TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABREVIATIONS

ACK	Actual Knowledge of the attribute
AEF	Actually/Excellent Factor
AEW	Actually evaluation weight
AKD	Actionable Knowledge Discovery
CEQ	Course Experience Questionnaire
D	Design
D3M	Domain Driven Data Mining
DM	Data Mining
E	External Utility
EA	Evaluation per Area
EEW	Excellent evaluation weight
EFT	Effort
EM	Education Management
ESPOL	Escuela Superior Politecnica del Litoral
F	Function
FAe_r	Frequency of the attributes in the antecedent
FCe_r	Frequency of the attributes in the consequent
FIM	Frequent itemset mining
GQE	General Question Evaluation
HUIM	High utility itemset mining
Is	Interest of students
It	Interest of teacher
LDA	Latent Dirichlet Allocation
LP	Learning Promotion
LSA	Latent Semantic Analysis

LSI	Latent Semantic Indexing
OOA	Objective Oriented Utility based Association Mining
PA-AKD	Postanalysis-Based AKD
Pe	Expected agreement
PM	Production of Learning Materials
Po	Observed agreement
Q1 to Q23	Question 1 to Question 23
QE	Question Evaluation
Re	Recuperation effort
SEEQ	Student Evaluation of Educational Quality
SPTE	Student Perception of Teaching Effectiveness
T	Transaction Utility
U	Utility
VIF	Variance inflation factor
VSM	Vector Space Model
WARM	Weighted association rule mining
WUARM	Weighted Utility Association Rule Mining
Wxi	Weight of variable Xi
Qty	Quantity

INTRODUCTION

Universities around the world evaluate their teachers through the use of surveys that retrieve judgments and opinion from students. These surveys contain questions which assess different areas, such as communication, course design and material development that are enforced by the university administration as a mandatory practice. The teacher receives the results of these surveys with values indicating the level of achievement for each question along with statistical graphs to illustrate the highest and lowest evaluated questions. The objective of these evaluations is to provide feedback to teachers on those areas of the survey they may not be aware of.

Incidentally, the surprise comes when the teacher finds out he has not met his expectations about the evaluation grade. Due to lack of precise information, teacher may end up applying considerable effort to improve non-problematic aspects of his teaching style without addressing the real source of the teaching problem. This study aimed to help the teacher compares the rated questions identified by the students with teacher's expectations, conciliating students' opinion with teacher's opinion. Also, we intended to help the teacher to better understand those surveys and give him better clues to improve his performance

0.1 Context: questionnaires and instruments

In our research, we are using teachers and students' data from a Latin American University. The database contains 64,138 survey questionnaires answered anonymously. It holds information about 798 teachers who, as a whole, have given courses to 13,000 students. The university is composed of twelve schools (faculties or schools) and institutes. Each of them provides services to around 218 and 2,300 students per year. Figure 1.1 presents the distribution of the surveys between faculties in the University.

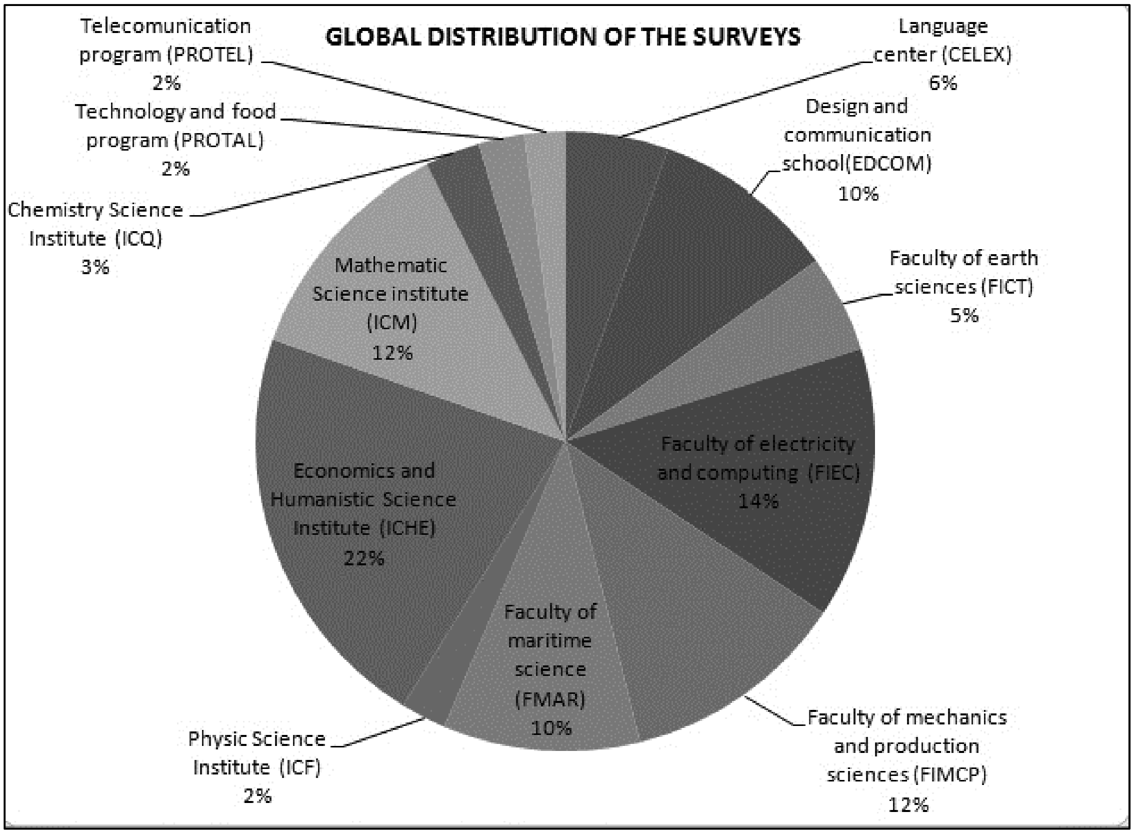

Figure 0.1 Global distribution of the surveys

Students usually take four to six courses each semester, resulting in eight to twelve courses per academic year. At the end of each course, a student has to fill out a survey about the teacher's performance. Each course has an average of 30 students who answer this survey. The data collected in these questionnaires include information about the year, the semester, the teacher's name and answers to the 24 questions that evaluate his teaching approach. The students' survey evaluates four areas: *Design*, *Learning Promotion*, *Production of teaching materials,* and *Management of education*. Each area has specific questions that help to evaluate the teachers. The survey has 24 questions where question 24 is an overall evaluation of the teacher.

The *Design* area of the survey evaluates the structure of the content and the evaluation of the course, and includes the following questions about the teacher:

Q1: Uses audiovisual help to support the content of the class
Q2: Fulfills the program proposed at the beginning of course
Q3; Evaluates student participation periodically in class
Q4: Evaluations fit the themes developed in class
Q5: Provides clear instructions for learning assessment (tests, quizzes, presentations, simulations, dramatic representation, role playing, etc.)
Q6: Motivates students to do additional research

The *Learning Promotion* area evaluates the materials constructed for the class and has the following questions:

Q7: Explains the course schedule at the beginning of the course
Q8: Explains class policies at the beginning of the course
Q9: Encourages active student participation in class
Q10: Summarizes key ideas discussed before moving to a new unit or topic
Q11: Establishes relationships between new concepts and those already known whenever possible
Q12: Motivates learning of the course material
Q13: Is willing to answer questions and offer advice within and outside of the classroom
Q14: Promotes reflection on topics covered
Q15: Maintains fluid communication with students
Q16: Is respectful towards students
Q17: Responds to questions in class about subjects related to the field
Q18: Delivers class content in an organized way
Q19: Develops class content in an understandable way

The two following areas are *Production of teaching materials,* and *Management of education* both concern management of the course activities. Therefore, we integrated these two areas into one, *Production of teaching materials,* & *Management of education.* The questions for this consolidated area are the following:

Q20: Prepares instructional, bibliographic or other resources to facilitate learning
Q21: Frequently uses schemes and graphics to support his/her explanations
Q22: Provides the results of the assessments on time
Q23: Attends classes on time

The survey uses a Likert scale from 0 to 10 and students assign a value to each question. The answers to this survey obtained from each course constitute what we called, the objective dataset of our research.

0.2 Aim of the thesis

When teacher receives his evaluation results, the data includes the application of descriptive statistics and the evaluation of each question belongs completely to the students' perception. The results of this instrument only offer the highest and lowest evaluated questions and the highest and lowest evaluated areas to the teacher, but no feedback is retrieved from the teachers' point of view about the evaluation. Consequently, the teacher is left on his own, with no additional support to reflect on the causes of the low rating in some questions, and the potential solutions to enhance his performance. Our methodology provides the teacher with a mechanism to incorporate his own views about the best ranked questions along with the views of other teachers.

As a solution to this problem, we proposed to take the results from these students' surveys compare them to the teacher's and his colleagues' insights in order to obtain a global view of the situation. This will help teachers review their strategies, understand relations among questions, and identify new areas to be evaluated that reflect students' interests regarding the themes that were not considered in the survey.

0.3 Research questions and overview of the methodology

The aim of the thesis is to present a combination of the objective analysis of students with the subjective considerations of teachers in the form of a methodology. This will expand the actions that can be taken from survey results, not only by knowing what specific questions are evaluated low or high by students, but also by finding what aspects are related to any specific question of the survey and how it is possible to improve the questions. In order to reach a higher evaluation in each question, different aspects can be related to survey questions such as: strategies, knowledge and teacher´s prior experience. We obtained further information about teacher´s perceptions by applying one interview and two surveys (S1 and S2) to teacher and using the students' perceptions from the university survey. The objectives of analyzing teachers 'and students 'perceptions were: first, finding relations between survey's questions that created new patterns that will become key patterns. Second, measuring the relation among strengths and weaknesses of the teacher and finally, finding new topics that were currently not measured by the survey but were actually introduced by teachers during the course.

We focused on three research questions throughout this methodology:
Research question 1: Are there differences or similarities between the students' perception and the teacher's perception?

Research question 2: Can we measure the relationships between survey's questions based on students' perceptions, teacher colleagues' opinions and individual teacher's beliefs?

Research question 3: Are there new topics to consider aside from the already measured topics?

In this research, we are using the data from the Electrical Engineering and Computer Science Faculty. We limited our analysis to this faculty for some important reasons. It has a highest number of students with an international certification from the Accreditation Board for Engineering and Technology, ABET, (ESPOL 2011) that guarantees their graduates are prepared to enter a global workforce; the teachers of this faculty showed interest to participate in this research. We applied this research in the following courses: Programming Fundamentals, Research Methods Applied to Computing, Software Engineering I, Entrepreneurship, Web Application Development and Digital Communications.

In this research, we applied a domain driven data mining methodology, D3M (Cao, Zhang et al. 2010) that uses objective and subjective data sets. The D3M is based on objective and subjective interestingness measures. The propose methodology was developed in three papers, and is explained in detail in the following paragraphs using our three contributions:

Our <u>first contribution</u> is presented in the <u>first paper</u>; here we proposed a method to build a representation based on actionable knowledge; this method generates relationships among questions representing teacher's perception. We evaluated these relationships using objective and subjective measures, and presented two groups of patterns: the teacher's old model and the teacher's new model. The first one, the teacher's old model, is a group of patterns representing rules of thumbs teachers apply in class. The second one, the teacher's new model, is a group of patterns representing students' perceptions. We integrated both in one representation. The patterns objectively and subjectively evaluated are thus grouped together. Additionally, we included the knowledge coming from teacher strategies. The resulting method can be used to present patterns, knowledge and meta-knowledge together in one unified model. In order to construct this model, we interviewed the teacher about the rules obtained from the students' while obtaining knowledge and meta-knowledge associated to the classes and the activities he accomplished. We focused basically on what teachers did to improve their performance in class, like anecdotes that reflect what teachers have done over the years. This type of information helped the teacher understand the rules automatically extracted from the students' survey. This knowledge in the sense of Herrmann and Kienle et al., (Herrmann, Kienle et al.), along with the information obtained from the students, was then presented graphically, offering a model with attributes or questions closely related to each other, and their connections with the meta-knowledge for each of the attributes in the association rule.

The <u>second contribution</u> is presented in the <u>second paper</u>, where we assessed the utility of these relationships using association rules based on objective and subjective measures, constructing a utility semantic measure to evaluate association rules and testing if they were good rules that improved the model. This semantic measure identified which were the most interesting rules to work with, focusing on the effort required to improve the topic of each question in the survey. In order to achieve this, we used linear regression analysis and created open, context and evaluation rules. Also, we constructed objective and predictive attributes to build the new dataset. This new dataset focused on teacher's interests. Finally, we evaluated association rules with this utility measure to obtain rules that have more value from the point of view of the teacher's applicability.

So far, we have presented association rules evaluated from three points of view objective, subjective, and semantic. In order to complete the model from the first paper, we needed to classify the knowledge extracted from each teacher's interview according to the questions from the students' surveys. For this purpose, our last contribution, presented in the third paper, uses a topic modeling analysis to obtain the general topics that teacher mentioned during the interviews. Also, we constructed a dictionary of the twenty-three questions of the survey and proceed to classify each teacher's interview themes. This dictionary presented all the variables of a teacher performance evaluation instrument. We constructed the dictionary by integrating abstracts from scientific pedagogical articles related to each survey's question and usual dictionary definitions of relevant concepts. Then, using Latent Semantic Indexing (Kuralenok and Nekrest'yanov 2000), we evaluated the effectiveness of the classification with the cosine similarity measure. Finally, we compared the machine classification with two human classifiers and presented the precision of the results.

As a result of the three research papers, we are now able to provide universities with a mechanism to improve students' evaluation surveys by combining them with the insights and experiences of teachers and obtaining a valuable tool to improve teaching

CHAPTER 1

LITERATURE REVIEW

Universities commonly use surveys to analyze teacher's performance in the different courses that are part of the careers curriculum. Around the world, different types of surveys are applied. In this chapter, we provide an overview of different survey instruments used worldwide. Additionally, we present the literature review for the components of our methodology for the analysis of teacher and students perceptions using the D3M Data Mining methodology, the Actionable Knowledge Discovery (AKD) framework and the use of objective, subjective and semantic interestingness measures.

1.1 Overview of students surveys around the world

The teacher performance evaluation is common practice at universities every semester. This process allows university management to obtain information from the students about their teachers. Universities analyse this information to gain knowledge through the analysis of the surveys and to take actions to reward teachers in order to improve their teaching abilities among other benefits.

Many instruments or tools have been proposed to evaluate teachers' performance in classroom. The major difference among these tools is the dimensions analysed. The point in common is that all of them work only with students' opinions. For example, the Course Experience Questionnaire or CEQ applied in Australian universities, evaluates the experience a student had during a course (Hirschberg, Lye et al. 2015). Students fill out the surveys evaluating different teacher's attributes, such as teaching attitude, teaching content, teaching evaluation, and other teaching aspects(Jiabin, Juanli et al. 2010).

Another example is the case of "Improving Learning of Higher Education, IDEA," a non-profit organization that, since 1975, provides the instrument and all the processes to evaluate and improve teacher's performance and all related services, including the application of the instrument and the analysis of the answers from the students' point of view. IDEA specializes in using student questioning to provide opportunities to improve teaching and learning processes (IDEA 2015).

Likewise, The Student Evaluation of Educational Quality, SEEQ, is an instrument from the Center for the Advancement of Teaching and Learning at the University of Manitoba, Canada, where students evaluate teachers through teaching dimensions that include: learning, enthusiasm, organization, group interaction and their overall impression of the instructor (University of Manitoba 2015).

Finally, the Student Perception of Teaching Effectiveness, SPTE, is an instrument from the Wichita State University, U.S., that measures students' perception of teaching. It is used for summative purposes to congratulate teachers who are doing well, and for formative purposes, to improve the teaching. (Wichita State University 2015). (Jackson, Teal et al. 1999)

In the previously described teacher's evaluation instruments, only students' point of view has been analysed. Teacher's opinions about the perceptions of students' feedback on their work has rarely been incorporated (Arthur 2009). The author (Arthur 2009) claimed that a teacher needs to consider students judgments to improve his development specially if a teacher gets low feedback. However, a teacher should consider changing his development based on his own feelings and professional judgment about what he is teaching. Hence, we observed an existing gap where comparison between students' evaluation and a teacher's perception needs to be made as referred by experts in the field (Hirschberg, Lye et al. 2015).

Besides the evaluation instruments, other data sources have been analysing to evaluate teachers' performance. These datasets refer to teacher's personal information, characteristics of the course, besides others. In the next section we include the analysis of these data sources.

1.2 Analysis of evaluation instruments

Most Universities perform their analysis of the data, taking into account four different sources: teacher's personal information, information not related to teacher's survey performance, the data related exclusively to the students' survey answers, and a mix of all the above. Four different investigations presented their findings referring to the above mentioned sources. In the first one(Wen, Rong et al. 2014), the information is related to teacher's age, experience, professional occupation and performance; it takes into account the teacher's personal information. Wen et al., selected some of these elements as the best characteristics for an ideal teacher. In the second one (Zhang and Wang 2012), the non-teacher factors or information not related to teachers' survey performance can affect the students' evaluations. These factors include the number of students in a classroom, the teaching hours in a term, the type of students, the grade and the excellence of the course. This information could affect the final result of an objective survey and is not related to teachers' abilities, but to students' abilities. In the third one (Jiabin, Juanli et al. 2010), the focus of the survey data was put into the analysis of the questions that students answer to evaluate the teacher's development; open and close questions were analysed, and different techniques were used to analyse them such as association rule mining and decision trees or linear regression (Badur and MARDIKYAN 2011). Finally, a mix of dimensions coming from teachers and students are considered to evaluate teacher's performance as teacher employment status, course workload (Badur and MARDIKYAN 2011), students attendance, percentage of survey students' fill up and the students' survey results.

All these research focus only in the analysis of the objective data and what it says about teachers using students' judgments, course difficulty and teacher's personal information. The analysis of teachers' opinion is very limited, as well as the actions to improve teachers' development. There is a lack of comparison among teaching approaches, the student understanding of the approaches, and clear description about what is measured in survey questions. Additionally, there is no clarity or certainty of which of the survey's questions should be the starting point to improve teacher's performance.

In our first paper, presented in Chapter 4, we identified the teacher's strengths provided by students' survey and found the strongest relationships among questions and teacher's using association rules. We illustrated this using a case study where six teachers applied our methodology. In this paper we incorporated teacher`s opinion and knowledge. In order to use knowledge, we needed a specific framework. This framework included besides the objective analysis of the survey students' data, subjective analysis including teacher`s opinion and additionally measures that help in the evaluation of objective and subjective patterns; this framework was D3M, Domain Driven Data Mining and AKD(Cao, Zhang et al. 2010). We explain these frameworks in next section.

