Business Intelligence supporting the teaching-learning process

MARIA BEATRIZ PIEDADE
School of Technology and Management, Polytechnic Institute of Leiria
Campus 2, Morro do Lena,2411-901 Leiria , PORTUGAL
bea@estg.ipleiria.pt
MARIBEL YASMINA SANTOS
Information Systems Department, Algorithmi Research Centre, University of Minho
Campus de Azurém, 4800-058 Guimarães, PORTUGAL
maribel@dsi.uminho.pt

Abstract: -. One of the activities usually pointed out as crucial to promote the students' success is the closely monitoring of the students' academic activities related with the *teaching-learning* process. Although important this activity does not take place in many higher education institutions. To support this complex process an adequate, conceptual and technological support is needed. In this sense, it is presented in this paper a *Student Relationship Management* (*SRM*) *system*. The *SRM system* supports the SRM *concept* and *practice* and is based in a *Business Intelligence* infrastructure. To demonstrate the *SRM system* relevance in the process of acquisition of knowledge about the students and theirs academic behaviour, in the support of the *decision-making* associated to the *teaching-learning* process and in the automatic interaction with the students, it is presented an application case, which was carried out in a real context.

Key-Words: - Business Intelligence, Data Warehouse, Data Mining, OLAP, Student Relationship Management

1 Introduction

Portuguese Higher Education has been characterized by a high rate of failure and abandon, mainly in the first year of the graduation courses [1]. Although this reality has changed due to several institutional actions, integrated activities need to be proposed and adopted. To be possible, it is necessary the identification of the factors and the measures that need to be monitored in the *teaching – learning* processes and in the *student-teacher* relationship. One of the activities usually pointed out as crucial to promote the students' success is the closely monitoring of the students' academic activities. Although important, this activity does not take place in many higher education institutions. Among the reasons, we point out the huge number of students with failure in the first graduation year, the huge number of new students in some courses and the work overload of the teaching staff. In the Portuguese higher education institutions, teachers are involved in lecturing, researching and management tasks. To help teachers and students in this complex process, an adequate, conceptual and technological support is needed. The conceptual framework and the technological infrastructure are in this work integrated in a *Student Relationship Management* (SRM) *System*. The system supports the *SRM concept* and *practice* and is based in a *Business Intelligence* infrastructure. To demonstrate the SRM system relevance and usability in the students' knowledge acquisition process, in the early identification of failure situations, in the decision-making support and in the automatic interaction with the students, this paper presents an application case that has occurred in a real context. It includes data gathering, data analysis, results interpretation and interaction. This paper is organized as follows: Section 1, refers the academic failure as the problem that motivates and justifies the *SRM system*; Section 2 includes an overview of the SRM principles and presents the *SRM concept* and *practice* (as understood in this work), and the methodology used to their validation; Section 3 describes the *SRM system* architecture and gives some details about the implementation; Section 4 describes the application case and the data analysis process using *OLAP* (*On-line Analytical Processing*) and *data mining* tools. This section also presents the interpretation of the obtained results and their support to interaction activities with the students; section 5 concludes with a summary of the undertaken work, the SRM system expected advantages and the upcoming future work.

