4878	4.3361	3.5131
Q2: Educational material	0.7892	0.453	0.9914	0.5622
Q3: Teaching effectiveness of the lecturer	3.8469	2.9443	4.3021	3.3646
Q4: Participation	4.3487	0.615	4.6575	4.6402
Q5: Study	1.7829	0.9199	1.8621	1.8387
Q6: Use of New Technology	3.0759	2.3937	4.2691	1.4779

Table IV Three clusters for the Faculty of Management and Economics.

Item	Full	C0	C1	C2
Questionnaires	(3474)	(490)	(2107)	(877)
Q1: Aim and objectives of the course	4.0104	3.002	4.2254	4.057
Q2: Educational material	0.6134	0.049	1	0
Q3: Teaching effectiveness of the lecturer	3.9609	2.8122	4.1063	4.2531
Q4: Participation	4.3175	3.6327	4.4514	4.3786
Q5: Study	2.0193	1.8041	2.0679	2.0228
Q6: Use of New Technology	3.6031	1.4041	3.8087	4.3375

Table V Three clusters for the Faculty of Technological Applications.

Item	Full	C0	C1	C2
------	------	----	----	----

Questionnaires	(4530)	(2758)	(564)	(1208)
Q1: Aim and objectives of the course	3.5956	3.9866	2.5656	3.1838
Q2: Educational material	0.6861	1	0.6206	0
Q3: Teaching effectiveness of the lecturer	3.534	4.0004	0.9149	3.6921
Q4: Participation	4.1318	4.442	2.4982	4.1863
Q5: Study	2.0625	2.1856	1.3245	2.1258
Q6: Use of New Technology	3.2031	3.6719	0.6596	3.3204

In the case of the Faculty of Management and Economics (see table IV) we can see that 3474 students evaluate using 1: agree strongly to 5: disagree strongly. Cluster 0 (C0) includes 490 students (14.1%), Cluster 1 (C1) includes 2107 (60.65%), Cluster 2 (C2) includes 877 (25.24%). In the case of the Faculty of Technological Applications (see Table V) we can see that 4530 students evaluate using 1: agree strongly to 5: disagree strongly. Cluster 0 (C0) includes 2758 students (61%), Cluster 1 (C1) includes 564 (12%), Cluster 2 (C2) includes 1208 (27%). Association rules were also extracted to study more the correlation of the items with the participation. Table VI depicts the ‘best’ Association rules extracted from the data of the three Faculties.

Table VI Association rules for the data of three Faculties

<p>Faculty of Health and Caring Professions</p> <p>1. Q3: Teaching effectiveness of the lecturer =(2.5-inf)' Q1: Aim and objectives of the course =(2.5-inf)' 6327 ==> Q4: Participation =(2.5-inf)' 6056 conf:(0.96)</p> <p>3. Q1: Aim and objectives of the course =(2.5-inf)' 6826 ==> Q4: Participation =(2.5-inf)' 6476 conf:(0.95)</p> <p>4. Q1: Aim and objectives of the course =(2.5-inf)' Q7: Educational material=(-inf-2.5]' 6411 ==> Q4: Participation =(2.5-inf)' 6073 conf:(0.95)</p>
<p>Faculty of Management and Economics</p> <p>1. Q3: Teaching effectiveness of the lecturer =(2.5-inf)' Q1: Aim and objectives of the course =(2.5-inf)' 2920 ==> Q4: Participation =(2.5-inf)' 2787 conf:(0.95)</p> <p>2. Q6: Use of New Technology =(2.5-inf)' 2735 ==> Q4: Participation =(2.5-inf)' 2605 conf:(0.95)</p> <p>3. Q3: Teaching effectiveness of the lecturer =(2.5-inf)' 3114 ==> Q4: Participation =(2.5-inf)' 2959 conf:(0.95)</p> <p>4. Q1: Aim and objectives of the course =(2.5-inf)' 3164 ==> Q4: Participation =(2.5-inf)' 2992 conf:(0.95)</p>
<p>Faculty of Technological Applications (all the ten ‘best’ rules are related to the participation)</p> <p>1. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q1: Aim and objectives of the course =(3.333333-inf)' Q7: Educational material=(0.666667-inf)' Q5: Study=(1.666667-3.333333]' 1081 ==> Q4: Participation =(3.333333-inf)' 1002 conf:(0.93)</p> <p>2. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q1: Aim and objectives of the course =(3.333333-inf)' Q5: Study=(1.666667-3.333333]' 1336 ==> Q4: Participation =(3.333333-inf)' 1234 conf:(0.92)</p> <p>3. Q1: Aim and objectives of the course =(3.333333-inf)' Q7: Educational material=(0.666667-inf)' Q5: Study=(1.666667-3.333333]' 1336 ==> Q4: Participation =(3.333333-inf)' 1222 conf:(0.91)</p> <p>4. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q7: Educational material=(0.666667-inf)' Q5: Study=(1.666667-3.333333]' 1269 ==> Q4: Participation =(3.333333-inf)' 1157 conf:(0.91)</p> <p>5. Q1: Aim and objectives of the course =(3.333333-inf)' Q5: Study=(1.666667-3.333333]' 1671 ==> Q4: Participation =(3.333333-inf)' 1520 conf:(0.91)</p> <p>6. Q6: Use of New Technology =(3.333333-inf)' Q1: Aim and objectives of the course =(3.333333-inf)' Q5: Study=(1.666667-3.333333]' 1051 ==> Q4: Participation =(3.333333-inf)' 955 conf:(0.91)</p> <p>7. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q5: Study=(1.666667-3.333333]' 1647 ==> Q4: Participation =(3.333333-inf)' 1493 conf:(0.91)</p> <p>8. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q1: Aim and objectives of the course =(3.333333-inf)' Q7: Educational material=(0.666667-inf)' 1789 ==> Q4: Participation =(3.333333-inf)' 1617 conf:(0.9)</p> <p>9. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q6: Use of New Technology =(3.333333-inf)' Q5:</p>