1.3 Domain Driven Data Mining (D3M) and Actionable Knowledge Discovery (AKD)

D3M is a methodology that is not only driven by the data but also by the domain (Cao, Zhang et al. 2010). This methodology looks for finding patterns that triangulate data, the domain knowledge and experts. It is based on the knowledge, the human interaction, the intrinsic knowledge and business expectations.

Moreover, AKD is part of the Domain Driven Data mining (D3M) frameworks. The main objective of AKD frameworks is to discover actionable patterns with immediate application focusing on domain knowledge, e.g. increase in profits and better efficiency, (Longbing 2008). Four different AKD frameworks have been proposed (Cao, Zhao et al. 2010). Each of them is specialized in one specific task: the first one is recommended when the amount of data sources to be accessed is big, MSCM-AKD(Cao, Zhang et al. 2010); The second one, is recommended when it is necessary to access the data several times in order to refine it and generate patterns, CM-AKD; the third one is applied when it is necessary to evaluate the patterns using only one interestingness measure (technical, business measures) UI-AKD; finally, when it is necessary to have separation between technical and business interestingness for the analysis, PA-AKD is recommended. For this research, we chose the latter, the Post analysis based AKD, PA-AKD framework, because it handles technical and business aspects of the problem in a separate way and it doesn't work with multiple data sources.

Other components of AKD are technical and business interestingness measures that help identify the interest of a pattern from the objective and subjective point of view (see Figure 1.1 and Figure 1.2). In other words, each pattern is analysed both objectively, using data, and subjectively, taking into considerations the opinions of the expert who worked with the data.

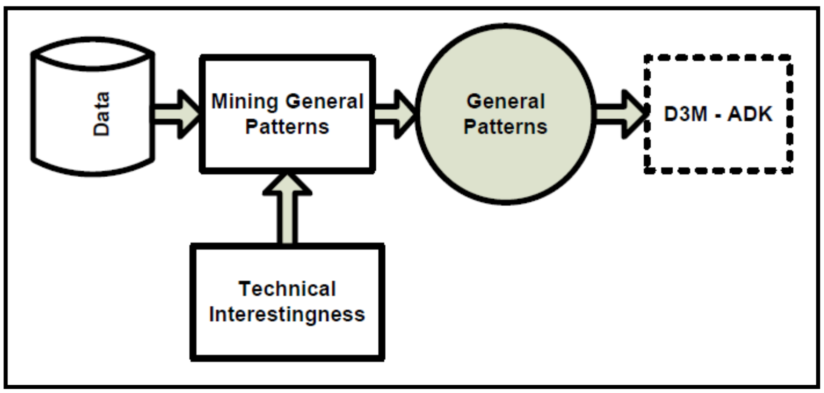

Figure 1.1 Data mining steps

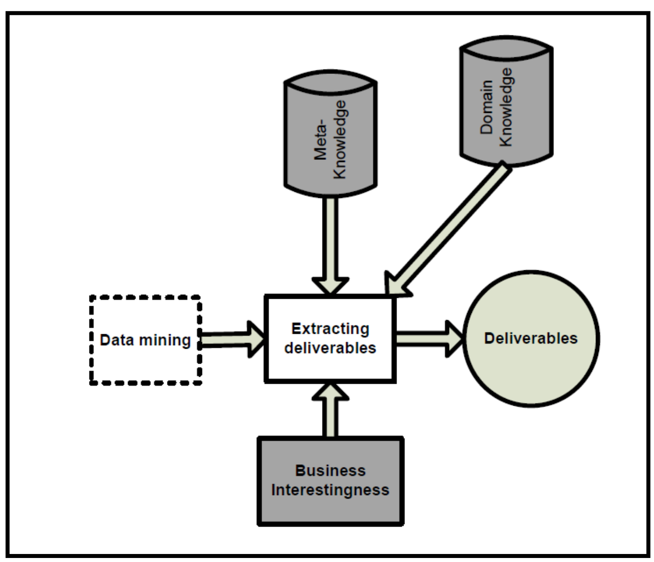

Figure 1.2 Post Analysis - AKD's approach
Taken from (Cao, Zhao et al. 2010)

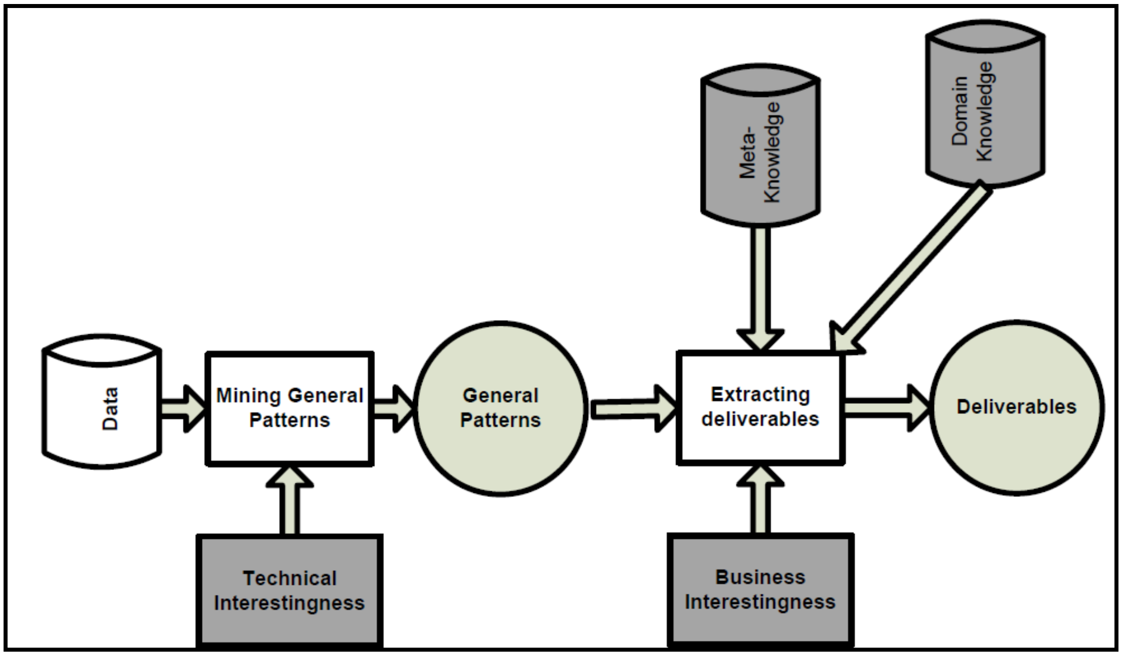

Figure 1.3 DM and AKD approach
Taken from (Cao, Zhao et al. 2010)

In Figure 1.3 we see that this framework works with three types of knowledge: domain knowledge, expert knowledge and meta-knowledge and a general concept of interestingness measures. We present the three knowledge types in section 1.4 and the interestingness measures in section 1.5.

1.4 Knowledge, domain knowledge, experts' knowledge and Meta - knowledge

The research within the AKD framework presents different components such as meta-knowledge, domain knowledge and experts' knowledge. We are going to define these concepts that are part of AKD.

First, Knowledge is a mix of constructed experiences, values, information or contextual data and expert insight that belongs to any enterprise in an implicit way and that transforms itself in assets for the organization over time (Davenport T. H 1999). These assets belong on the one hand, to people who manage the knowledge and learn in everyday experiences, and on the other hand, to the organization, which saves it in different forms including documents, processes and practices. People and organizations struggle to recognize and organize knowledge because of the large number and variety of business situations and solutions. An example of knowledge gathered during a course is everything related to students, tests, exercises, workgroups and teacher.

Second, Domain Knowledge (DK) is related to the experiences, values, and user insights implicit in user knowledge (Davenport T. H 1999). A simple definition about domain knowledge was given as "the knowledge of the subject area (domain), what you know about a subject or topic" (Paquette et al. 2011). DK is therefore related to the user experiences in a specific domain. For example, domain knowledge for a teacher might include academic problems that students face during the discussion of a specific topic (including insecurities and fears), feedback about the difficulties with a topic and specific abilities that the teacher has to apply to reduce students' fears in order to reach academic objectives.

Third, in any industry, domain knowledge experts are individuals who have reached high levels of expertise in a particular domain. They specialize in a specific problem resolution. They gain expertise doing similar tasks or resolving the same problem in different contexts, and store and apply these rules of thumb depending on the situation they need to solve. For example, teacher in programming course find constantly that students confuse the assign command (=) with the equal command found in math courses; to correct this he needs to teach students about the assign command in programming course.

To elicit domain knowledge from an expert, it is necessary to use techniques to retrieve the tacit knowledge and transform it into explicit knowledge in the form of rules (Flavell 1979, Yi-Dong, Zhong et al. 2002). This knowledge can then be standardized and applied in future similar situations.

Finally, Meta-knowledge is knowledge about knowledge, what we know about our cognitive abilities or how we learn. We focused on a teacher's knowledge, cognitive tools, abilities, limitations and the use of strategies (Valot and Amalberti 1992). It identifies persons, variables or values that intervene in the activity as well as strategies, steps and actions needed to accomplish the activity. If we consider the academic context, MK include how students learn, how a specific topic is taught, what is the easiest and what is hardest topic for students to learn. To identify meta-knowledge, three questions have to be answered (Flavell 1979, Yang, Yu et al. 2009): *why, what and how*. The *why* identifies the reasons associated with performing a specific activity; the *what* defines the objective to be reached in performing an activity; finally, the *how* refers to the strategy or steps required to accomplish the activity. The following example shows how to apply these questions in relation to the academic evaluation. If the activity is "*to make a student evaluate a teacher by filing out a survey*" to obtain the relevant meta-knowledge, we will ask the following questions: *why* is this evaluation done?; *what* concept do people interested in the results of this questionnaire learn?; *how* are learned concepts going to be used?. In this process of identifying meta-knowledge, teachers have to talk about why they make students do specific activities, what activities and concepts they do in class and how the teacher tries to have those activities done in the best way. These activities are related to each of the survey questions. In this research, we called the survey questions *attributes* to standardize the terminology.

1.5 Interestingness: Objective, Subjective and Semantic measures and Actionability

The interestingness of a pattern reveals how interesting it is for the user. According to Geng and Hamilton (Geng and Hamilton 2006), Interestingness is determined through nine criteria: conciseness, coverage, reliability, peculiarity, diversity, novelty, surprisingness, utility and actionability. These criteria are classified in three categories: objective, subjective and semantic. Namely, conciseness, coverage, reliability, peculiarity and diversity are considered objective criteria and are measured using objective measurements. Some of these objective measurements are support, confidence and lift. Whereas, novelty (also known as usefulness) and surprisingness (also known as unexpectedness), are considered subjective criteria and are measured using the user's domain and his background knowledge. Conversely, utility and actionability are considered semantic measures. Semantic measure takes into account the semantic of a pattern and the explanations that the user can give about this pattern. It is possible that the interestingness is represented by some of the above mentioned criteria, for example, something could be interesting if it is unexpected and actionable (Silberschatz and Tuzhilin 1995, Silberschatz and Tuzhilin 1996). Objective, subjective and semantic measures try to measure the Interestingness of a pattern but from different points of view. Additionally they are used to sort patterns, using interesting objective measures first and then subjective measures (Lenca, Vaillant et al. 2007).

In AKD, patterns are measured using two types of interestingness measures: objective and subjective measures.

We defined objective and subjective interestingness measures in detail as follow: Objective measures concern technical interestingness (Bing, Wynne et al. 2000)–rule structure, predictive performance, and statistical significance–of the data. Researchers have proposed many objective measures, and have studied their characteristics and properties (Tan, Kumar et al. 2002), (Geng and Hamilton 2007); other researchers have studied the suitability of a measure with respect to a certain domain (Xuan-Hiep, Guillet et al. 2006). Among these metrics, we applied in our research three of these measures: *support*, *confidence* and *lift*.
Subjective measures are related to the needs and interests of the user and the domain (Bing, Wynne et al. 2000, Oliviera Rezende Solange 2009). Many authors have proposed subjective measures (Bing, Wynne et al. 2000, Geng and Hamilton 2006), (Silberschatz and Tuzhilin 1996) (Geng and Hamilton 2006); Our research is based on the subjective and semantic criteria of Gen and Hamilton, and on Oliviera et al.'s knowledge criteria (Oliviera Rezende Solange 2009).

Gen and Hamilton refers a pattern as *Usefulness* if it is not outstanding but could help the user in decision making and is also called as accepted belief; a pattern is considered *Unexpected* if it is completely new to the user also mentioned as a contradicted belief; Utility is the characteristic of a pattern that helps to reach a specific objective; finally, actionability measures the ability of a pattern to suggest taking some concrete action for user advantage.

Oliviera et al. (Oliviera Rezende Solange 2009) proposed five knowledge categories, *unexpected*, *useful* , *obvious*, *previous* and *irrelevant.* Oliviera et al. stated that a pattern should fall in one of this categories; the knowledge category is selected by the user based on his knowledge about the pattern; a pattern has a range of evaluation from very well known (we can call it irrelevant) to a pattern completely new (we can call it unexpected).

We mix Gen and Hamilton criteria's with Oliviera et al. criteria to have a wide range of knowledge categories for the user in one hand; in the other hand we use some of these criteria for the construction of our utility measure. This brings us to the second paper, where we constructed the utility measure. We explain the Utility measure in the next section.

1.6 Utility mining and utility formula

The general utility mining approach (Wang, Liu et al. 2007) states that utility mining should contained two types of utility: transactional and external utility. Transactional utility refers to the one obtained from the working database and external utility refers to additional datasets around the transactional information environment. For example, we can obtain transactional utility from students' dataset and external utility from teacher's surveys (as we recall teacher's surveys are S1 and S2). The detail of the instruments S1 and S2 are in Chapter 3.

We constructed a teacher's survey (S1) that contained five components. We give a quick glance of the components. The first component was the selection of attributes from students' survey. From the survey's that students filled up, the teacher selected the ones he considered the most relevant from teacher's point of view. Then, he constructed association rules based on his selection. The constructed rules are the open association rules. The second component was the selection of attributes per areas from the students' survey. After, the teacher chose his preferred questions per area and constructed association rules using these preferred questions as attributes. The constructed rules are the context association rules. In other words, we obtained two different evaluations for each attribute, the one obtained from the complete survey and the other one obtained from selection per area. The third component was the section of the reaction to low evaluation where the teacher evaluates the actionability of the low rated questions. The fourth component was the evaluation using the knowledge categories from Oliviera. The fifth and the sixth component has already been considered in the first and second components as the construction of the open and context association rules. These components helped evaluate the perception of the teacher about the students' survey answers.

We described how we integrated the survey components in the construction of the utility measure in the methodology in Chapter 5.

For the construction of the utility formula for association rules, if the attribute appeared in the association rule, then the attribute had a utility value. For example, if communication attribute is important for the teacher, or if the development of the audiovisual material is important, then we defined the attribute frequency when it appeared in the antecedent or in the consequent of the rule. If the attribute was more frequent in the antecedent, then that attribute was considered as predictive attribute. We placed the predictive attributes in the antecedent side of the rule. If the attribute frequency was frequent in the consequent side of the association rule, then it was considered objective attribute. For the objective attributes, we used the objective oriented utility based association mining (OOA) (Yi-Dong, Zhong et al. 2002), that focuses on the semantic sense of the attributes considered by the user. We placed the objective attributes in the consequent side of the rule.

Additionally, because it was difficult to quantify the attributes that needed improvement, we included the effort as part of the utility measure to improve the low evaluated attributes. The utility measure helped to quantify subjective attributes which were not quantifiable such as: communication, organization, and respect. We will see the development of this measure in Chapter 5.

Once we measured patterns utility based on the objective and subjective datasets, we were interested in analysed the interviews from teachers to identify the wide variety of topics teacher talks in class. This bring us to our third paper, therefore, we applied topic modeling to the interview data of all teachers and obtained categories in teachers' interviews. We explained this in section 1.7.

1.7 Probabilistic Topic modeling and measuring the model

Teacher answered the interviews and gave their perspective during it. For the analysis of this information we applied a technique called probabilistic topic modeling to understand the relation between documents representing the interviews. Applying topic modeling we obtained the most representative topics and classified the interviews sentences into them. We asked two experts to help us classifying by hand the same sentences from the interviews to test the effectiveness of the constructing model. We applied Cohen Kappa measure to do this.

1.7.1 Probabilistic Topic Modeling.

We applied probabilistic topic modeling to the interviews from teachers. Probabilistic topic modeling are algorithms that work with statistical methods to identify the most relevant theme inside a group of documents (Blei 2012). They are also called mixed membership models in automatic content analysis methods (Grimmer and Stewart 2013). The theme or topic found by the topic modeling algorithm is a distribution of words over a fixed vocabulary (Blei 2012) where the topic is represented with the group of distributed words from the vocabulary along the documents. One of the statistical methods that topic modeling works with is LDA or Latent Dirichlet Allocation; the basic concept behind LDA is the use of joint and conditional distributions. This method is very useful when there is a group of documents with different themes or there is no categorization (Grimmer and Stewart 2013).

Using probabilistic topic modeling, the technique classified the topics among the interviews. In order to evaluate the machine classification, we asked two experts to classify too the interviews ideas as the machine classification did. We used Cohen Kappa measure to evaluate how good the classification was made by our knowledge experts against the machine classification. We explained Cohen Kappa measure in next section.

1.7.2 Cohen Kappa

Cohen's Kappa statistic is a measure for computing the inter-rater reliability coefficient, or the reliability between the opinions of two human categorizers. It is assumed that the categories should be disjoint (do not overlap) (Gwet 2015). It is used to compare the classification of a subject made by different human evaluators into different categories. It is constructed from the observed and expected frequencies on the diagonal of a square contingency table.

In Chapter 6, we explained deeply the use of LDA to obtain the categorization of themes and the use of Cohen Kappa measure to evaluate the relation between human categorizers' opinions.

1.8 Conclusion of the review

After the analysis of the literature reviewed, we decided to use the framework from D3M methodology along with techniques from data mining. In our <u>first contribution</u>, we use objective and subjective information; part of the analysis includes experts' knowledge that helps better understand the patterns obtained from the data mining process. For this reason, we needed a methodology that includes the use of knowledge. In this case, Crisp-DM or SEMMA methodologies were not appropriate since they only work with objective datasets; knowledge or meta-knowledge information could not be incorporated to the analysis. We chose the Actionable Knowledge Discovery framework, AKD, specifically PA-AKD, Post analysis based AKD, since it can handle objective and subjective aspects of the problem in a separate way, working with only one data source. Post Analysis Based AKD works with the knowledge and the meta-knowledge of the problem and proposes a mean to incorporate the experts in the loop; it also uses objective and subjective measures to evaluate the quality of the patterns retrieved. To filter out our association rules, we use three objective measures, support, confidence and lift because they allow to analyze the attributes that are in the consequent and in the antecedent of the association rule; for the subjective measure we used unexpected, usefulness, obvious, previous, irrelevant, and actionable as actionable knowledge categories.