2 Conceptual framework

The *SRM system* was inspired in principles underlying to the *CRM* (*Customer Relationship Management*) systems, which are used nowadays used in a business contexts to support and to manage processes related with the customers'. These systems allow the identification of knowledge about the customers, using the information and business transactions available in the organization databases. Using this *knowledge*, the organization defines the activities and actions that allow maintaining a close and strong relationship with its customers [2] . The *SRM system* is based on these principles, but supports processes and activities concerned with the *teaching-learning* process, mainly activities that allow the monitoring and the supervision of the students' academic activities. Associated to the *SRM* concept is the scholar success promotion, as it is generally accepted that exists a high correlation between the closely monitoring of the students and their academic success. To exemplify the similarity between the *CRM/SRM* actions, one could compare the actions developed by the customer's manager, that on the scope of the banking activity, alerts the customer when he/she exceeds his/her credit account, and the actions developed by the student's tutor/teacher, that on the scope of the monitoring processes send an alert message to the student when detects he/she misses several lessons. To refer that both the "*Student Relationship Management*" and the *"SRM"* designation, were already used in a technological/commercial environment to designate solutions mainly dedicated to support processes related with the students in academic areas (students' management information, courses management, admissions management, enrolment and registration management) and areas related with the institution available services (communications, marketing, financial aids, accommodation). The proposal of the *SRM concept*, in the scope of this work, was focused on the students' academic success promotion. The *SRM concept* is understood as a process based on the students' acquired knowledge, whose main purpose is to keep an effective *student-institution* relationship through the closely monitoring of the students' academic activities. This concept, and as already stated, is base on the premise that there exist a strong correlation between the closely monitoring of the students' academic activities and their academic success promotion. The *SRM practice* is understood as a set of activities/actions, which should guarantee the student's individual contact, and an effective, adequate and closely monitoring of his/her academic performance. To validate the *SRM concept* and the set of activities included in the *SRM practice* it was adopted a research methodology based on the *Grounded Theory* principles, which included the interviews realization. The selected interviewers were teachers with institutional responsibilities (courses' directors, institution directors', council members). Each interview was recorded, transcribed and analyzed. The analysis process was done following the *Grounded Theory* principles and supported by the *NVivo* software (a *Computer Assisted Qualitative Data Analysis Software*). Each interview was guided by a script, prepared beforehand, including also open questions (semi-structured interviews). The interview question included topics like: academic success/failure, activities to promote the success; *student-institution* relationship, practices to maintain an effective *student-institution* relationship, relevant monitoring indicators, behavior patterns, and identification of activities that can be automatically supported, among others [3].

3 SRM System

The *SRM system* is the technological infrastructure that supports the *SRM concept* and *practice*. To undertake an S*RM practice* it is necessary: (i) to have adequate, consistent and complete information about the students. This information must be stored in an appropriate data repository, which allows maintaining a single vision of students' data; (ii) the analysis of such data in order to obtain knowledge about the students' and their academic behavior; (iii) the starting of automatic actions whenever specific situations are detected; and (iv) to assess the impact of all the implemented actions. The structural issues related with the data repository (*data warehouse*) and the data analysis tools envision that the *SRM system* must be implemented using the concepts and the technological infrastructure that traditionally support the *Business Intelligence System* [4]. The *SRM system* architecture includes four main components: the *Data Acquisition and Storage* component; the *Data Analysis* component; the *Interaction* component and the *Assessment* component (Figure 1). The *Data Acquisition and Storage* component is responsible for storing the students' data in a *data warehouse*, which structure was modeled for this purpose. The students' data exist in different data sources. All the data is stored in the *data warehouse* after the ETL (*Extraction*, *Transformation* and *Loading*) process. The *Data Analysis* component is responsible for obtaining knowledge about the student/students. The stored data is analyzed using appropriate data analysis tools, allowing patterns identification. The obtained knowledge is stored in an adequate data repository (*knowledge database*). The *Interaction* component is responsible for maintaining an adequate and effective relationship with the students, using the obtained knowledge. The system must allow the definition and the automatic execution of adequate actions over the student/students. The *Assessment* component is responsible for the assessment of all the actions carried out and their impact, by monitoring the students' academic behavior, verifying different rates (assiduity, marks, among others). The *SRM system* prototype implementation was done using database management tools, *Business Intelligence* tools and *web* development tools. Considering the context in which this project takes place, the selected development tools integrate the *Microsoft* environment: *SQL Server Business Intelligence Development Studio* (*Database Engine*, *Integration* and *Analysis Services*) and *Visual Studio .NET* [5]. The *SRM system* prototype validation has been done through the execution of a set of application cases, taking place in different Higher Education Institutions.