Study=(1.666667-3.333333)' 1072 ==> Q4: Participation =(3.333333-inf)' 966 conf:(0.9)
 10. Q3: Teaching effectiveness of the lecturer =(3.333333-inf)' Q6: Use of New Technology =(3.333333-inf)' Q1: Aim and objectives of the course =(3.333333-inf)' Q2: Educational material=(0.666667-inf)' 1252 ==> Q4: Participation =(3.333333-inf)' 1127 conf:(0.9)

3. Discussion, Conclusions and future work

The length of studies of the students who “linger” in Higher Education has not been justified in many countries, and the Higher Education Institutes try to solve the problem using various methods. In a previous study Chalaris et al. compared the results of the cluster analysis of two schools, the Faculty of Health and Caring Professions and the Faculty of Management and Economics of the TEI of Athens, and concluded that there is correlation among the use of the library (Q0) and study (Q5) (Chalaris et al., 2012). Also, the participation of the students of this Faculty at the courses seems to be higher than the participation of students at the Faculty of Technological Applications. In the following tables VII-IX the correlation tables for the six items and the related calculations for the three Faculties are depicted. In Table X we could see a comparison of the average values for the questions in the three Faculties.

Table VII. Correlation Table for the Faculty of Health and Caring Professions

Faculty of Health and Caring Professions						
	Teaching effectiveness of the lecturer	Use of New Technology	Aim and objectives of the course	Educational material	Participation	Study
Teaching effectiveness of the lecturer	1	0,35	0,63	0,14	0,17	0,04
Use of New Technology	0,35	1	0,33	0,2	0,07	0,08
Aim and objectives of the course	0,63	0,33	1	0,16	0,21	0,04
Educational material	0,14	0,2	0,16	1	-0,09	-0,02
Participation	0,17	0,17	0,21	-0,09	1	0,02
Study	0,04	0,08	0,04	-0,02	0,02	1

Table VIII. Correlation Table for the Faculty of Technological Applications

Faculty of Technological Applications						
	Teaching effectiveness of the lecturer	Use of New Technology	Aim and objectives of the course	Educational material	Participation	Study
Teaching effectiveness of the lecturer	1					
Use of New Technology		1				
Aim and objectives of the course			1			
Educational material				1		
Participation					1	
Study						1

Teaching effectiveness of the lecturer	1	0,32	0,54	0,08	0,19	0,1
Use of New Technology	0,32	1	0,31	0,06	0,07	0,09
Aim and objectives of the course	0,54	0,31	1	0,09	0,21	0,15
Educational material	0,08	0,06	0,09	1	0,04	-0,04
Participation	0,19	0,07	0,21	0,04	1	0,04
Study	0,01	0,09	0,15	-0,04	0,04	1