For our second contribution, we constructed our utility formula using two types of utility, transactional and external utility. Additionally, we used the frequency of the attributes presented in the association rules and focused in the fulfillment of the partial utility value from each attribute. We retained some concepts of Objective Oriented Utility Based Association Mining and identified objective and predictive attributes to generate association rules to be evaluated with the utility measure. Another consideration for the utility measure is the application of the analogy of a machine lever example to obtain the Effort teacher needs to improve an attribute. We based our utility formula on these elements.

Finally, for our last contribution, we used probabilistic topic modeling algorithms to obtain some clusters of words that represent topics; we applied LDA with Gibbs Sampling. LDA looks for the posterior distribution between the words in the interviews and the hidden topics. For the comparison between the interviews and the topics, we used Vector Space Model because it is the baseline model applied to compare text documents without using complexes approaches such as Natural Language Processing. We used Latent Semantic Indexing (LSI) to measure the Cosine similarity between a topic vector and an interview-paragraph vector. Cosine Similarity measure is a standard similarity measure in Vector Space Model problems. We worked with two human experts to test the classification of the model. We applied Cohen Kappa measure to evaluate the similarity among people and the model classification as well as between experts' classifications. Cohen Kappa is one of the most commonly used to measure inter-expert agreement.

CHAPTER 2

GENERAL METHODOLOGY

This chapter presents the general methodology and the objectives we defined during the research. In the research, we used the students' surveys and teachers' surveys and interviews. We used machine learning tools to discover patterns, create a utility measure and apply LDA to understand better teachers' interviews.

2.1 Methodology overview

The objective of this research was to create a methodology to integrate the objective analysis of students with the subjective considerations of teachers. With this objective we applied D3M that uses objective and subjective data sets. The D3M is based on objective, subjective and semantic interestingness measures. In order to achieve the aim of this thesis, we accomplished the general objective focusing the following three specific objectives:

- To build a representation using students' and teacher's perceptions based on actionable knowledge
- To construct a utility semantic measure to evaluate the usefulness of association rules within the model
- To discover new topics from the analysis of interviews to improve the teacher evaluation

We present in Figure 2.1 the conceptualization of our research objectives; the figure shows the data sources we used: teacher survey, students' survey and the teachers' interview. The figure includes the three specific objectives: the first objective is the building representation, the second one is the semantic measure and finally the third one is the new topic discovery; each objective is represented using circles with a number

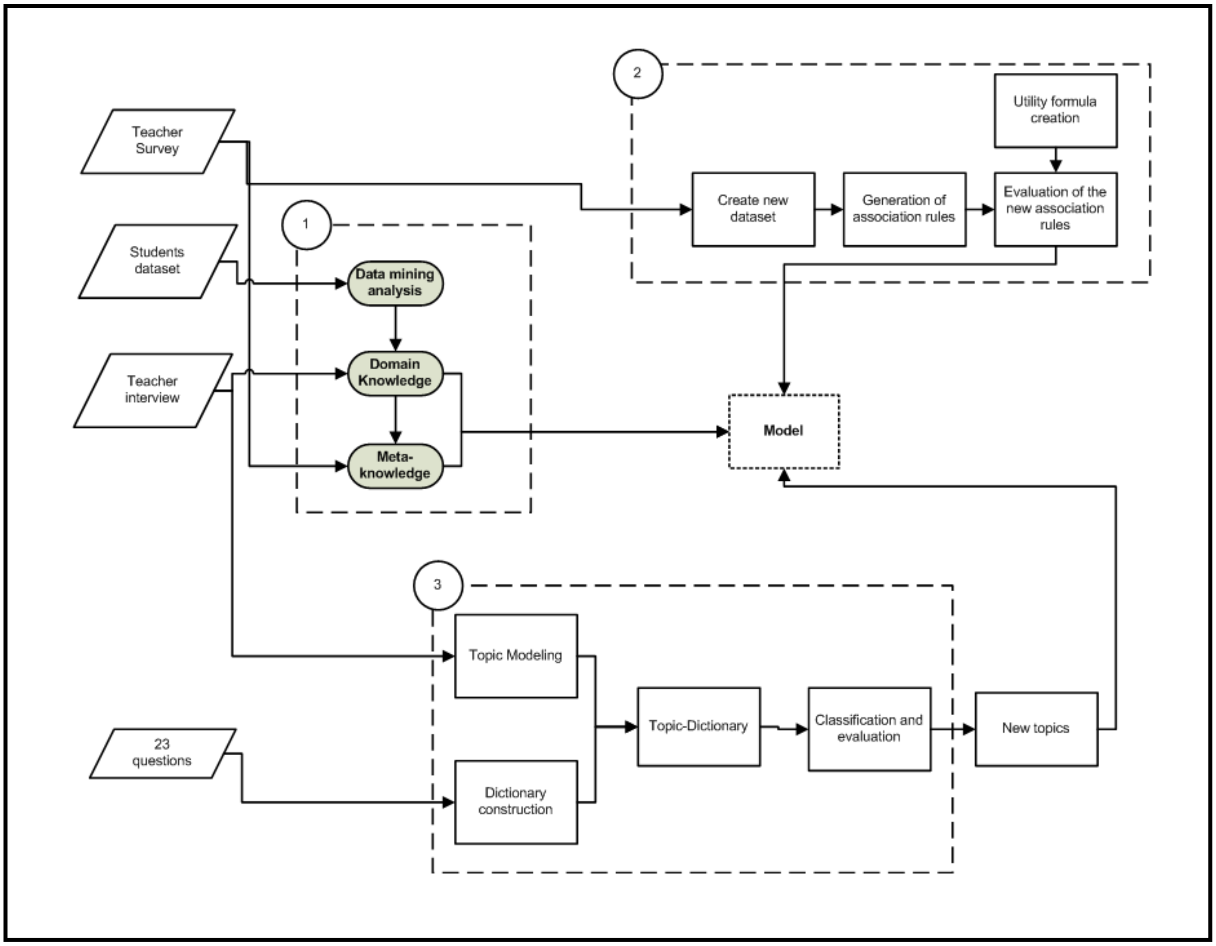

Figure 2.1 Methodology and Research objectives

To obtain the model resulting from Objective 1, the domain driven data mining methodology integrates the data mining analysis, the domain knowledge and the meta-knowledge. To achieve the first objective, we used the data from two datasets, students' surveys and the teacher's survey. For the second objective, we used teachers' surveys and generated the new dataset to feed Apriori. Then, we obtained association rules that contain predictive and objective questions that belong to a specific professor. Finally, to accomplish the last objective, we used teacher's interviews and the information from the 23 questions. In this last objective, we applied the topic modeling to construct the dictionary from the 23 questions, evaluated the results from the model and from the human evaluators and obtained the new topics. We will explain in detail each objective in the next section.

2.2 Methodology description

We are going to describe the three specific objectives and how we filled the existing gaps by achieving each of them. Later on, we will mention the relation between each objective and the literature review presented in Chapter two.

2.2.1 Objective 1: To build a representation using students' model and teacher's model based on actionable knowledge

There is substantial research (Marsh 1984, Zhang and Wang 2012, Eshach, Dor-Ziderman et al. 2014, Wen, Rong et al. 2014) dealing with teacher evaluation, student evaluation, and analysis of perception, to improve teaching and learning. Among the DM tools that have been used to evaluate the teaching and learning, there are decision trees (Jiabin, Juanli et al. 2010), association rules (Pan, Qu et al. 2009, Lanfang, Qingxian et al. 2010) and clusters (Burton, Morris et al. 2014). Until now, there is no explicit representation to capture teacher and students' perception. A representation of perception can help a teacher understand what he did in class and how the students received it. To work with D3M three important components are needed: the data mining analysis, the domain knowledge, and the meta-knowledge extraction. Our representation presented these three components in the form of rules integrated among them. These rules reflect the background knowledge teacher generally uses to teach. This knowledge is important because it relates questions from the survey evaluation to actions taken by teacher and vice versa.

2.2.2 Objective 2: To construct a utility semantic measure to evaluate the usefulness of association rules within the model

Some authors (Silberschatz and Tuzhilin 1995) (Padmanabhan 1998, Padmanabhan and Tuzhilin 1999, Bing, Wynne et al. 2000) consider that in addition to the analysis of objective patterns, it is also important to perform a subjective analysis. Subjective analysis provides tacit knowledge which is not formerly captured, but it is used all the time during the decision making process. The subjective analysis always relies on subjective measures where the interestingness of a pattern depends on the decision maker and not only on the statistical analysis of a pattern (Padmanabhan and Tuzhilin 1999). The objective analysis is based on objective measures that support the obtained objective patterns. The limitation of this objective analysis is that most of the patterns obtained are not interesting since most are already known. There is a wealth of measures that evaluate objectively patterns, or subjectively patterns but there is no measure that integrates both objective and subjective considerations using the evaluation of objective and subjective data. Therefore, we considered important to create a measure that adds the subjectivity of the patterns and objectivity of the patterns. This measure is called utility semantic measure and includes the statistics of teacher's survey, the students' evaluation and the subjectivity of the domain knowledge retrieved. We used the utility mining approach to construct the objective-subjective measure.

The utility semantic measure is applied to each association rule to evaluate if this rule has a high utility. With a high utility, the rule is considered a key pattern. On the other hand, if the utility semantic measure evaluates the rule as low, the measure provides the effort component. This component helps the teacher identify how much work he needs to do in order to improve the low graded attributes.

2.2.3 Objective 3: To discover new topics from the interviews analysis to improve teacher evaluation

Clark & Yang et al. (Clark 2000, Majid, Yang et al. 2014) considered that most students, besides learning from class materials, acquire other abilities such as: organizing and expressing ideas, participating in discussions, and doing additional research. Thus, course objectives are related not only to the content per se but also to additional abilities of the students to react positively, learn happily, and share their opinions in open discussions.

When the teacher promotes the demonstration of the above-mentioned abilities, other topics that he considers important for student's professional preparation are discussed. These topics, that are not part of the evaluation in the students' survey, expose a gap between what is taught and what is evaluated. The teacher in his interview described all activities he did in class to help students reach the desired course objective. He also shared his perceptions of the different students' reactions to topics that are not related to the course content. Very little research has been done to analyze teachers perceptions (Arthur 2009), from students reactions in the application of different teaching strategies.

We interviewed the teacher about his students' perceptions during the course and how students responded to the activities he planned. In the interview, the teacher revealed aspects related to what students do, feel, think, and how the teacher perceived students' feedback. We applied topic modeling to the group of interviews to extract all the topics in the teachers' interviews.

2.3 Objective description and associated results

2.3.1 Objective 1: To build a representation using students' model and teacher's model based on actionable knowledge

The representation is constructed using D3M methodology and PA-AKD framework. The PA-AKD framework considers objective and subjective measures and it was applied to students' dataset. The students' dataset contained the students' survey with the teacher's and course evaluation.

In the objective analysis, we wanted to identify the attributes from the dataset that best represented the student's model. For this purpose, we used multi-linear regression analysis and obtained a group of attributes that represented the model. We fed the Apriori algorithm with these selected attributes and obtained association rules. Then, we evaluated the association rules with objective measures: support, confidence and lift. These measures evaluated the relation between the attributes in the antecedent and the attributes in the consequent in the association rules. Rules with stronger support, confidence and lift represented strong rules.

The next step was the subjective analysis of the teacher's perspective and opinions. We prepared two surveys and one interview for the teacher to retrieve this information. The first survey, S1, was designed to obtain the domain knowledge from the teacher and the second survey, S2, was designed to obtain rules evaluated by him. In S1, we asked the teacher to analyse the questions presented in the students' survey. Therefore, we requested him to choose the most representative questions in the students' survey in general (section A) and in each area of the students' survey (section B). Then, the teacher wrote down possible association rules that he believed are followed during his classes. These steps constructed the domain knowledge.

The S2 contained the association rules with high support, confidence and lift, obtained by the Apriori Algorithm. After, we grouped the rules that had the same question in their consequent. Then, we asked the teacher to classify each group of rules using our extended list of subjective measures (interesting, useful, unexpected, obvious, previous and irrelevant).

As a result of the classification, the teacher identified rules that represented interesting rules, accepted beliefs (usefulness), and patterns that contradicted beliefs (unexpected). The same extended list offered three additional knowledge classifications: obvious, previous and irrelevant. These were used when the rule was already known (obvious), represented old knowledge (previous), or was a rule without importance (irrelevant) (Section E).

After the classification, we applied the meta-knowledge interview that would help us to obtain the teacher meta-knowledge from each of the association rules. For obtaining the meta-knowledge, we applied Flavell's theory (Flavell 1979) who applied question words (who, what, why), to retrieve meta-knowledge from a specific activity. We added the question words when and where. By asking these five question words, we obtained deeper information about a specific association rule, which included activities and strategies the teacher applied during his courses. These activities and strategies provided descriptions of the actions, elements, steps and approaches that were not in the students' survey. This new knowledge is called rule-meta-knowledge that helped augment the description of a rule, providing more meaning to it. When we added more elements associated with the description of an attribute, we enriched the rule-meta-knowledge by giving more detail. The integrated model included the “Improve Model” that contained interesting, unexpected and useful classified rules and the “Real Model” that included obvious and previous rules in the form of domain knowledge (Chapter 4).

2.3.2 Results associated to Objective 1, “To build a representation using students’ model and teacher’s model based on actionable knowledge”

An applied example is presented here. In Figure 2.2, we represent with white circles the association rules that correspond to the domain knowledge. We obtained these rules from the objective analysis of the students’ dataset and the subjective analysis teachers did regarding the association rules obtained from the objective analysis results.

All black circles represent interesting association rules and all circles with black and white represent association rules that connect domain knowledge with the interesting rules. Antecedents and consequents in each association rule are connected with a connector (small black circle)

For example, rule “X3, X14 -> X6” or rule “X9, X22 -> X7” represent association rules from the domain knowledge. Where X3 and X14 are considered attributes in the antecedent of the association rule, and X6 is the attribute in the consequent side of the association rule. The same happen with the rule “X9, X22 -> X7”, but X9 is also in the consequent of the association rule “X21, X1 ->X9”. Nodes X9 and X12 have double circles (white and black) representing the fact that these attributes are part of both DK and the group of interesting rules. We have two interesting rules represented by black and white circles: “X9, X22 -> X7” and “X22, X12 -> X16” where attributes X9, X22 and X12 are in the antecedent of the interesting rules, and X7 and X16 are in the consequent of the same interesting rules. These two rules complement the DK.

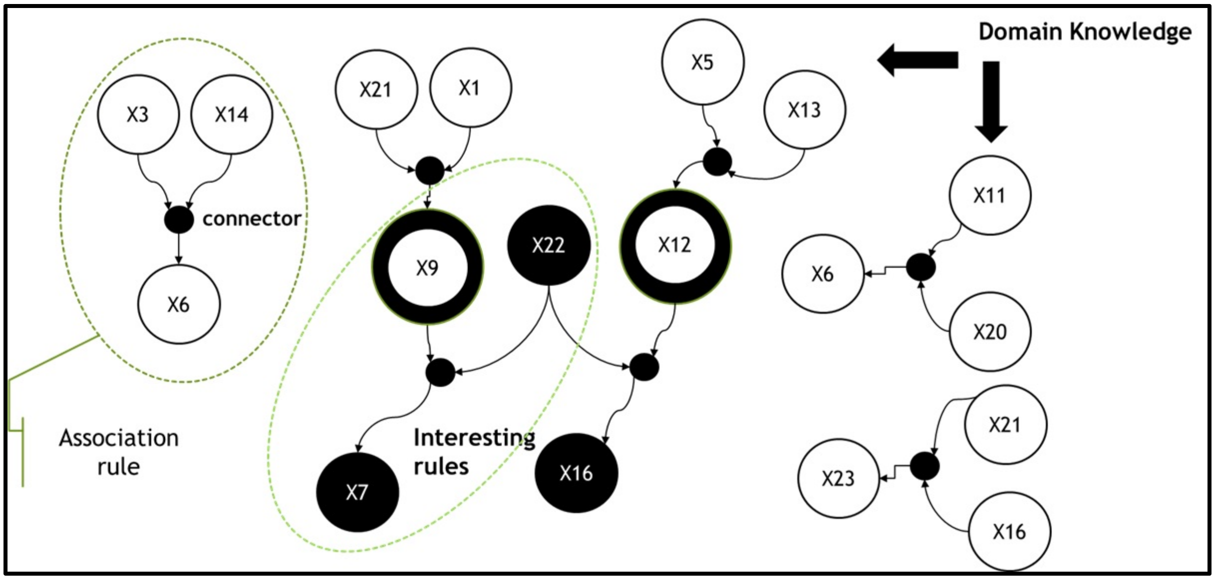

Figure 2.2 Domain Knowledge and Interesting rules representation

With this diagram in hand, we interviewed the teacher. As an example, let’s choose the interesting rule “X9, X22 -> X7” (see Figure 2.3). We asked the teacher about each attribute. Attributes X9 (“Teacher encourages students’ participation in class”), X22 (“Teacher provides the results of the assessment on time”) and X7 (“Teacher explains the course schedule”) represent attributes of the association rule that surprise the teacher because there should be no relation between these attributes. We applied the meta-knowledge questions, “What”, “Why” and “How”, to each attributes in the rule.