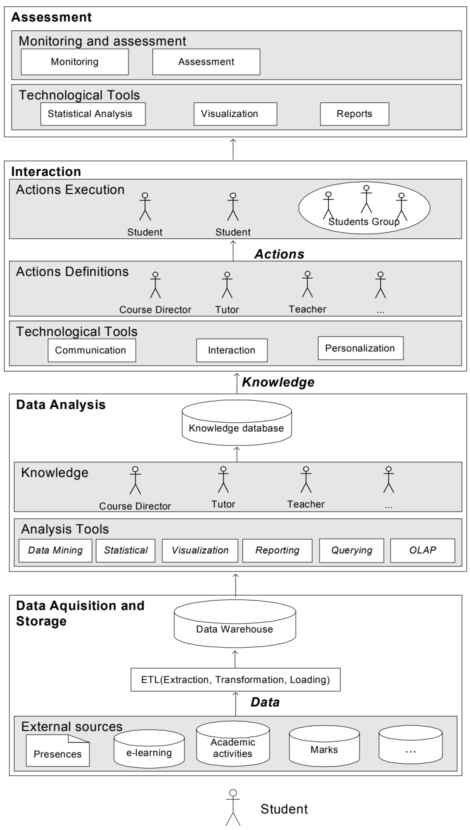

Fig 1 - The SRM system's architecture

4 Application Case

The data selected in this application case was gathered in a graduate course of a Portuguese Higher Education Institution. This course, in the engineering area is composed by a set of curricular units. The selected unit, with 70 students, is integrated in the first year. The teaching method adopted is based in a presential component, and on a component supported by the *e-learning* platform available in the institution (e-learning component). The presential component includes different types of classes (theoretical, practical and tutorial orientation). The activities included in the theoretical classes are related with the curricular subjects' presentation and explanation; in the practical classes are related with exercises and problems solving in a laboratory environment; and in the tutorial orientation classes' activities include students' individual work support and orientation and subject clarification. The *e-learning* component includes activities related with the distribution of relevant information and materials (unit general information, curricular contents, exercises and project guidelines), and, also, communication activities (messages that are automatically or manually sent and discussion forums). The unit assessment includes two distinct methods: the normal assessment period and the exam assessment period. The normal assessment period integrates a written test and a project with individual discussion, with weights of 40% and 60%, respectively, in the final mark. The quantitative mark scale comprises values among 0 to 20. In both assessments, 8 is the minimum mark that needs to be obtained by the students in order to be possible the calculus of their final mark. This final mark results from the application of the weights associated with the test and the project and needs to be equal or higher than 10 to guarantee success in the unit. The exam assessment period only comprises a written exam. To pass the unit, the student must obtain a positive mark ($\geq$ 10). In the exam evaluation, we frequently have students who have failed the normal assessment and/or students that missed the normal assessment.

The available data, about each student and his/her involvement in the teaching-*learning* process, was: (i) provided by the institutional academic system (like students' personal information and unit information); (ii) provided by the unit teachers (include students presences in classes, developed activities and the corresponding student marks); (iii) provided by the *e-learning* system (information related with the *student-unit* interaction using the *e-learning* platform. The analysis of all the available data allowed the identification of the data subset considered in this application case. This data subset includes: (i) Student information: student number; student registration year; worker/full time student information; phase of admission to higher education and students' origin local. The phase information is only related with first year students and can have the values *first* or *second*. This attribute is also used to identify the students that are repeating the unit (value *rep*), as a consequence of a previous failure. The origin local was also classified in function of the distance to the institution, using for that purpose the descriptions: *near* (≤ 25 Km), *far* (≥ 25 and < 50 Km*), veryFar* (≥ 50 Km). In order to maintain the students' privacy, all the information that allows his/her identification is ignored or codified. (ii) Unit information: unit identification; unit name; curricular year and semester; associated course (iii) Class type information: class type identification; class description; class start hour; class duration; and, class week day (iv) Classes assiduity information: percentage assiduity rate associated with each student and each class type. The assiduity value was transformed in classes. In this particular application case, the used values were: *Low* (< 50%), *Acceptable* (≥ 50 and < 70%) and *High (*≥ 70% and ≤ 100%). (v) Assessment information: assessment activity identification; description; weight in the final mark, mandatory (yes/no); and, marks (obtained by each student). To represent some specific situations, negative values were used. In the project assessment results the value *-1* mean that the student misses the project individual discussion; the value *-3* mean that the student did not implemented the project work. In the unit final results, the value *-1* mean that the student was not submitted to any type of evaluation (test, project or exam); *-2* mean that the student *failed* the unit, but he/she was submitted to any one of the activities includes in the unit assessment (test/ project/exam). The final values marks were also classified in qualitative terms, using for that purpose the following attributes: *Satisfactory* (between 10 and 13), *Good* (between 14 and 16) and *VeryGood* (between 17 and 20) (vi) *E-learning* platform interaction information: number of distinct days that each student interacted with the unit using the *e-learning* platform. In qualitative terms, it was considered that from 0 to 16 corresponds to a *low* interaction, 17 to 32 corresponds to a *reasonable* interaction, 33 to 49 corresponds to an *expressive* interaction and values greater or equal to 50 correspond to a *high* interaction (the values distributions was analyzed in order to be possible the definition of this limits). Considering all the relevant data, it was designed the *data warehouse* (Figure 2) afterwards implemented and loaded with operational data. The loading process followed the *ETL* process steps, in which the relevant data was *extracted* from the source databases, *was cleaned* (when errors in data were detected) and *was transformed* in order to accomplish the format of the target system (the *data warehouse*).The *data warehouse* exploration has been done using *OLAP* and *data mining* techniques. *OLAP* techniques allow the analysis of data under different perspectives, taking part from the *data warehouse* multidimensional model.