In this section, we discuss the experiments of section 3 and the correlation tables in order to present the potential impact of the described method and its results in Higher Education. As the main conclusion, we claim that the problem of students who “linger” in their Departments beyond the six or seven years in Greek Higher Education could be solved by a mixed technique based on the use of clustering and the extraction of association rules. The rules and the clustering techniques offer to the experts a tool for investigating the problem. The rules imply that students’ participation in the class is closely related with two factors: 1) The lecturer explains the content of the Course in a clear and distinct way and 2) The learning goals of the Course are clear with respect to the future profession. Such ‘best’ rules appeared in the case of the Faculties, and it means that these factors form a condition, which is sufficient for higher rates of participation in the class. Focusing on the tables III-V, and specifically, in the cluster with the higher percentage per school, we should conclude that in the cases of the Faculty of Health and Caring Professions and the Faculty of Management and Economics there is correlation among the Participation (Q4) and Aim and objectives of the course (Q1), Teaching effectiveness of the lecturer (Q3) and Study (Q5). In the case of the Faculty of Technological Applications Participation is also related to the study (Q5), specifically, in cluster C1. It is an indication that students participating in the classes are the students that study more. Also, Tables show that the participation of the students of the Faculty of Health and Caring Professions and the Faculty of Management and Economics at the courses seems to be higher than the participation of students of the Faculty of Technological Applications. The analysis of the correlation matrix indicates that few of the observed relationships were very strong. The strongest relationship exist among "teaching effectiveness of the lecturer" and the "aims and objectives of the course" for the Faculties. This observation could lead to the conclusion that lecturers with teaching effectiveness can make clear to students the aims and the objectives of the course. We can also reach the same conclusion based on cluster analysis, which indicates that more than 70 % of the students belong to the clusters, which achieve high score at these questions. From Table X it seems that the problem of “lingering” in TEI of Athens is not associated with the teaching effectiveness of the lecturers but is correlated with the time that students spend for studying. As we could see in Table X the questions that are related with the lecturer achieve high scores in contrast with the questions for the "Study", which has value from 2.14 to 2.04. Another correlation (0.35-0.32) is observed among "The use of new technologies" and the "teaching effectiveness of the lecturer". We could conclude that the importance of using new technologies in the class has rather positive effects at the teaching effectiveness. The participation in the classroom is slightly correlated with both the teaching effectiveness of the lecturer, and the "aims and objectives of the course", implying that slightly more students attend in the classrooms when the lecturer is effective and the aims and the objectives of the course are clear. From the analysis of the association rules we could conclude that there is a difference between the students of different Faculties. For example, the participation in the classroom in the Faculty of Technological Application could be achieved (93%) when the students have also high scores at the other four questions (Q1, Q2, Q3, Q5). In the other two Faculties the high level of participation is mainly ‘influenced’ by two factors/questions (Q1, Q2). Such an observation shows that it is rather more complicated to

achieve higher participation of the students of the Faculty of Technological Applications than in the case of the other two Faculties.

Table IX. Correlation Table for the Faculty of Management and Economics

Faculty of Management and Economics						
	Teaching effectiveness of the lecturer	Use of New Technology	Aim and objectives of the course	Educational material	Participation	Study
Teaching effectiveness of the lecturer	1	0,35	0,61	0,14	0,25	0,07
Use of New Technology	0,35	1	0,32	0,13	0,14	0,11
Aim and objectives of the course	0,61	0,32	1	0,12	0,27	0,08
Educational material	0,14	0,13	0,12	1	0,05	0,07
Participation	0,25	0,14	0,27	0,05	1	-0,03
Study	0,07	0,11	0,08	0,07	-0,03	1

In the near future, we shall conduct further analysis based on more factors of the questionnaire. We shall also examine the influence of organisational culture and job satisfaction on the quality of services provided in Higher Education (Trivelas and Dargenidou, 2009), and shall try to combine the results with further studies based on strategies for attracting students in the class. We shall also conduct a semiautomatic way in order to discover useful association rules “focusing the search by predefining what should appear on the right hand side and what may appear on the left hand side of a rule” (Pechenizkiy et al.). A similar approach to the proposed by the Oprea, Zaharia, and Enachescu will be adapted in our research. They used data mining techniques to identify the profile of students who have difficulties adapting in first year of college (Oprea et al.).

Table X. Comparison of the average values for the questions in the three Faculties

	Teaching effectiveness of the lecturer	Use of New Technology	Aim and objectives of the course	Educational material	Participation	Study
Faculty of Management and Economics	4,15	3,88	4,17	0,73	4,5	2,14
Faculty of Health and Caring Professions	3,96	3,28	4	0,78	4,56	2,05
Faculty of Technological Applications	3,53	3,2	3,6	0,68	4,13	2,06

References

Chalaris M., Chalaris I., Skourlas C., Tsolakidis A. (2013) Extraction of Rules based on Students' Questionnaires, *Procedia - Social and Behavioral Sciences*, 73, 510–517

- Trivellas P., Dargenidou D (2009) Organisational culture, job satisfaction and higher education service quality: The case of Technological Educational Institute of Larissa, *The TQM Journal*, 21, 4, 382–399
- Pechenizkiy, M., Calders, T., Vasilyeva, E., De bra, P. (2008) Mining the Student Assessment Data: Lessons Drawn from a Small Scale Case Study International Conference on Educational Data Mining, Montreal, 187-191
- Oprea C., Zaharia, M., Enăchescu, D. (2010) Knowledge discovery using data mining techniques: A case study *Annals of the Oradea University. Fascicle of Management and Technological Engineering* vol. IX, 543-550 ISSN 1583 - 0691