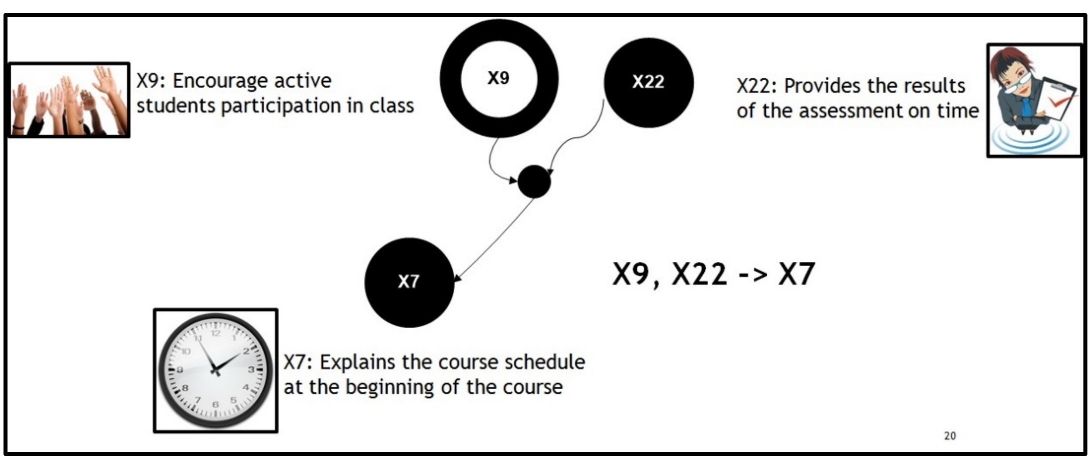

Figure 2.3 Interesting rule example

We obtained from the interview with teacher that attribute X9 ("Teacher encourages participation in class") is not only related to the action of teacher asking questions and students answering them, but also to other activities or behaviors such as "Reinforcing students' self-confidence", "Speaking openly of every topic" or "Always contrast ideas". The same occurs with attribute X22 ("Teacher provides the assessments of students on time"); in this case, that attribute is not only related to the action of returning the exams corrected, but also to the actions of presenting the solutions for the exam, finding mistake in the answer, relating the answer to the content learned in class, and acknowledging the students' difficulty to learn, or teacher's difficulty to explain the topic. Finally attribute X7 ("Teacher explains the schedule at the beginning of the course") is not only a reminder of the activities students have to accomplish during the semester; this particular teacher explains the schedule every two weeks; doing so, he found that students were understanding their position and their role within the course's timeline. Figure 2.4 provides a complete description of this interesting rule with the rule meta-knowledge associated to each attribute.

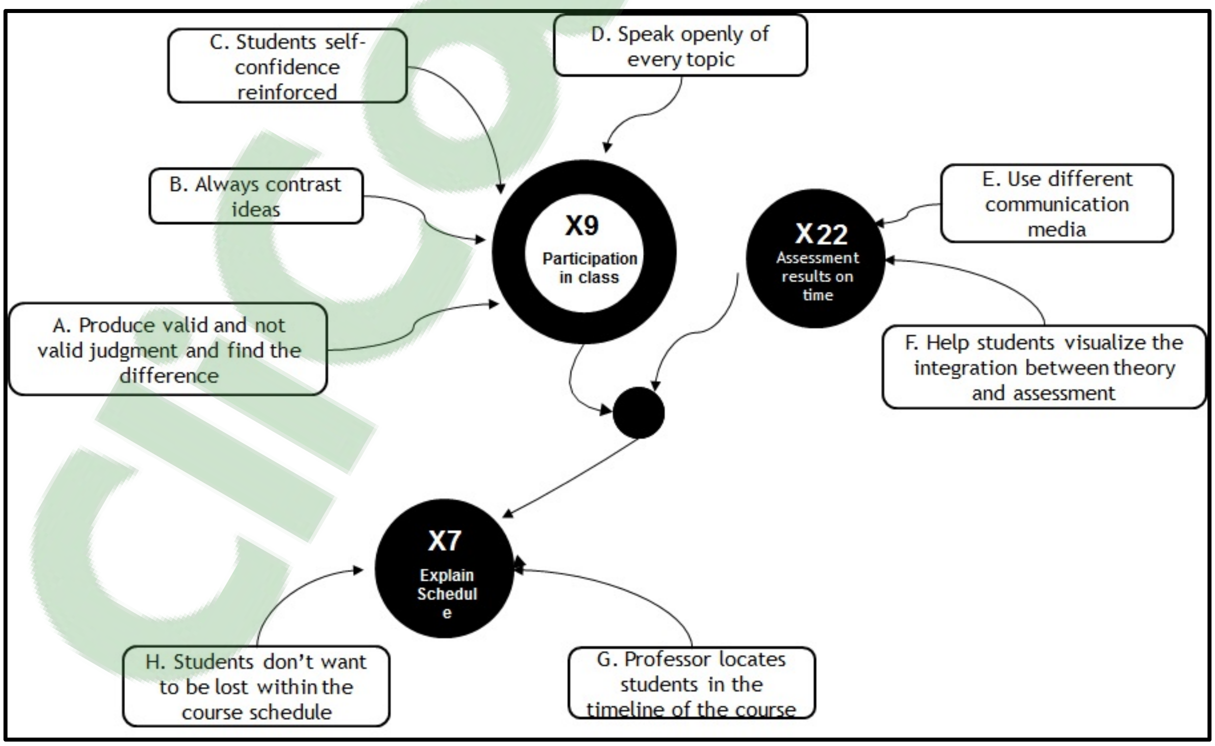

Figure 2.4 Improved model with the meta-knowledge included

Lastly, teacher relates all the attributes using the rule and their “meta-knowledge rules”, A, B, C, D, E, F, G and H. The teacher reported that the interestingness of this rule is that these three questions together, presented here as attributes of the association rule, encourage the participation in class (X9) and teacher’s assessment (X22), in order to make students understand their role within the course’s timeline (X7), which help them to be focused during the course.

The technical details about this contribution are presented in Chapter 4 along with the related methodology to obtain these results.

2.3.3 Objective 2: To construct a utility semantic measure to evaluate the usefulness of association rules within the model

We constructed the utility measure using a general utility mining approach that contains transaction utility and external utility, incorporating information from students’ survey and teacher’s surveys. The semantic aspect covers three components: students’ interest, teacher interest and recuperation effort.

The teacher’s surveys, S1 and S2, helped us define some necessary elements for the formula, such as the frequency of open and context association rules, along with their attributes. These open and context attributes were retrieved using the opinion of the teachers and their colleagues. Rules were labeled as interesting, unexpected and useful applying the extended list of knowledge classification.

First, the *Student interest* (Is) represents the teacher’s performance as a weight corresponding to the students’ point of view. This component contains three elements: the mean of the evaluations students gave to the teacher, a factor that shows how well the teacher did in comparison to how much he should have done, and finally, frequency of answered surveys.
Second, the *teacher interest* (It) represents the teacher's perspective, insights and experience about the evaluated attributes. This component contained five elements: the teacher’s attributes evaluation (taken from section A), the teacher’s attributes evaluations per area

(taken from section B), the frequency of each attribute in the antecedent from the categorized rules, the frequency of each attribute in the consequent from the categorized rules, and the perspective of the colleagues. We obtained this information from the survey applied to the teacher.

Finally, the *Recuperation Effort* (Re) is the effort the teacher has to make to improve an attribute. We determined the *Recuperation effort* based on the knowledge teacher had about the attribute and the complexity of the actions to improve it. During the teacher's survey, we asked teacher to classify the usefulness and actionability of survey questions hypothetically evaluated as low. The ranges go from very useful and straightforward to not useful and not straightforward. Depending on these categories, we defined our actual knowledge of the attributes in a Matrix of Recuperation Effort (Actionability vs Knowledge). See Figure 2.5 below:

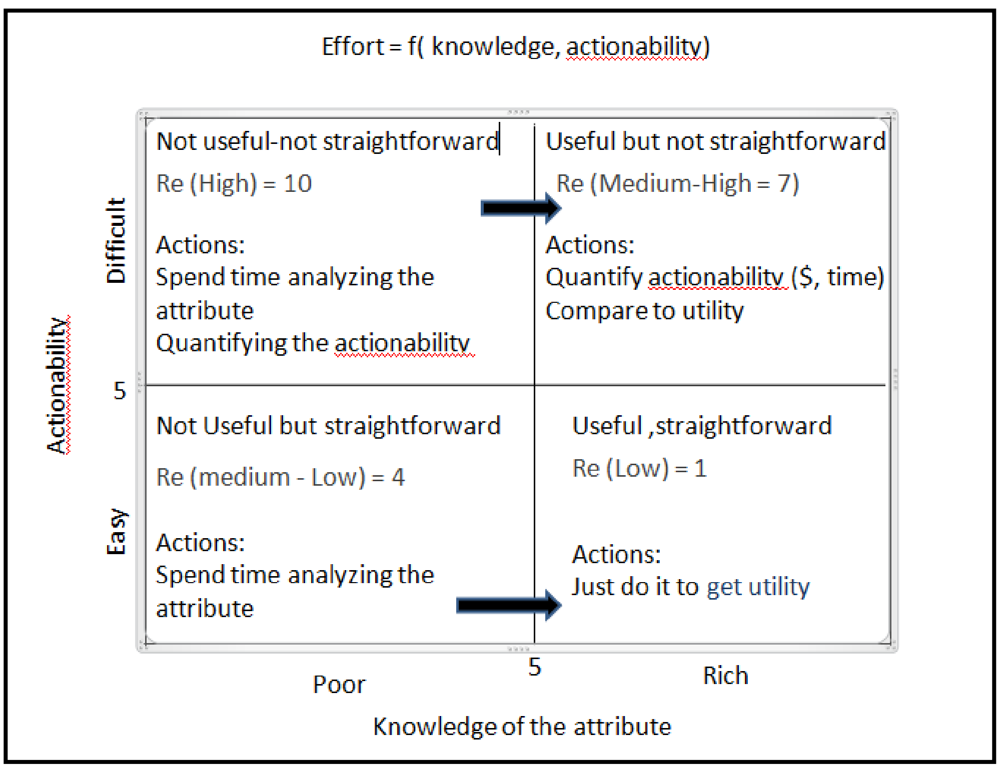

Figure 2.5 Matrix of Recuperation effort (Actionability vs knowledge)

We identified the actual knowledge about an attribute based on the Re teacher has to make in order to improve his evaluation. If the attribute was very difficult to improve, Re was very high (equal to 10). This amounts to say that the teacher was not aware of all the elements that affected the evaluation of the question. One example is when students' or teachers' attitudes in class were not addressed. When Re is medium high, 7, the teacher considers he can improve it, perhaps with additional training or learning new techniques to better explain his knowledge. This Re is more related to how he can improve his techniques and abilities. When the Re is medium low, 4, it is not simple to improve. It is possible that this is related to something that is out of the teacher's expertise. Finally, if the Re is low, 1, it means that maybe the teacher made some mistakes, but they are easy to fix. The total knowledge the teacher should have on an attribute minus the recuperation effort he needs to make represents the actual knowledge.

The utility formula is a function of the interest of the students (Is), the interest of the teacher (It) and the recuperation effort (Re). That is:

$$U = f(Stud.Interest, Teach.interest, Recuperat\_Effort) \tag{2.1}$$

2.3.4 Results associated to objective 2, "To construct a utility semantic measure to evaluate the usefulness of association rules within the model"

In Objective 2, we created a utility semantic formula that includes 3 elements: Interest of students (Is), Interest of teachers (It) and the Effort. Interest of students or "Is" is related to students` evaluation. Interest of teacher or "It" is related to teacher auto-evaluation obtained from the survey we applied to teacher; "It" contains the frequency of the attributes in the antecedent and consequent coming from the interesting rules teachers evaluated and the opinion of other colleagues about the attributes. Finally, the Effort is related to the knowledge teacher has about the attributes and the difficult to improve them; to obtain it, we used the analogy of levers, where a rigid bar is supported over a fulcrum and in each extreme there is an individual force, the Effort force, and the Actual Knowledge force. The Effort force is the quantification of the effort teacher has to do, and the Actual knowledge is the knowledge teacher has about the attribute based on the Recuperation effort or how hard it is to improve an attribute. The interest of the students (Is) and the interest of teacher (It) are represented in the analogy of levers as the distance between the fulcrum and the forces. Figure 2.6 present all these elements.

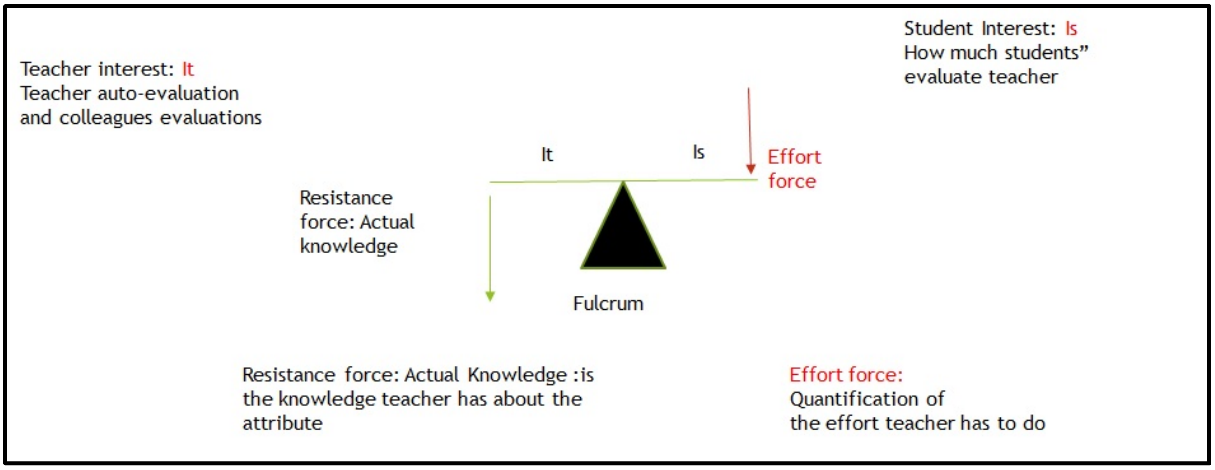

Figure 2.6 The Effort force, the Actual knowledge, Is and It

The Actual Knowledge is based on the Recuperation effort (Re). We mentioned that the Recuperation effort is constructed based on the actionability and the knowledge teacher has about the attribute. This is an example of how teacher from course 004 identified his Re associated to each of the attributes.

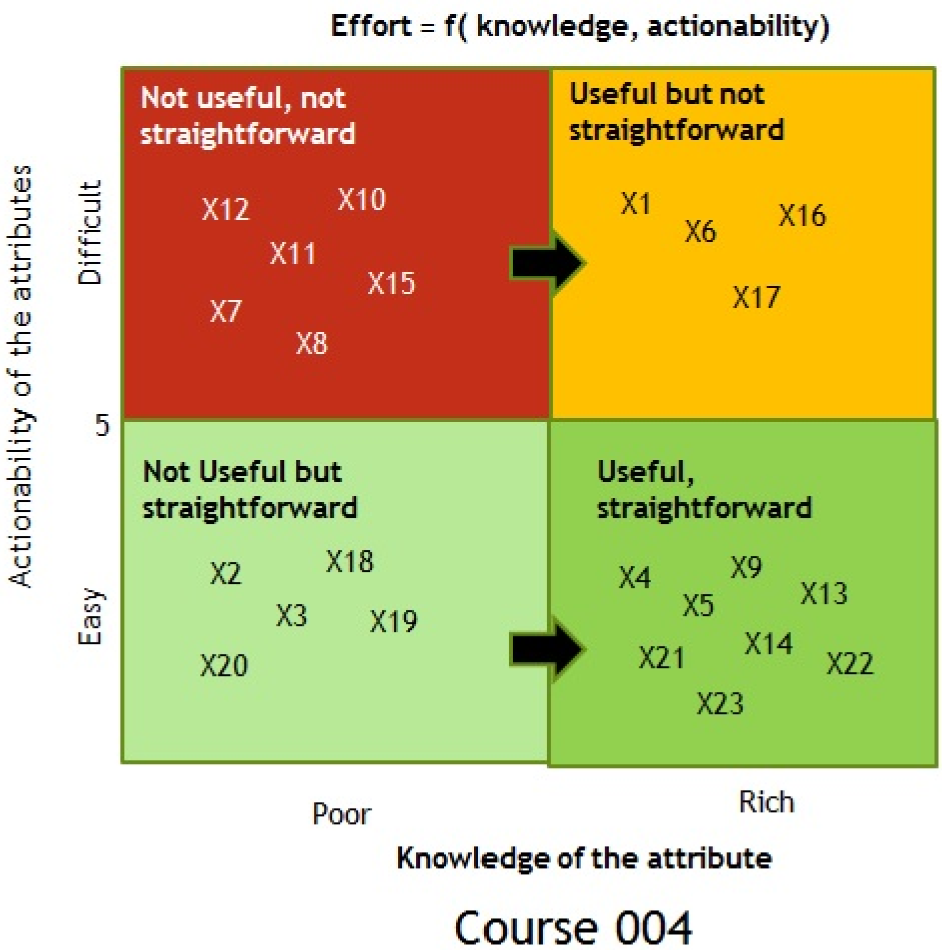

Figure 2.7 Matrix of Recuperation effort

In Figure 2.7, the teacher placed in the "Useful and straightforward" area, the following attributes: X4 "Evaluation fit themes", X5 "Assessments instructions", X9 "Encourages Participation", X13 "Gives Advise", X14 "Reflection", X21 "Uses Graphics and schemes", X22 "Returns Assessment results" and X23 "Punctuality". This means that these attributes are easy to be changed, which amounts to say that if this teacher receives low evaluation on these attributes, he will know exactly what to do and consequently, his Re is very low.

Teacher placed in the square labeled as "Not useful but straightforward" the attributes: X2 "Fulfills program", X3 "Evaluates participation", X18 "Gives Class Content organize", X19 "Gives Class content understandable", and X20 "Prepares Material". In this case, although the teacher understands the attributes, there is very little he can do since these attributes are related to elements out of teacher's control or expertise.

Then the teacher placed in the area labeled as "Useful but not straightforward" the following attributes: X1 "Uses Audiovisual help", X6 "Gives Motivation", X16 "Respectful" and X17 "Answers questions". These attributes are not easy to improve, but they are related to elements that teacher can improve, doing something like, taking courses, learning teaching techniques, etc.

Finally, the teacher placed in the area labeled as "Not useful, not straightforward" the following attributes: X7 "Presents Course schedule", X8 "Presents Class policies", X10 "Practices Summarization", X11 "Establishes Relation between concepts" and X12 "Motivates learning material". These attributes are very difficult to improve, and consequently, if the teacher has low evaluations on them, he has no clue about where to start.

It is important to notice that the classification of these attributes is based on the actionability and the knowledge of the questions for each teacher, and this example belongs to a particular teacher. Each teacher has his own classification.

Once a teacher has his own "Recuperation effort", it is possible to calculate the "Actual knowledge" of the attribute and then the "Effort". With the "Effort", teacher can apply the formula to find each of the partial utilities for each attribute.

The technical aspects of the utility semantic formula are detailed in Chapter 5.

Figure 2.8 presents a comparison between the students' evaluation and the partial utility for each of the attributes belonging to the teacher of the Programming Fundamental Course. The gray line presents the survey evaluation results from students' evaluations. The line shows a value between 8 and 8.5 as the teacher evaluation. It is pretty flat and therefore it is not easy to recognize any particular aspect that requires attention. The black line represents the partial utilities obtained from the utility semantic formula proposed by us. With this formula, we can see that teacher has peak values on attributes X1, X7, X8 and X10. The utility formula encourages teacher to use and apply the interesting rules that contains these attributes in particular.