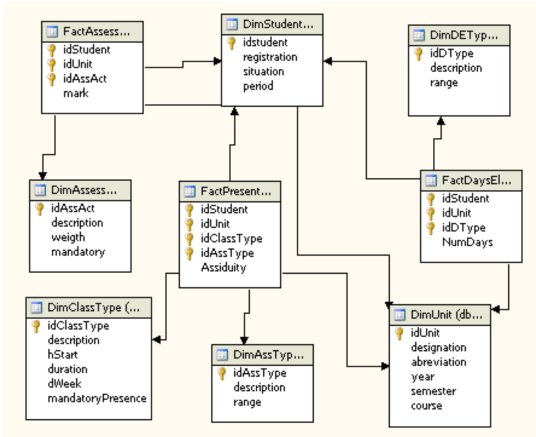

Fig 2 - The *Data Warehouse* model

Data mining techniques allow the identification of models that exhibit patterns and trends in data. In this application case *OLAP* cubes were created to analyse the students' results verifying both the *teaching-learning* experiences and the assessment methods influence. To analyze, particularly the unit results, their correlation with the theoretical classes presences, the unit interaction through the *e-learning* platform and the project work a cube was created. From its analysis it is clear that all the students that did not implement the project or misses the individual project discussion *fail* the unit. It can also be verified that there exists a large number of students with *low* assiduity to theoretical classes. Many of them are repeating the unit and others are differentiated by the phase of admission to the University. In these cases, only the students that are repeating the unit pass. These students also have a *reasonable* or an *expressive* interaction with the *e-learning* platform. The students that *fail* have few interactions with the unit, existing only one exception, a student (id 15) with *many* interactions. Figure 3 shows an extract of the analyzed data.

				Description		
				low		
Dim Student	Phase	Situation	Descriptior	Assiduity	Num Days	Mark
⊟ 36	⊟ rep	⊟ full-time	Final	0	19	12
			Project	0	19	8
⊟ 58	⊟ rep	⊟ full-time	Final	0	13	-1
			Project	0	13	-3
⊟ 62	⊟ rep	⊟ full-time	Final	0	18	10
			Project	0	18	9
⊟ 64	⊟ rep	⊟ full-time	Final	0	0	-1
			Project	0	0	-3
⊟ 68	⊟ 1phase	⊟ full-time	Final	0	1	-1
			Project	0	1	-3
⊟ 7	⊟ rep	⊟ worker	Final	8.33	21	-1
			Project	8.33	21	-3
⊟ 47	⊟ rep	⊟ full-time	Final	8.33	0	-2
			Project	8.33	0	-3
⊟ 48	⊟ rep	⊟ full-time	Final	8.33	37	15
			Project	8.33	37	14
⊟ 12	⊟ 2phase	⊟ full-time	Final	16.67	8	-1
			Project	16.67	8	-1
⊟ 15	⊟ rep	⊟ full-time	Final	16.67	52	-2
			Project	16.67	52	8
⊟ 29	⊟ rep	⊟ full-time	Final	16.67	11	-1
			Project	16.67	11	-1
⊟ 30	⊟ rep	⊟ full-time	Final	16.67	38	12
			Project	16.67	38	12
⊟ 33	⊟ rep	⊟ full-time	Final	16.67	23	11
			Project	16.67	23	8
⊟ 40	⊟ rep	⊟ full-time	Final	16.67	10	-2
			Project	16.67	10	-3
⊟ 13	⊟ rep	⊟ full-time	Final	25	16	-1

Fig 3 – Students data extract, grouped by *Low* assiduity rate

Another analysis allows us to verify that many students with *high* assiduity rates, and *expressive* or *many* interactions, *pass* the unit with *good* marks. Students with *acceptable* assiduities and *few* or *reasonable* interactions, *fail* or *pass* the unit with a *satisfactory* mark. Preoccupant situations occur with students that go to the University in the second phase, as many of them *fail* the unit. Figure 4 represents some of these situations as an extract of the data analysis.