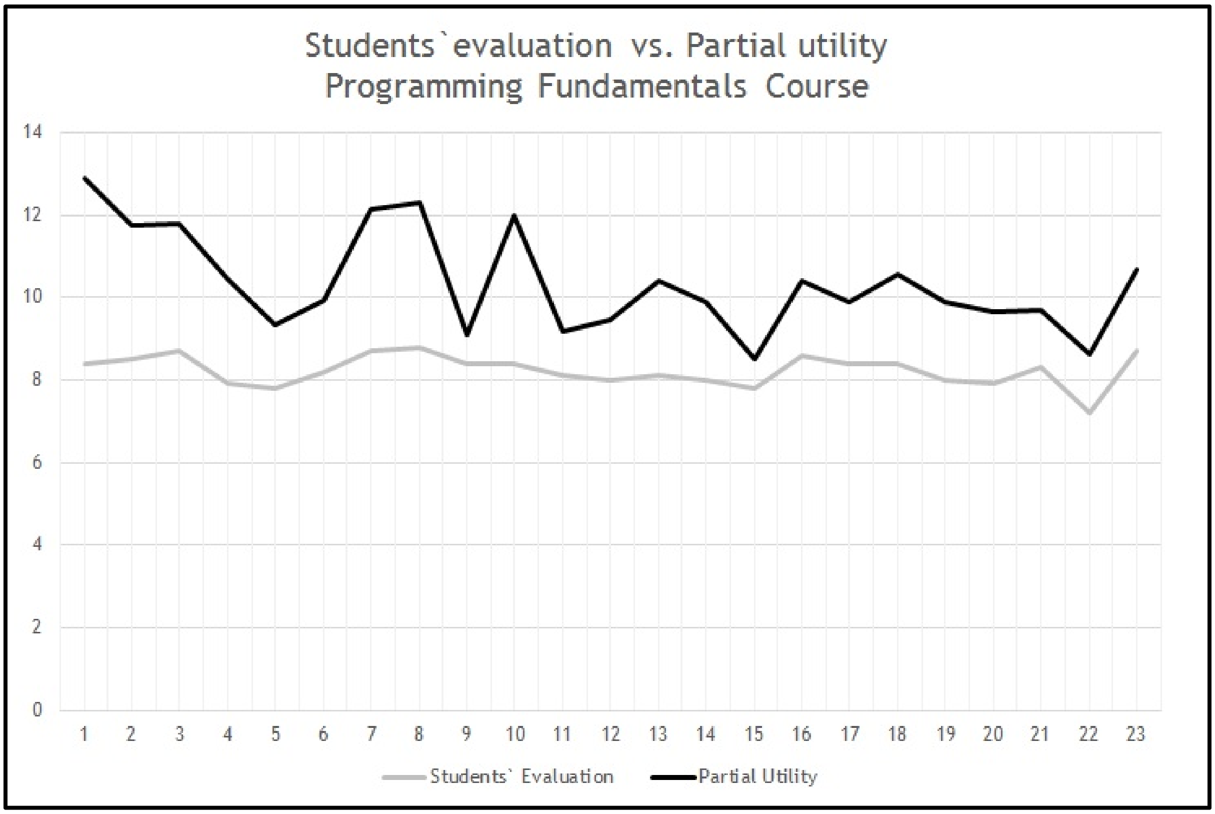

Figure 2.8 Students' Evaluation vs Partial utilities for Programming Fundamentals Course

Figure 2.9 shows another example for the teacher of the Research Methods Course. In this case, students' evaluation line and teacher's partial utility line coincide in different attributes as X7, X8, X10, X11, X12, and the highest values are in X9 and X14 as the highest attributes the teacher is encouraged to use in an interesting rule. The matches represent the questions where the teacher considered as his weaknesses, while the students considered them as the teacher's strengths.

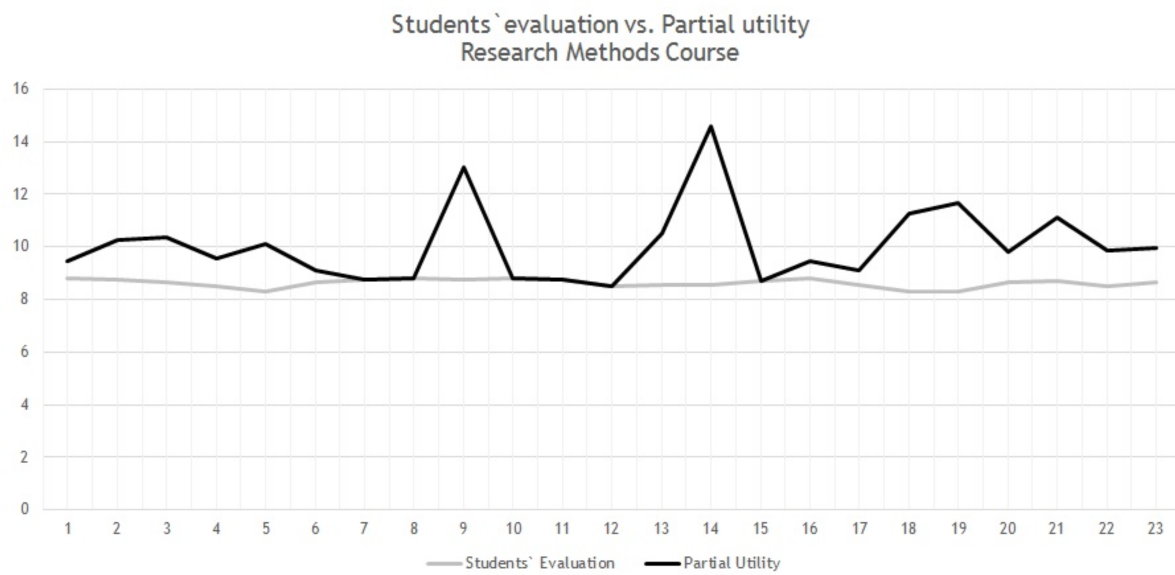

Figure 2.9 Students' Evaluation vs Partial utilities for Research Methods Course

Next we need to compare the results obtained from objective 1 and objective 2. We want here to compare both to see if teachers' perception is the same as students' perception. It is useful to recall at this point that the result of Objective 1 is a representation obtained from students' dataset, while the result of Objective 2 is the utility measure obtained from the teacher's perception.

2.3.5 Results of the comparison between Objective 1, "To build a representation using students' model and teacher's model based on actionable knowledge", and Objective 2, "To construct a utility semantic measure to evaluate the usefulness of association rules within the model"

We compare the two results obtained from the realization of objectives 1 and 2 to see if the model constructed from students' dataset has some relation to the utility semantic results obtained from teachers.

To do so, we constructed the representation for one of the course, "Web Application Development". We are not including the meta-knowledge analysis here since we only want to compare the presence of attributes in the rules for the two different perceptions.

In Figure 2.10, we present five rules, four are part of the domain knowledge and one is considered interesting.

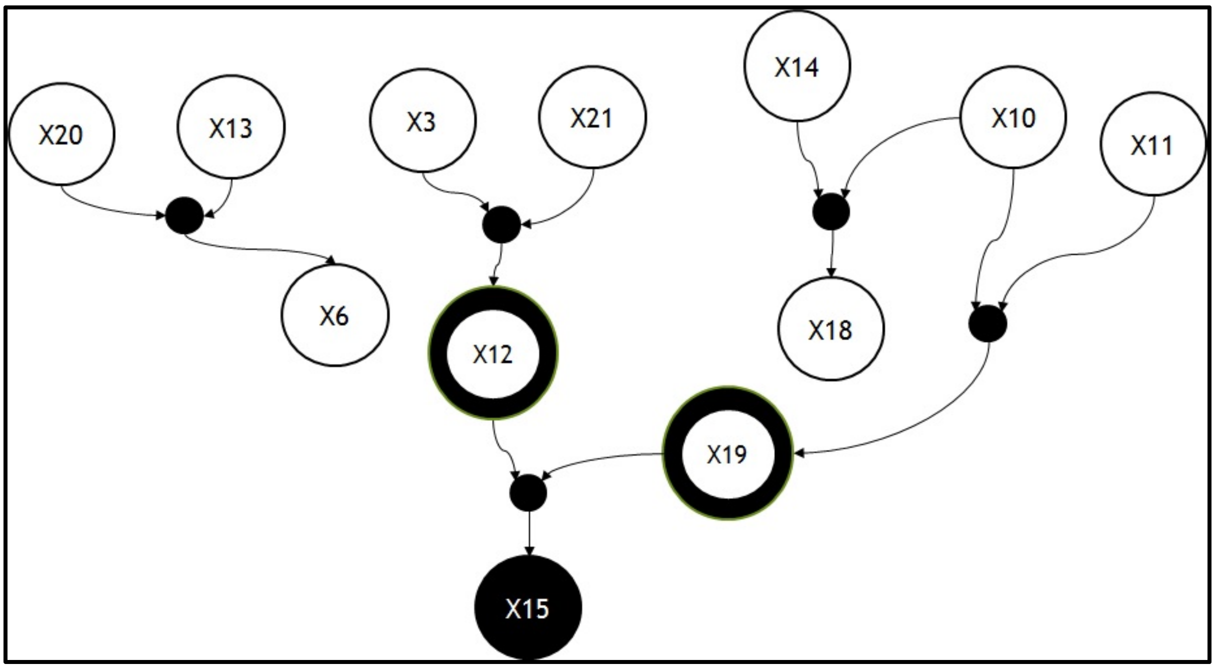

Figure 2.10 Web Application Development Course Representation

We obtained the partial utilities for this course. The partial utility in general has a range from 0 to 180. The higher utilities for this teacher attributes are in the range 61 – 80. In Figure 2.11, we present in dark gray, the attribute X21, X14 and X18. These attributes appeared in the students' representation, and in the table, with utility values between 61-80. The only attribute that was in the table but not in the representation of the rules is X9. There is a coincidence between the attributes appearing in the utility table with 61-80 ranges and the attributes in students' representation with the exception of X9.

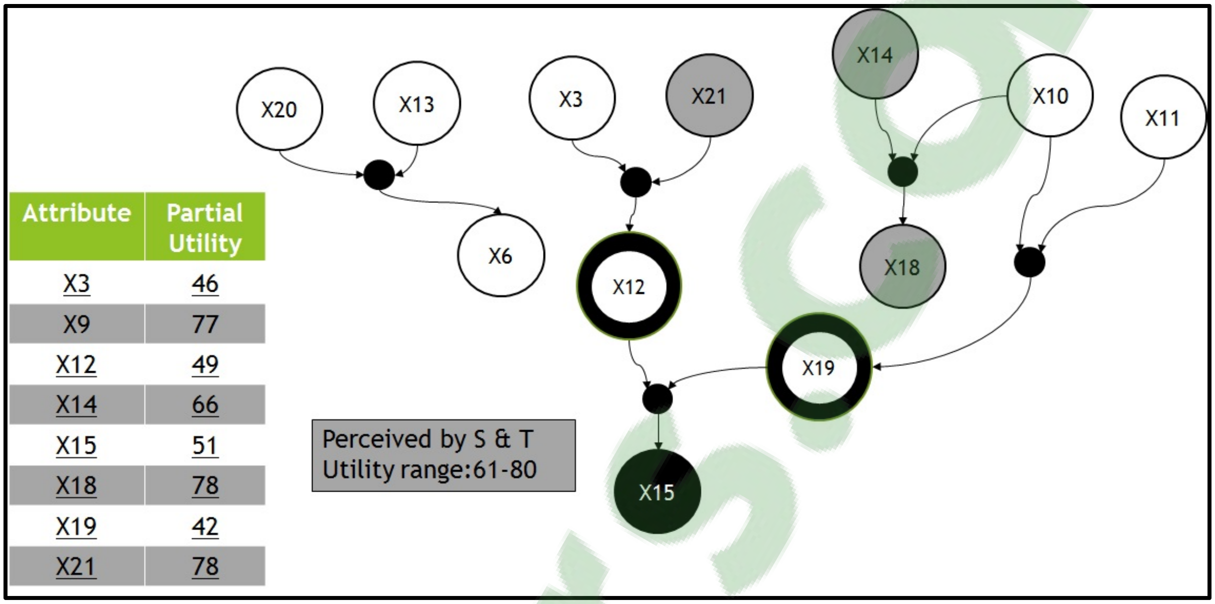

Figure 2.11 Web Application Development Course Representation with utility range 61-80

In Figure 2.12, we present in light gray, attributes with utility values within the 41-60 range. These attributes appeared too as part of the attributes in the students' representation.

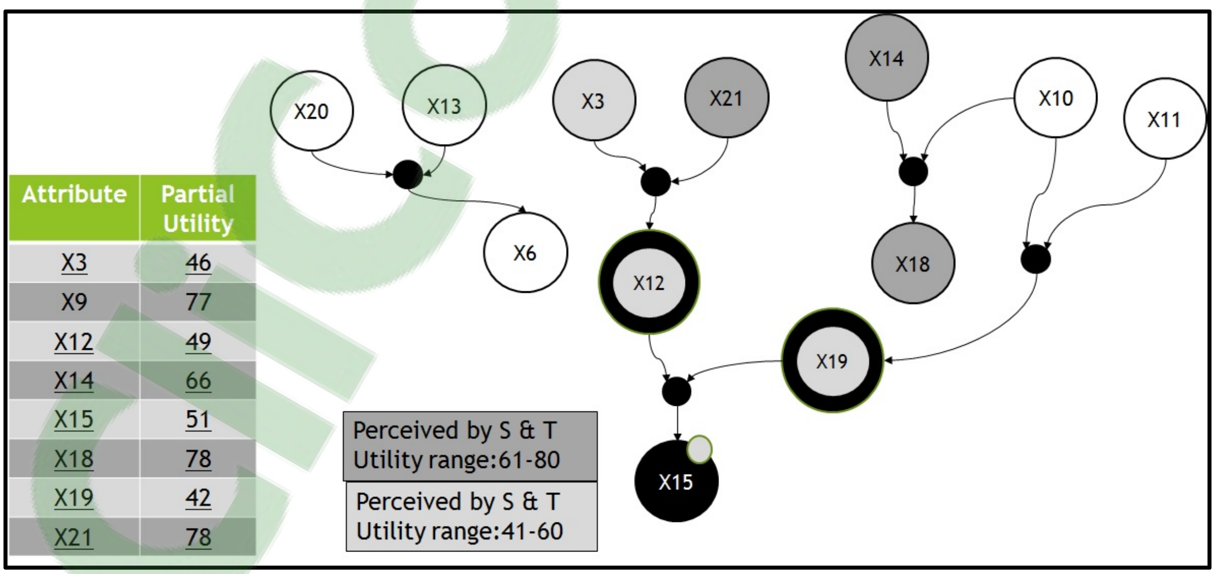

Figure 2.12 Web Application Development Course Representation with utility range 41-60

Only one of the attributes with utility range between 61-80 didn't appear in the students' representation, X9. This amounts to say that attribute X9 was considered a strength only by the teacher.

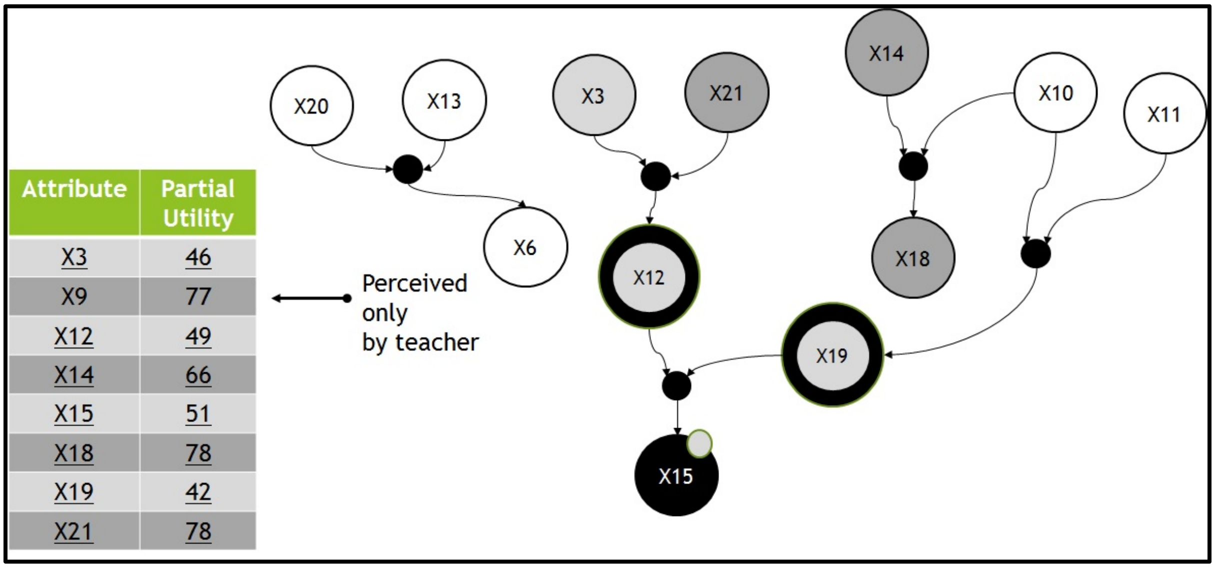

Figure 2.13 Web application development Course- Teacher perception

Figure 2.14, shows the distribution between all the attributes and their utilities in a condensed form. Most of the attributes have utilities in the 21-40 range.

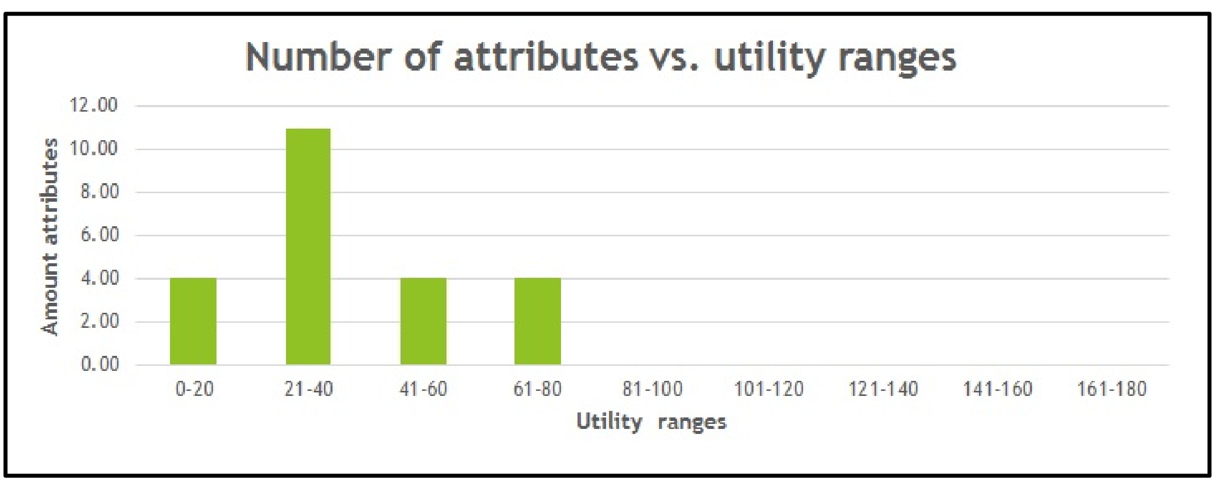

Figure 2.14 Distribution of the partial utilities

Figure 2.15 shows the details for each partial utility of the attributes for the "Web Application Development" course

Attribute	Partial utility	Attribute	Partial utility
X1	22.40	X13	34.51
X2	18.14	X14	66.29
X3	46.03	X15	50.69
X4	12.72	X16	34.11
X5	22.62	X17	34.37
X6	19.85	X18	78.31
X7	28.40	X19	41.90
X8	27.43	X20	28.33
X9	77.25	X21	78.13
X10	39.69	X22	13.67
X11	34.28	X23	28.97
X12	48.56		

Figure 2.15 Partial utilities details.

To test the application of the utility measure, the following association rules have to be evaluated. The main idea here is that each rule will be evaluated using the partial utility for each attribute. The association rule with the higher mean utility, after adding and obtaining the mean, is the one that is encouraged to be followed:

(a) X12(49), X1(22) -> X9(77) then *Urule*=49
(b) X12(49), X11(34) -> X9(77) then *Urule*= 53
(c) X12(49), X13(35) -> X9(77) then *U*rule=54
(d) X12(49), X18(78) -> X9(77) then *U*rule=68

Rule (d) contains the highest utility value. This rule contains the attributes X12, X18 and X9. Using Table 2.1, we can better explain rule (d); X12 "Teacher motivates the learning of the course materials" and X18 "Teacher delivers class content in an organized way" are attributes that can help teacher reach a high evaluation in X9 "Teacher encourage active students' participation in class", based on his own knowledge about his best abilities.

Table 2-1 Attributes' description

Attribute	Description
X1	Uses audiovisual help to support the content of the class
X9	Encourage active students' participation in class
X11	Establishes relationships between new concepts and those already known whenever is possible
X12	Motivates learning of the course material
X13	Is willing to answer questions and offer advice within and outside the classroom
X18	Delivers class content in an organized way

2.3.6 Objective 3: To discover new topics from the interviews analysis to improve teacher evaluation

First, we looked for new topics inside the teacher's interviews that were not expressed in the students' survey. We fed a topic modeling tool with the interview information (textual comments). The objective of this step was to obtain clusters of words that could represent the topics the teacher considers important. With this in mind, we did several tests ranging from 3 to 20 topics, and obtained a final number of topics. Finally, we matched the questions from students' survey and its descriptions to the final topics revealed by the topic analysis. These descriptions contained information found on Internet, research papers related to the question, and abstracts; they constituted what we called the dictionary and are used to classify the themes. To exemplify the theme classification, we will use question 15 of the students' survey (communication). Thus, we constructed the description of question 15 (communication concept and types of communication) and classified all the themes from the interviews related to communication.

Second, we tested each of the theme from interviews and observed how the model classified them. For this purpose, we compared the similarity between the topic description and the interview theme. If the interview theme was similar to the topic, then the theme could be classified as that topic category. To measure the similarity, we applied cosine. The six topics and each of the 146 themes in the interviews were converted to vectors. Each topic was compared to all the interviews' theme. If the cosine value between the topic vector and the theme vector was very near to 1, the topic and the themes were similar, but, if the cosine was lower than 0.7, it meant that there was little or no similarity between the topic and the theme.

Finally, once all the interviews' themes were classified, we tested the precision of the classification model with experts. We asked two experts to classify manually the themes coming from the interviews. The experts found the same theme could represent more than one topic at a time, so we allowed them to classify a theme in two or more possible topic categories. We compared the model's classification with the experts' classification using Kappa measurement. Most of the comparisons between these two, resulted in a Kappa coefficient greater than 40 on a scale from 0 to 100

2.3.7 Results of Objective 3, "To discover new topics from the interviews analysis to improve teacher evaluation"

From the topic analysis using LDA, we obtained six groups of words that have to be named based on the words each topic presented:

Table 2-2 Topic analysis - LDA results

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6
understand	express	time	research	practical	code
concepts	confidence	attention	career	advice	project
exam	person	assistant	interested	steps	structures
feedback	opinion	suggest	motivated	easily	programming

evaluation	talk	focus	friends	jokes	instructions
asks	behaviour	part	people	material	program
complicated	likes	styles	interest	remember	abilities
easy	speak	regular	motivate	exercises	writing
applied	afraid		moment	application	
resolve	personal		topic		
problems			curiosity		

A group of teachers revised the words and assigned a name to the topic based on them. Table 2.3 shows the group of words with the assigned topic name.

Table 2-3 Topic names

Topic 1 **Evaluation and feedback**	Topic 2: **Feeling and communication**	Topic 3 **Time and effort**	Topic 4 **Critical thinking**	Topic 5 **Content and delivery**	Topic 6 **Technical aspects**
understand	express	time	research	practical	code
concepts	confidence	attention	career	advice	project
exam	person	assistant	interested	steps	structures
feedback	opinion	suggest	motivated	easily	programming
evaluation	talk	focus	friends	jokes	instructions
asks	behaviour	part	people	material	program
complicated	likes	styles	interest	remember	abilities
easy	speak	regular	motivate	exercises	writing
applied	afraid		moment	application	
resolve	personal		topic		
problems			curiosity		

The construction of the dictionary was done with the description of each of the questions. This includes what the question represents, how the question is accomplished, some synonyms about the question and some paper abstracts from academic journals that contain related description of the question. In Figure 2.16 we present an example of the question X6: "Teacher motivates students to do additional research"

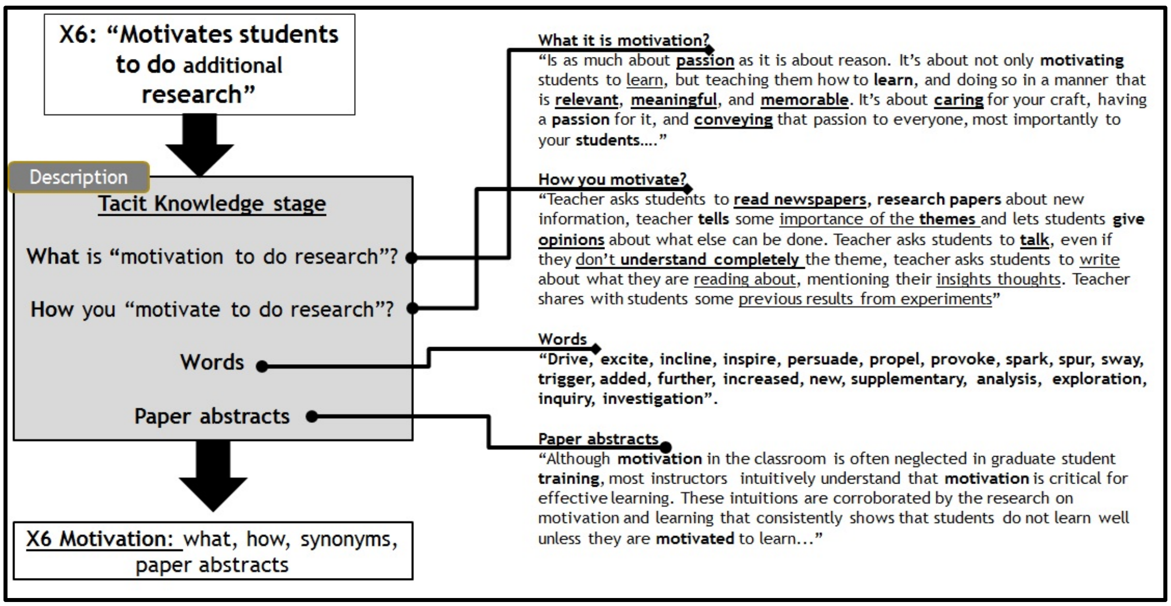

Figure 2.16 Dictionary description Question X6

After this, we matched the topics already generated with the best suited survey questions and their descriptions. Figure 2.17 shows topics being matched with some questions. For example, topic 1 was matched with question X3, X4, X5 and X22.

Next step is to perform the classification step using the dictionary and the found topics.

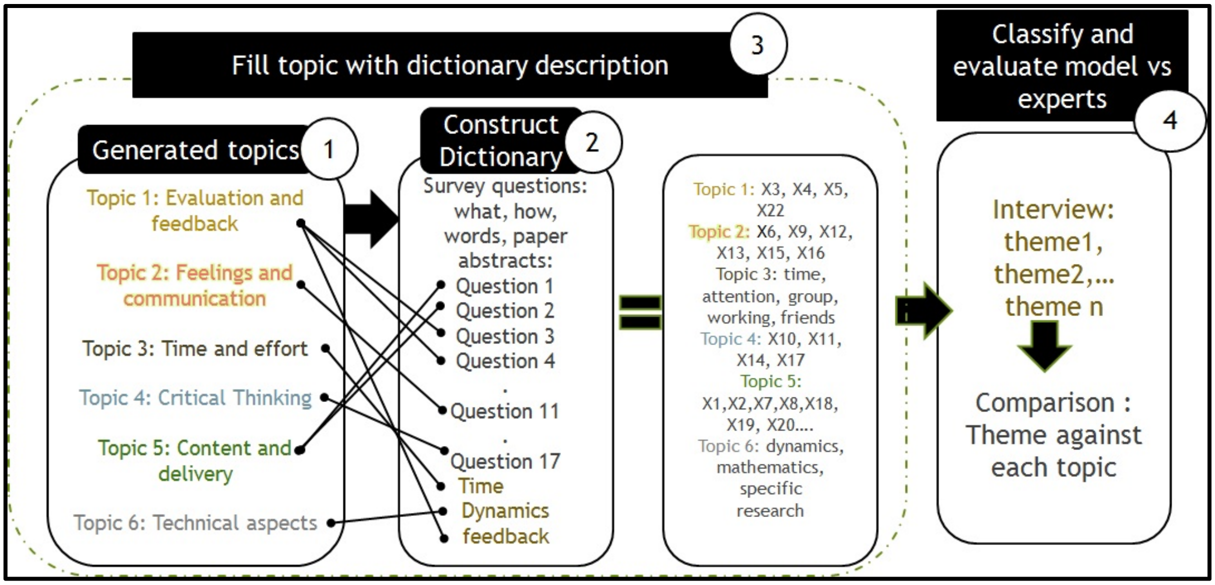

Figure 2.17 Topic to questions assignment

We divided each interview in paragraphs and used LSI to classify each paragraph into a topic; we compared them (topic vs paragraph) using cosine similarity; human experts did the same classification using their judgments. We gave them the topics and asked them to classify each interview paragraph into one topic. We evaluated the results between experts, and between each expert and the model with Cohen Kappa measure. The following table presents the results

Table 2-4 Cohen Kappa's results

	Expert 1 vs. Model	**Expert 2 vs. Model**	**Expert 1 vs. Expert 2**
Interview 1	0.46	0.52	0.70
Interview 2	0.63	0.63	0.78
Interview 3	0.53	0.40	0.50
Interview 4	0.56	0.72	0.60
Interview 5	0.28	0.78	0.51

	Expert 1 vs. Model	Expert 2 vs. Model	Expert 1 vs. Expert 2
Interview 6	0.23	0.46	0.52

As we see in Table 2.4, most of the coefficients illustrate a moderate or substantial agreement, with the exception of Interview 5 and 6. These interviews were the first being distributed to the experts. We had two experts, one from the local university and one from an external university. Expert 1 was from the external university and was not use to this kind of evaluations. We had to explain in detail how the classification works to obtain some results. Expert 2 was from the local university. This difference of origin between experts could explained the lower coefficients for interview 5 and 6 from expert 1 against the others interviews results.

We present some examples of the topic classification made by the machine. The bullets points contain extracts from the interview paragraphs classified as "Feelings or communication", "Time and effort", "Critical thinking", "Content and delivery", "Evaluation and feedback", and "Technical aspects".

Feelings or communication

- "Students don't want to contradict teacher because of shyness"
- "Students are overwhelmed with the amount of work they have to do during a project"

Time and effort

- "Students have to put a lot of effort to understand projects already done"
- "Students work with friends to be more effective with their time"

Critical thinking

- "Teacher leads the reflection during class"

- “Reflection is something that is not perceived like an individual activity in class but is considered implicit in students` activities”

Content and delivery

- “Course organization is part of the content of the course”
- “Digital tools, as content managers or blogs help teacher to integrate the content with students’ doubts”

Evaluation and feedback

- “Teacher gives feedback during exercises to students”
- “Students have to include more than one opinion in their judgments”

Technical aspects

- “Structure programming course are very related with C programming courses but students could have forgotten what they learnt from one semester to another”
- “Teacher talks about perseverance, knowledge and perception to explain creativity”

More technical details about this objective are available in Chapter 6.

2.3.8 Conclusions

In section 2.3.2 and 2.3.4, we presented preliminary individual results for the first two contributions: the representation using students’ model and teacher’s model and the construction of a utility semantic measure to evaluate the usefulness of association rules within the model. In section 2.3.5, we showed the comparison between these two contributions. The comparison showed how both models are integrated and what means the presence or the absence of the attributes in each model.

In section 2.3.7, we showed the preliminary results for contribution 3. We found three new topics not considered in the survey questions. Additionally, we noticed that the interviews contain feelings and expressions of communication that is not possible to see using the regular survey.

CHAPTER 3

DESCRIPTION OF THE INSTRUMENTS TO OBTAIN KNOWLEDGE: INTERVIEW AND SURVEYS

3.1 Introduction

In chapter two, we presented the proposed methodology with the general research objective and the three particular objectives. As general research objective, we aimed to create a methodology to integrate the objective analysis of students with the subjective considerations of teachers. For this purpose, we applied a D3M methodology based on AKD framework and interestingness measures (objective, subjective and semantic). With this intention, we performed the construction of a representation using students' perceptions and teacher's perception based on actionable knowledge discovery; the construction of a utility semantic measure to evaluate the usefulness of association rules within the model and finally the discovery of new topics from the interviews analysis to improve teacher evaluation. To accomplish these objectives, we designed three tools to retrieve knowledge from teachers. These tools were one interview and two questionnaires, S1 and S2.

The total duration of the two surveys` application and the interview is between three and four hours. The two surveys were sent by email and the answers were received in the same way. All the interviews were conducted by video conference using Skype or in person. We describe the time spent per instrument in the next section. The results of these surveys are presented in the section 3.3 of this chapter.

3.2 Questionnaires for Teacher's subjective perspective (S1 and S2)

We constructed survey S1 that contained five sections. In Section A, teacher identified questions that were of his interest based on his course experience. In Section B, teacher selected the questions per area that were more important for him marking them. In section C, the teacher identified the actionability and usefulness of the answer suggested when it has a low evaluation. The teacher used the following four categories of "actionability" and “usefulness” to assess each question: "do not know how to use the results," "useful, but not easy to change "," perhaps the outcome is bound to something else" and "useful and simple”. In section D_a, the teacher created open association rules using variables from Section A; in section D_b, the teacher created contextual association rules using variables from Section B; the evaluation time S1 was 1.5 hours. Finally, survey S2 is presented in section E, the teacher evaluated students` association rules using 6 knowledge categories. We present each section of the questionnaire for teacher’s subjective perspective with and example. The Figure 3.1 shows teacher’s survey sections and the results obtained in some sections.

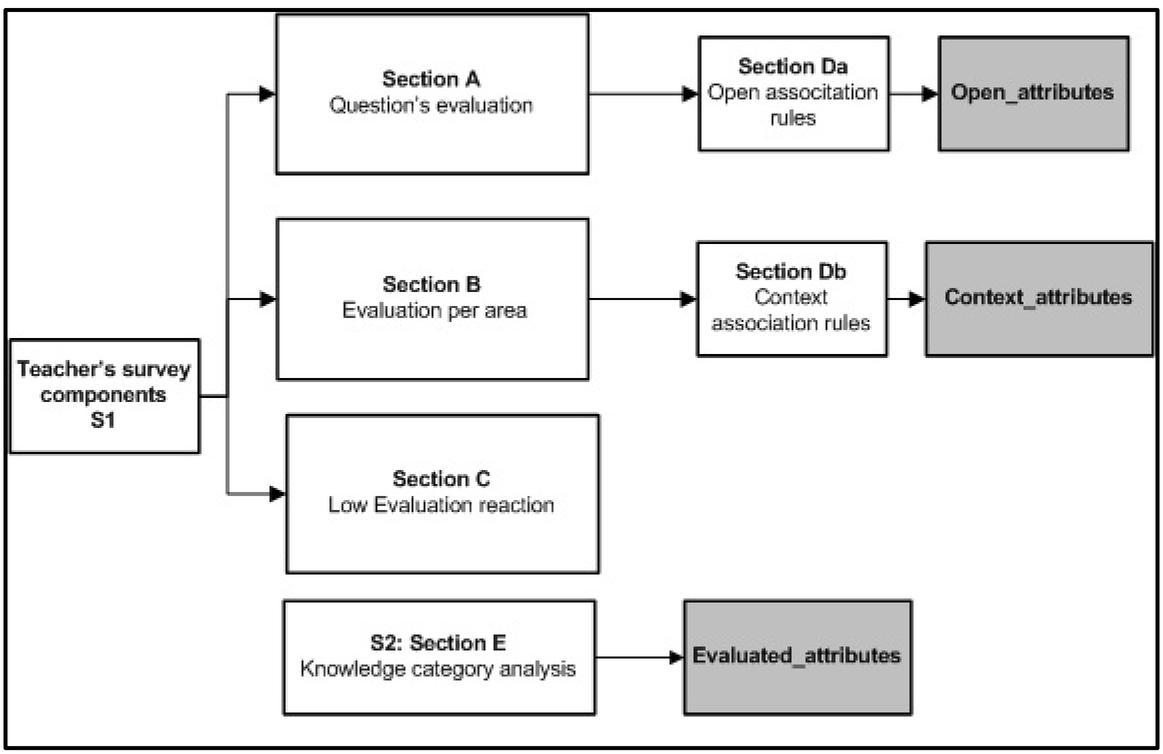

Figure 3.1 S1 Survey’s Components

3.2.1 Section A: Question evaluation

The goal of this section is to obtain from teacher a more precise understanding of what is evaluated in each survey question. In this section experienced teachers express their opinion about each survey question assigning a value between 0 (when he considers that a statement is irrelevant) and 10 (when he considers that it is extremely important). This value depends only on the subjective thinking of the teachers and their experience about each survey question. For example, question number 9 of the Figure 3.2, "Teacher develops class content in an organize way" is evaluated high if teachers consider it as fundamental for their courses. Figure 3.2 shows the question evaluation sheet.