				Description					
				acceptable			high		
Dim Student	Phase	Situation	Descriptior	Assiduity	Num Days	Mark	Assiduity	Num Days	Mark
⊟ 52	⊟ 1phase	⊟ full-time	Final	58.33	43	10			
			Project	58.33	43	8			
⊟ 8	⊟ 1phase	⊟ full-time	Final	66.67	24	-2			
			Project	66.67	24	8			
⊟ 18	⊟ 1phase	⊟ full-time	Final	66.67	68	12			
			Project	66.67	68	13			
⊟ 32	⊟ rep	⊟ full-time	Final	66.67	32	-1			
			Project	66.67	32	-1			
⊟ 41	⊟ 2phase	⊟ full-time	Final	66.67	68	-2			
			Project	66.67	68	8			
⊟ 55	⊟ 1phase	⊟ full-time	Final	66.67	6	-1			
			Project	66.67	6	-1			
⊟ 57	⊟ rep	⊟ full-time	Final	66.67	25	10			
			Project	66.67	25	8			
⊟ 11	⊟ 1phase	⊟ full-time	Final				75	35	-2
			Project				75	35	-1
⊟ 43	⊟ 1phase	⊟ full-time	Final				75	44	15
			Project				75	44	14
⊟ 53	⊟ 2phase	⊟ full-time	Final				75	19	-2
			Project				75	19	8
⊟ 56	⊟ 1phase	⊟ full-time	Final				75	45	15
			Project				75	45	16
⊟ 60	⊟ 1phase	⊟ full-time	Final				75	73	-2
			Project				75	73	8
⊟ 2	⊟ 1phase	⊟ full-time	Final				83.33	42	15
			Project				83.33	42	15
⊟ 20	⊟ 1phase	⊟ full-time	Final				83.33	43	16
			Project				83.33	43	18
⊟ 21	⊟ rep	⊟ full-time	Final				83.33	19	14
			Project				83.33	19	14
⊟ 31	⊟ 1phase	⊟ full-time	Final				83.33	27	15

Fig 4 – Students data extract, grouped by *Acceptable* and *High* assiduity rate.

It was verified that some students *fail* and others *pass* the unit with different marks. Now our objective is to identify the behavior of the students differentiating them by marks.The main purpose is to identify the students' profile. Through it, mainly in the case of *fail*, adequate actions can be implemented to minimize the failure rate in future unit editions. *Data mining* algorithms were used to identify the students' profile considering the assessment results: *Fail*, *Satisfactory*, *Good*, *VeryGood*. The objective is to find a model that describes the predictable attribute, *Mark*, as a function of the input attributes *phase*, *situation*, *numDays, origin* and the assiduity rates for the *theoretical (theo), practical and tutorial* classes. The traditional steps of the knowledge discovery in databases process were fallowed: *Data Selection, Data Treatment*, *Data Pre-Processing*, *Data Mining* and *Results Interpretation*. The first three steps were supported by the *data warehouse* implementation process. In the *Data Mining* step, it was selected a decision tree algorithm to carry out the classification task previously defined. The obtained model integrates a set of rules. From that it was selected the following set of rules (Table 1) that explicitly the patterns *Fail*, *Satisfactory* and *Good*. To refer that no rule was obtained to characterize the *VeryGood* pattern, due to the fact that only a student was achieved this mark.

If `Theo`= *'Low'* and `NumDays`= *'Few'*	*Fail*
If `Theo`= *'Acceptable and* `Origin` = *'Near'* and `Phase`= *'2'* or *'1'*	*Fail*
If `Theo` = *'Low' and* `NumDays` *not = 'Few'* and `Origin` = *'Near'* and `Phase` not= *'Rep'*	*Fail*
If `Theo` = *'Low'* and `NumDays` not = *'Few'* and `Origin` = *'VeryFar'*	*Fail*
If `Theo` = *'Low'* and `NumDays` not = *'Few'* and `Origin` = *'Near'* and `Phase` = *'Rep'*	*Satisfactory*
If `Theo` = *'Acceptable'* and `Origin` = *'Near'* and `Phase` = *'Rep'*	*Satisfactory*
If `Theo` = *'Acceptable'* and `Origin` = *'Far'* or *'VeryFar'*	*Satisfactory*
If `Theo` = *'High'* and `Origin` = *'Near'* and `Phase` =*'1'*	*Good*

Table 1 – Set of rules

The *OLAP* and *data mining* analyses allow us to verify that to decrease the failure profile it is necessary to take special attention to the students, differentiating them by the admission phase and also by their origin local, since several students are away from their familiar environment. The students that are at the first time in the unit must be encouraged to go in a regular basis to the different type of classes. These students are, in many cases, influenced by older students that say to them to avoid classes, mainly the theoretical classes.