Questions	NA	Not important	Less important	+ o - important	Very important	Extremely important
9. Teacher develops class content in an understandable way	☐	☐	☐	☐	☐	☑
10. Teacher prepares instructional, bibliographic or other resources to facilitate learning	☐	☐	☐	☑	☐	☐
11. Teacher responds to questions in class about subjects related to the field	☐	☐	☐	☐	☐	☑
12. Teacher explains class policies at the beginning of the course	☐	☐	☑	☐	☐	☐
13. Teacher encourages active student participation in class	☐	☐	☐	☐	☐	☑
14. Teacher summarizes key ideas discussed before moving to a new unit or topic	☐	☐	☐	☑	☐	☐
15. Teacher establishes relationships between new concepts and already known concepts, whenever possible	☐	☐	☐	☑	☐	☐
16. Teacher motivates learning of the course material	☐	☐	☐	☐	☐	☑

Figure 3.2 Example of Section A – Question evaluation

Figure 3.2. shows de content of “Section A – Questions evaluation”. Each question present six boxes, where teacher selects only one of them, depending on the importance he gives to each question. In the figure, the teacher considered extremely important only questions 9, 11, 13, and 16.

3.2.2 Section B: Evaluation per area

Section B evaluated the three areas that appeared in the student survey (*design*, *learning promotion,* and *production of learning materials & education management).* Table 3.1 shows the number of questions per area and the maximum numbers of questions we asked the teacher to choose as the most relevant from each area. For example, for the Design area comprising six questions, the teacher could only choose up to three questions with higher relevance.

Table 3-1 Questions per area and maximum number of questions

Area in Student Survey	Questions per area	Max. # of questions chosen by the teacher
Design	6	3
Learning promotion of the course	13	6
Production of learning material & education management	4	2

First, the teacher had to select the most important area for him. For example, in Figure 3.3 teacher chose *Learning promotion of the course* and *Production and teaching materials for the course & Education management*

2. Statements from 1 to 23 (question 1), focus on the evaluation of specific areas of the courses ("Design", "Learning Promotion", "Production and Material &Education Management" area). Which of these areas are more important for you? Choose all the areas that are important for you (You can choose one or more)

☐Design of the course
☑Learning promotion of the course
☑Production and teaching materials for the course & Education management
☐None of the above

Figure 3.3 Example of Section B – Evaluation per area

Once he selected the areas, he had to select the questions per area based on Table 3.1. The table presented the maximum number of questions he could choose. In Figure 3.4, we present the *"Learning promotion of the course"* area selected questions.

6. If you checked the **Learning promotion of the course** area in **question 2**, which of the following statements are more important when you analyze your own development as professors. Select up to 6 statements.

☐7 Professor explains the course schedule at the beginning of the course
☐8 Professor explains class policies at the beginning of the course
☑9 Professor encourages active student participation in class
☐10 Professor summarizes key ideas discussed before moving to a new unit or topic
☐11 Professor establishes relationships between new concepts and already known concepts, whenever possible
☐12 Professor motivates learning of the course material
☐13 Professor answers questions and offers advice within and outside the classroom
☑14 Professor promotes reflection on topics covered
☐15 Professor maintains fluid communication with students
☐16 Professor is respectful to students
☐17 Responded to questions in class about subjects related to the field
☑18 Professor delivers class content in an organized way
☑19 Professor develops class content in an understandable way
☐None of the above

Figure 3.4 Example of Section B – Selection of attributes per area

In *Learning Promotion of the Course,* the max number of questions the teacher could choose was up to six questions. In figure 3.4, the teacher chose four questions 9, 14, 18 and 19.

As we can see, *section A* and *section B* evaluated the same survey questions but in different conditions. *Section A* did it in a general way, leaving freedom to the teacher to choose among all questions. In *section B,* the teacher focused on one specific area to be evaluated. He chose which questions were more interesting within this area, and left out other questions that he considered irrelevant or not useful. At this point, the teacher assigned values to each question in section A, from the students' survey. These values might change when we asked the teacher to evaluate again the same question focusing on the area (Section B). Therefore, if a question had a value of 8 (very important) in *section A*, the teacher was allowed to give another value to the same question in *section B (zero if teacher didn't choose it, ten if he chose it)*. In Figure 3.2, teacher chose questions 9, 11, 13 and 16; but in figure 3.4, teacher chose questions 9, 14, 18 and 19. Only one question was considered always important, question 9.

3.2.3 Section C: Reaction to low evaluation

In this section, we confronted the teacher to the hypothetical situation of receiving a low evaluation in the students' survey questions and he determined how hard it was to improve these results for next semester. The aim was to understand what the teacher knew about the questions' results from the survey and if he could interpret and improve the results he retrieved from this section. Thus, we asked the teacher to quantify what he could do with these results using four different categories. These value categories went from 10 to 1, where ten means lack of understanding of the results, and 1 means total comprehension of what to improve. **Table 3.2** shows the categories for this *reaction to low evaluation*:

Table 3-2 Categories of reaction to low evaluation.

Categories	evaluations
a) "I don't know how to use these results"	10
b) "Results are useful but not easy to change"	7

Categories	evaluations
c) “Maybe the results do not depend only on the teacher”	4
d) “Results are useful and straightforward”.	1

When the teacher identified low evaluated question as difficult to improve, he assigned a value of 10. This means, the teacher didn’t know why this question was low evaluated. When the teacher knew why he received a low evaluation but he realized that fixing it would require more resources as time, money, knowledge, attitude, etc. he assigned a 7. In other words, he needs to improve something in class or in his teaching style. When the teacher knew that the low evaluation occurred due to causes that were above the course domain, he assigned a value of 4. For example, a cause could be the students’ personal situation. Finally, if the teacher knew exactly where the problem was and what he needed to modify in his class, he assigned a value of 1. This means he could apply the changes in order to get a higher evaluation next time, Figure 3.5 shows an example of this reaction to low evaluation section.

Statements	I don’t know how to use the results	Useful but not easy to change	Maybe the result is related to something else	Useful and straightforward
1. Professor seldom uses audiovisual help to support the content of the class	☐	☐	☐	☑
2. Professor seldom fulfills the program proposed at the beginning of course	☐	☐	☑	☐
3. Professor doesn’t evaluate periodically student participation in class	☐	☐	☐	☑
4. Evaluations used in class don’t fit the themes developed in class	☐	☐	☑	☐

Figure 3.5 Example of Section C- Reaction to low evaluation

In figure 3.5, if teacher received a low evaluation in question two, he considered that the results are related to something else like in question four.

3.2.4 Section D_a: Creation of open association rules using variables from section A

In this section, we created open association rules. Hereafter, we will refer to the survey questions as attributes. To construct the open association rules, we used the 23 attributes from the survey. We defined *open association rules* (henceforth *open rules*) as the rules constructed using attributes from all areas; these open rules were of the form a →b (where "a" and "b" were attributes or a group of attributes that belong to the attributes group X1 to X23. Thus, attributes used to create an *open rule* were called *open attributes*.

In section D_a, we invited the teacher to create up to five rules with the attributes that were evaluated high in *section A* and placed them in the consequent side of the association rule. We asked them to construct only five rules because, otherwise, it would be too overwhelming for the expert to construct and then to explain each of the association rules and each of the attributes. Then, from the entire group of attributes, he chose two attributes and placed them in the antecedent side of the open rule. In other words, the teacher created association rules of 3-itemset (two attributes in the antecedent and one attribute in the consequent). We instructed teacher not to use in the antecedent the same attribute that he had placed in the consequent. The antecedent → consequent open rule construction denoted which attribute placed in the antecedent appeared at the same time with an attribute place in the consequent based on general considerations. The teacher may try different options of attributes in the antecedent for the selected attribute in the consequent. In Figure 3.6, we see that *Attribute X1* and *Attribute X2* belong to the antecedent side of the association rule and *Attribute Y* belongs to the consequent side of the association rule.

	ANTECEDENT			CONSEQUENT
	ATTRIBUTE X1	**ATTRIBUTE X2**		**ATTRIBUTE Y**
RULE A	X3	X15		X14
RULE B	X20	X13		X6
RULE C	X15	X13	**THEN**	X9
RULE D	X3	X21		X12
RULE E	X10	X11		X19

Figure 3.6 Example of open rules creation

Once each rule was constructed, the teacher selected only one of the two attributes in the antecedent of each rule. The final rule had one attribute in the antecedent and one attribute in the consequent. The reason for eliminating one of the antecedents was the teacher weighted implicitly the relative importance of the attributes in the antecedent and chose only one. The remaining attributes, those in the antecedent or in the consequent of these rules, were called *open attributes*. Figure 3.7 shows the remaining attributes. The teacher eliminated attributes X15, X13, X15, X3 and X10 respectively.

	ANTECENDENT ATTRIBUTE X		CONSEQUENT ATTRIBUTE Y
RULE A	X3	THEN	X14
RULE B	X20		X6
RULE C	X13		X9
RULE D	X21		X12
RULE E	X11		X19

Figure 3.7 Example final attributes (open rules)

3.2.5 Section D_b: Creation of contextual association rules using variables from section B

We defined *Context association rules* (henceforth *context rules*) as the rules constructed using attributes that belong to specific areas. These context rules are of the form c → d, where c and d are an attribute or group of attributes that belong to the *design area* (X1 to X6); f → g, where f and g are an attribute or a group of attributes that belong to the *learning promotion of the course area* (X7 to X19), and r →t , where r and t are an attribute or a group of attributes that belong to the *production of learning materials & education management area* (X20 to X23). Each attribute that was part of a *context rule* was called a *context attribute*.

In section D_b, the teacher created *context rules* for each area, using only the attributes selected in *section B*. In *Section B - evaluation per area,* the teacher selected a group of variables per area that were the most interesting for him. The teacher placed one of the selected attributes in the consequent side of the rule and two attributes in the antecedent side of the rule. For example, if the teacher selected in the *design area* attributes X2, X3, and X5 as the most important attributes for that area, he could choose one of these three as consequent of a *context rule*, while the two other attributes for the antecedent were chosen from the complete set of attributes of the *design area (*X1, X2, X3, X4, X5 and X6). We instructed the teacher not to use in the antecedent the same attribute that he already placed in the consequent. These rules were 3-itemset (two items in the antecedent and one item in the consequent). The antecedent → consequent context rule construction denoted which attribute placed in the antecedent appeared at the same time with an attribute place in the consequent based in the area consideration. Figure 3.8 presents the example of the context rules from *learning promotion of the course area*

	ANTECEDENT			CONSEQUENT
	ATTRIBUTE X1	ATTRIBUTE X2		ATTRIBUTE Y
RULE A	X15	X13	THEN	X9
RULE B	X11	X17		X12
RULE C	X14	X10		X18
RULE D	X10	X11		X19

Figure 3.8 Example of context rules creation

The teacher could try different combinations of attributes in the antecedent side of the rule for the consequent attribute selected. Next, the teacher chose only one of the two attributes in the antecedent. We used only rules with one attribute in the antecedent and one attribute in the consequent. This step was repeated for the three areas. The reason for eliminating one of the attributes in the antecedent side was to encourage the teacher to weight implicitly the relative importance of the attributes that were in the antecedent and chose only one. In summary, during this step, the teacher weighted the attributes and identified which were more important per area, which ones were more interesting for him as results (consequent), and which ones were more interesting/useful to appear in the antecedent of the same rule. Figure 3.9 shows the final attributes in the antecedents with its respect consequent.

	ANTECEDENT		CONSEQUENT
	ATTRIBUTE X		ATTRIBUTE Y
RULE A	X13	THEN	X9
RULE B	X17		X12
RULE C	X10		X18
RULE D	X11		X19

Figure 3.9 Example final attributes (context rules)

In Figure 3.9 teacher eliminated attributes X15, X11, X14 and X10 respectively. This teacher considered attributes X15 and X10 as less important attributes in open rules and in context rules as can be seen in Figures 3.6, 3.7, 3.8 and 3.9

3.2.6 Section E: Questionnaire for Teacher's evaluation rules (S2)

To construct survey S2, we used rules with higher support, confidence and lift values. We obtained these rules from the linear regression analysis applied to students' dataset and the Apriori algorithm. (We explained the linear regression application in Chapter 4). Then, the teacher analysed the association rules obtained from the regression attributes using knowledge categories. For this purpose, the teacher classified these association rules using six knowledge categories: "interesting", "unexpected", "useful", "obvious", "previous" and "irrelevant". We presented an example of the association rules evaluated in Figure 3.10**.**

	RULES EVALUATED AS <u>EXCELLENT(ALWAYS), GOOD(SOMETIMES) OR REGULAR(RARELY)</u> (ANTECEDENT AND CONSEQUENT SIDE)	INTERESTING	UNEXPECT. knowledge LS	RS	BS	USEFUL knowledge	OBVIOUS knowledge	PREVIOUS knowledge	IRRELEVANT knowledge
1	IF X15 Professor **ALWAYS** maintains fluid communication with students **THEN** Professor is evaluated as she/he… **X12 ALWAYS** motivates learning of the course material	☑	☐	☐	☐	☐	☐	☐	☐
2	IF X19 Professor **ALWAYS** develops class content in an understandable way **THEN** Professor is evaluated as she/he… **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☑	☐	☐	☐
3	IF X15 Professor **ALWAYS** maintains fluid communication with students **AND** X19 Professor **ALWAYS** develops class content in an understandable way **THEN** Professor is evaluated as she/he… **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☑	☐	☐	☐
4	IF X22 Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he… **X12 ALWAYS** motivates learning of the course material	☐	☑	☐	☐	☐	☐	☐	☐

Figure 3.10 Example Section E - Knowledge categories classification

After the teacher categorized the association rules, we used the rules classified as *interesting, unexpected* and *useful* and obtained the frequency of the attributes in the antecedent and in the consequent. The frequency results from the attributes represented teacher's preferred attributes. Then, we chose the attributes with higher frequencies in the antecedents and in the consequents. These attributes were called *evaluated attributes*. The evaluation time for this instrument was 1.5 hours.

3.3 Results from the survey

In this section, we presented the results from the teachers' survey of the Sections A, B, C, D and E. We applied the Survey S1 and S2 to each of our 6 teachers participating in this research.

3.3.1 Results from Section A: Attribute evaluation

From Section A, we obtained the general attribute evaluation. Figure 3.11 presents the results where most of the attributes were evaluated above 8 with the exception of X1. The figure illustrates that the teachers considered most of the attribute very important.

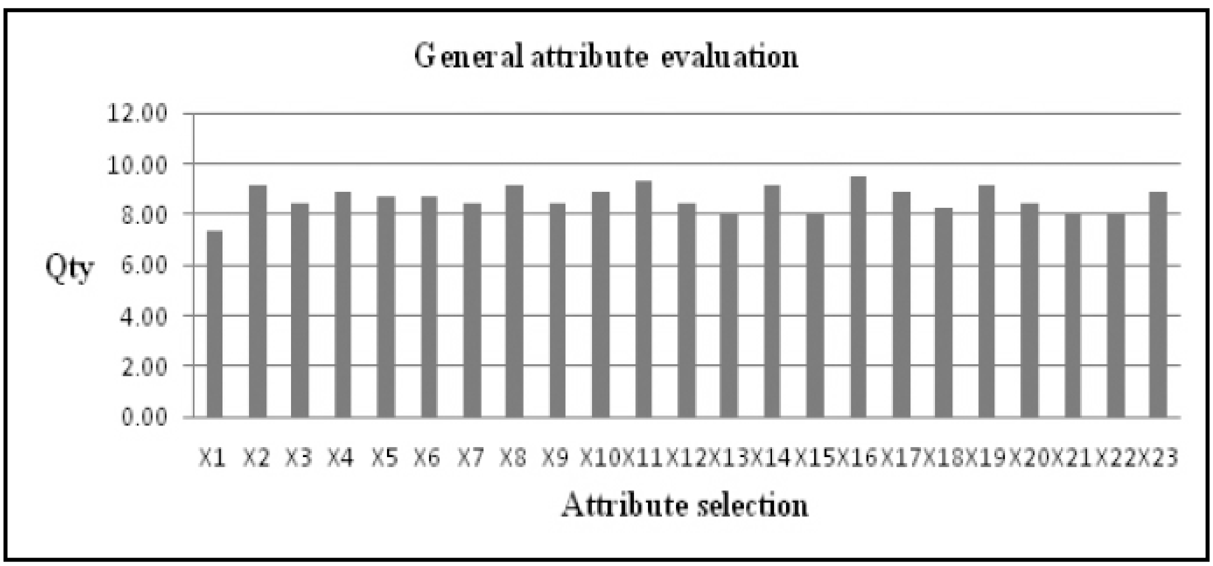

Figure 3.11 Results from Section A- Attribute evaluation

3.3.2 Results from section B: Most selected areas

In Figure 3.12, we show the teachers selected as the most important area the Learning Promotion area.