For the second phase students, an additional support must be given as they arrive to the University when half semester has passed. Due to this situation, these students lose the initial curricular contents explanation and the subsequent curricular content comprehension, fact that could help to explain their failure. For these cases, the institution could adopt special activities or procedures, as extra classes or tutorial orientation, providing additional support to the students. It is also necessary, for all the students, to verify the evolution of the project work implementation, motivating the students and providing additional support when necessary. This support can be achieved using tutorial classes. This was not the case of the current edition of the unit, helping to explain the lack of presences and interest in this king of lectures. In what concerns repeating students, it is also necessary to motivate them to go to the presential classes, although in many situations these students have timetable incompatibilities. To overcome this limitation, the institution could adopt the schedule differentiation, like classes in the morning for the first year students and in the afternoon for the second year students. In complement, it is also necessary to motivate the use of *e-learning* platform increasing the interaction of the student with the unit. These different situations need to be evaluated without neglecting the students that present what we could say is a "success profile". With regard to the students that are away from home, the course director must take a special attention to these students, verifying if they have some integration difficulties and providing them additional support, if necessary.

The previous results allow us to conclude that data analysis supports the process of obtaining knowledge about the students and their academic behavior, through the students' behavior pattern identification. The obtained students' information allows the definition of a set of actions (included in the *SRM practice)* to closely follow the students with a view to decrease the failure rate. The identified actions, to integrate in future editions of the unit, include the presences/interaction automatic monitoring and an adequate support to the project work.

The presences/interaction automatic monitoring needs the automatic recording of this data and the subsequent automatic sent of alert messages to the student when deviating behaviors are detected. The main purpose of these actions is to alert the students to both the continuous unit interaction and the daily study as a way of increase their academic success. Between the teachers and the course coordinator/director could also exist periodically exchanges of information, which state how the semester is ongoing. Based on these reports specific actions can took place, as for instance verifying if the student has any problem.

The implementation of activities, like the described above is always supported and executed by a *web application*. It is now possible to confirm that the automatic *e-mailing* of alert message to the students that miss some classes has a positive effect on the students, once in general they attend to the next classes. Next steps, in this project include:

i) The integration of additional information in the data analysis process, related with the students and their activities in the unit scope

ii) The implementation of other activities that enable a closely monitoring of the students (activities included in the *SRM practice*);

iii) The SRM practice assess, through the analysis of its impact in the students behavior and their final results.

4 Conclusion and Future Work

In Portuguese Higher Education institutions persists a high rate of failure and abandon (mainly in the first graduation year). With the new formative process implementation aligned with the *Bologna Process*, the number of hours of contact between the teachers and the students decreased. This requires high student autonomy in the learning process. In this scenario, it is essential the design and implementation of mechanisms that facilitate the monitoring of the students' academic activities. In this context, we believe that the *SRM concept* and *practice* implementation, supported by the *SRM system*, creates an advantage towards the students success promotion, and, therefore, in the institution success, ensuring an effective *student-institution* relationship. The development of this project has occurred in different stages. In the first stage, it was proposed the *SRM concept* and *practice* and its validation. It was also verified the lack of an adequate technological support to this concepts (such as defined and understood on this work scope). The second stage was associated with the definition of the structural framework, which allowed the definition of *SRM system* architecture and its main functionalities. The third stage was associated with the prototype implementation and validation. Future work includes the fully implementation of the prototype and its validation with more application cases.

References:

[1] *Statistical results can be consulted in the official web page*: http://www.gpeari.mctes.pt.

[2] Payne, A., *Handbook of CRM. Achieving Excellence in Customer Management*. Elsevier-BH, 2006.

[3] Piedade, M.B. Santos, M.Y. *Student Relationship Management: Concept, Practice and Technological Support*. in *IEMC 2008 - IEEE International Engineering Management Conference* June 28 to 30, Estoril, Portugal, 2008

[4] Negash, S. Gray, P. *Business Intelligence*. in *Ninth Americas Conference on Information Systems*, 2003.

[5] Mundy, J., Thornthwaite, W. Kimball, R., *The Microsoft Data Warehouse Toolkit with SQL Server 2005 and the Microsoft Business Intelligence Toolset*. Wiley Publishing, Inc. 2006.