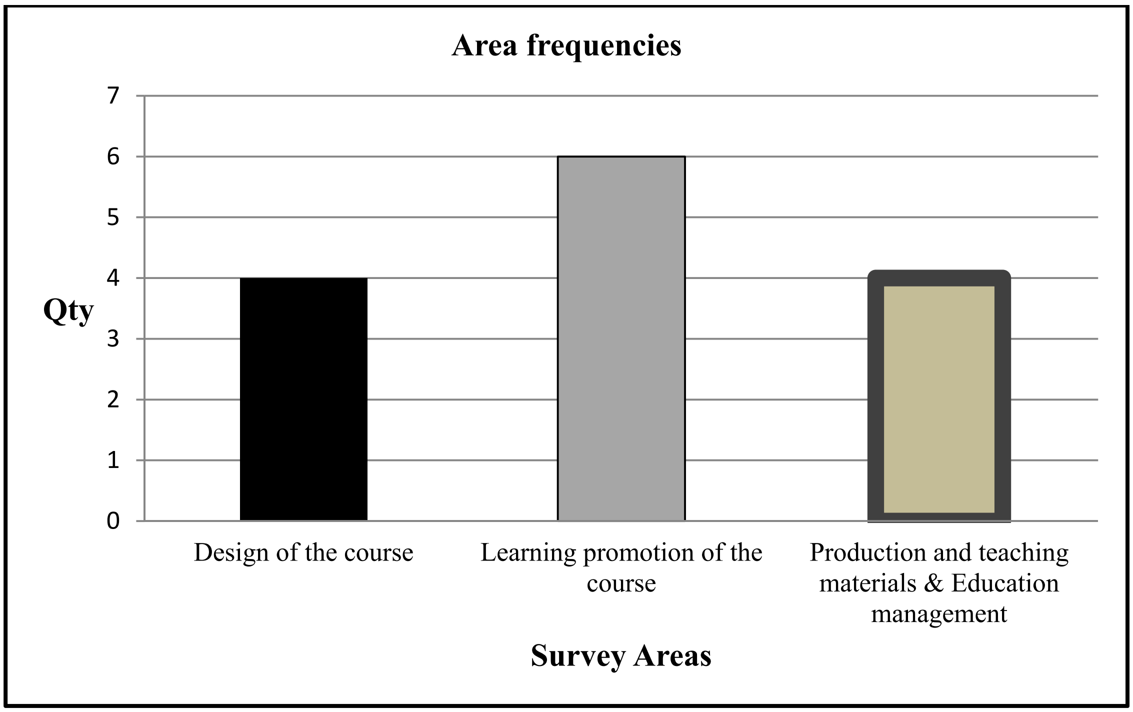

Figure 3.12 Results section B frequency per areas

In Figure 3.13, each attribute is evaluated inside of each area (namely Design of the course; Learning Promotion of the course; Production & Teaching materials and Education Management); Using Table 3.2 the teachers chose the three, six or two attributes correspondingly from each area. The attributes selected in Section B as the most important were different from those in section A, (See Figure 3.13).

We see that inside the Design area (X1-X6) the most frequently chosen attributes (four out of six) was X3; inside the Learning Promotion of the course (X7-X19) the most frequently chosen attribute were the X14 and X19 and in the last area, the most common question was X21.

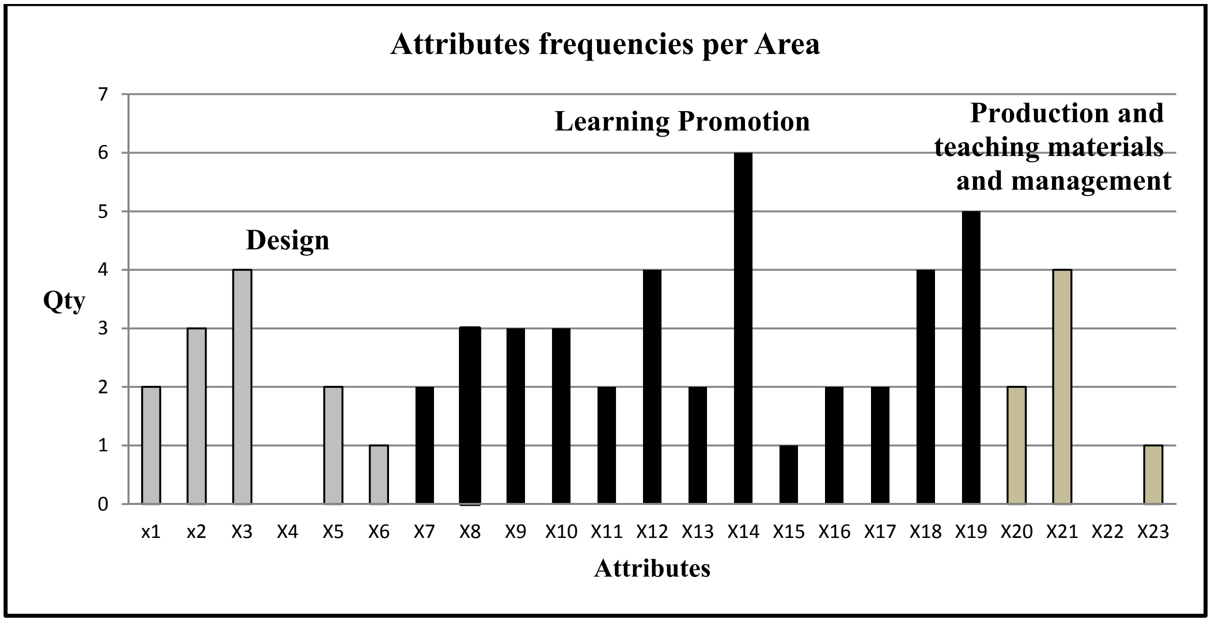

Figure 3.13 Results Section B-Selected attributes per area

None of the teachers chose attributes X4 and X22 when they were analysing attributes per area. The descriptions of the most frequently chosen attributes were:

X3: Evaluates students' participation in class

X14: Promotes reflection on topics covered

X19: Develops class content in an understandable way.

X21: Frequently uses schemes and graphics to support his or her explanations.

3.3.3 Results from section C: Reaction to low evaluation

This section shows the results from section C. In Figure 3.14, attribute X10, X16 received the evaluation of 10 ("I don't know how to use these results) and attribute X6, X11, X15 and X19 received the evaluation of 7 ("results are useful but not easy to change"). Half of the teachers considered these attributes difficult to interpret based only on the students' survey. The other attributes received an evaluation lower than 7 ("Maybe the results do not depend only on the teacher"). Attributes evaluated with this value, reflect that these low graded attributes are related to other aspects like students' family situations or realities (arriving late, arriving tired to class, etc.) that are out of the range of the teacher.

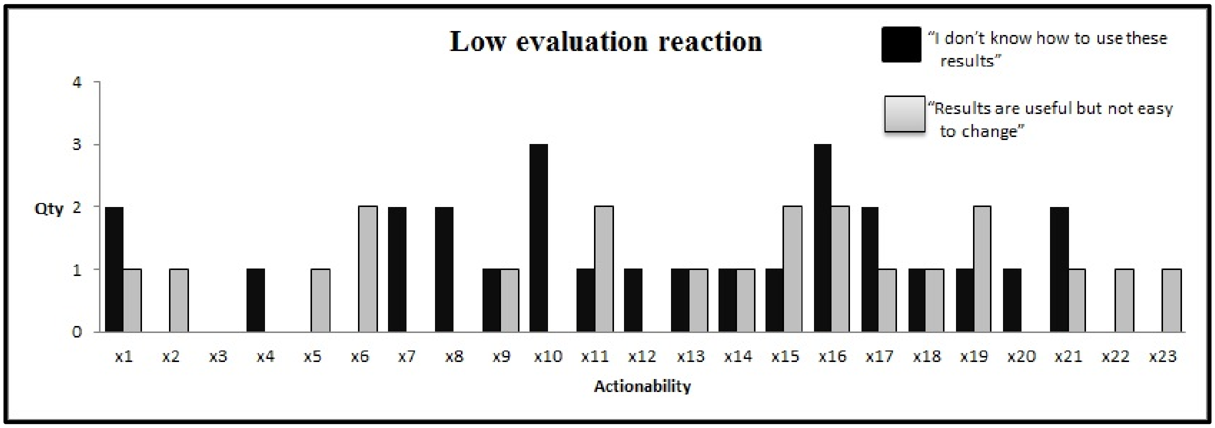

Figure 3.14 Results Section C - Low evaluation reaction

3.3.4 Results from section D_a: Creation of open association rules using attributes from section A

In section D_a, the teacher created open association rules using up to five of the attributes considered as the most interesting in section A. The teacher constantly placed the following attributes: X6, X10, X12, X14, X19 in the consequent, as can be seen in Figure 3.15**.**

The description of the attributes frequently placed in the consequent of the association rules were:

X6: Motivates students to do additional research

X10: Summarizes key ideas discussed before moving to a new unit or topic

X12: Motivates learning of the course material

X14: Promotes reflection on topics covered

X19: Develops class content in an understandable way

The teacher never used attributes X7, X15, X16, X20 and X21 in the consequent for open association rules.

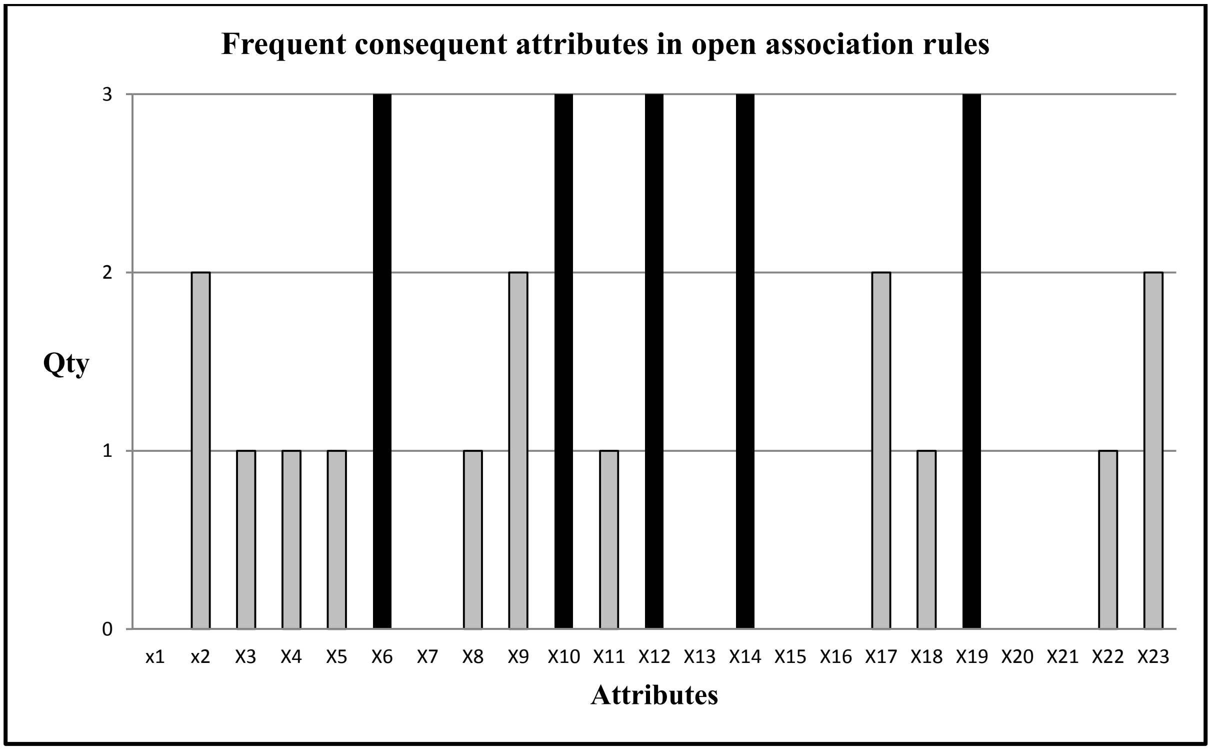

Figure 3.15 Frequent consequent attributes use in open association rules

The attributes frequently placed in the antecedent to construct open association rules were X9, X11, X13, X16, and X21 as can be seen in Figure 3.16. The descriptions of these attributes were:

X9: Encourages active student participation in class
X11: Establishes relationships between new concepts and those already known whenever possible.
X13: Is willing to answer attributes and offer advice within and outside of the classroom
X16: Is respectful towards students

X21: Frequently uses schemes and graphics to support his or her explanations

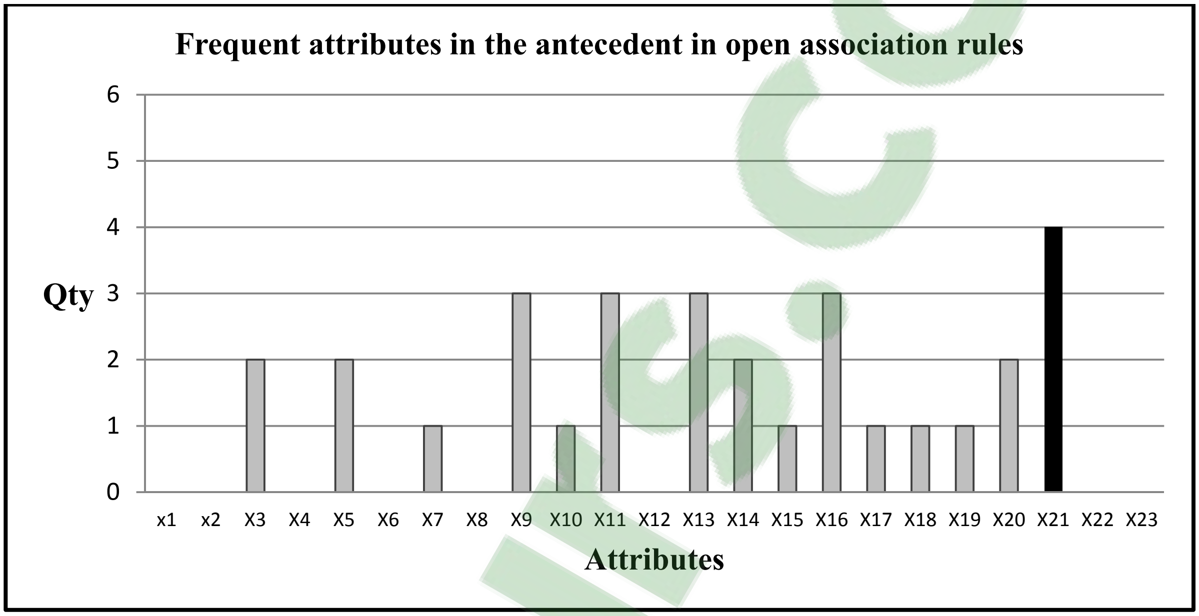

Figure 3.16 Frequent antecedent attribute use in open association rules

The attributes never used in the antecedent were X1, X2, X4, X6, X8, X12, X22 and X23

As previously stated, the teacher eliminated one attribute from the antecedent in the association rule. The most frequently eliminated attributes from the antecedent were X13 and X21 as illustrated in Figure 3.17**.** The descriptions of the commonly eliminated attributes were:

X13: Is willing to answer attributes and offer advice within and outside of the classroom.
X21: Frequently uses schemes and graphics to support his or her explanations.

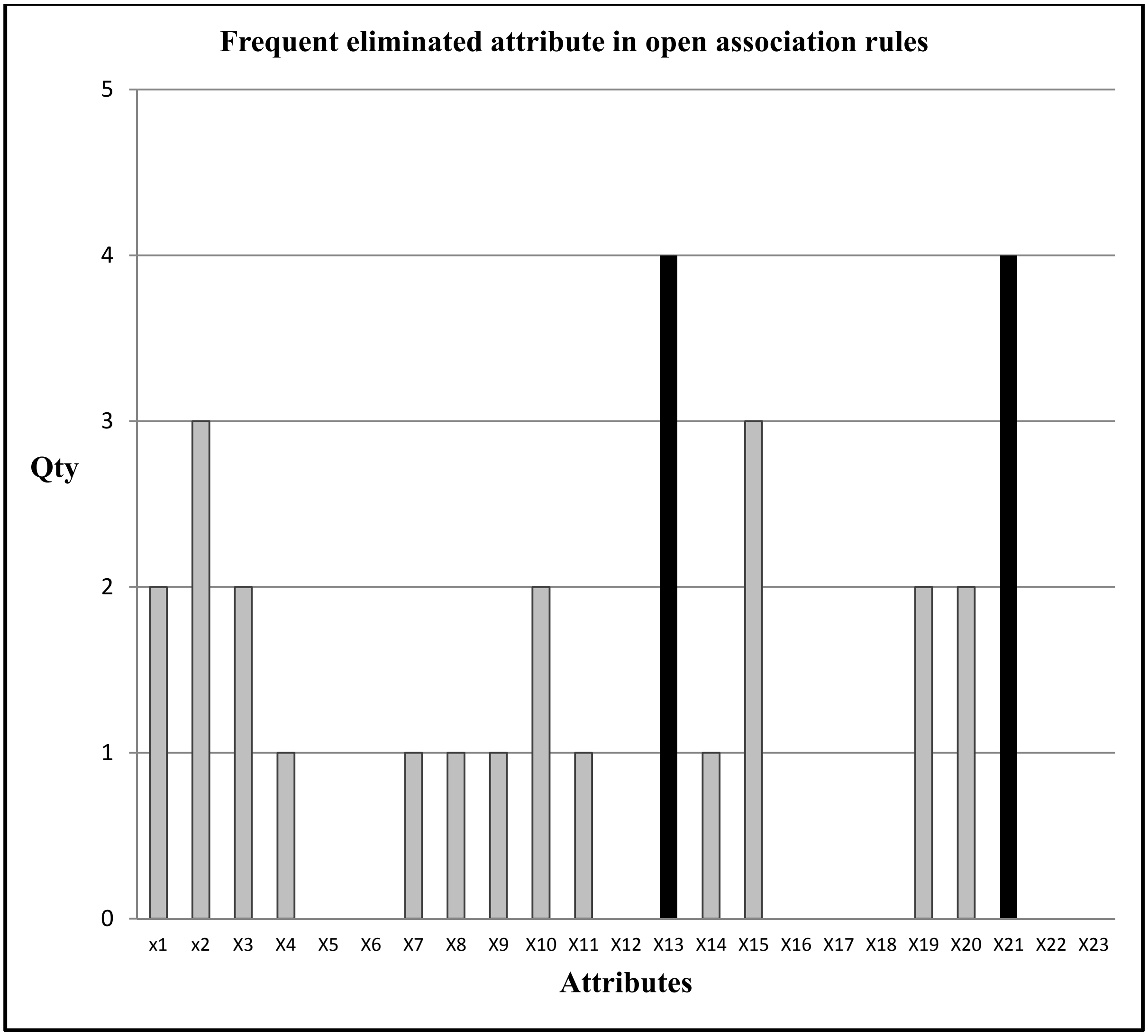

Figure 3.17 Frequent eliminated attribute from the antecedent of open association rules

Finally, we analysed open teachers' rules where only one rule appeared constantly.
This rule was X16 then X23:

X16	→	X23
Is respectful towards students	→	Attends classes on time

In other words, attending on time to class was an expression of Teacher's respect to students.

3.3.5 Results from section D_b: Creation of context association rules using variables from section B

In these results, we present three types of graphs related to the three groups of attributes for each area: the most frequent attribute in the consequent, the most frequent attribute in the antecedent and the most frequent attribute eliminated from the antecedent. The three areas were Design, related to attributes X1 to X6; Learning promotion related to attributes from X7 to X19, and finally Production and Teaching materials & Education management related to attributes from X20 to X23.

Bar charts for the design area, Figure 3.18 shows teachers constantly placed in the consequent attribute X5. The description of the attribute X5 is:

X5: Provides clear instructions for learning assessment (test, quizzes, presentations, role playing, etc.)

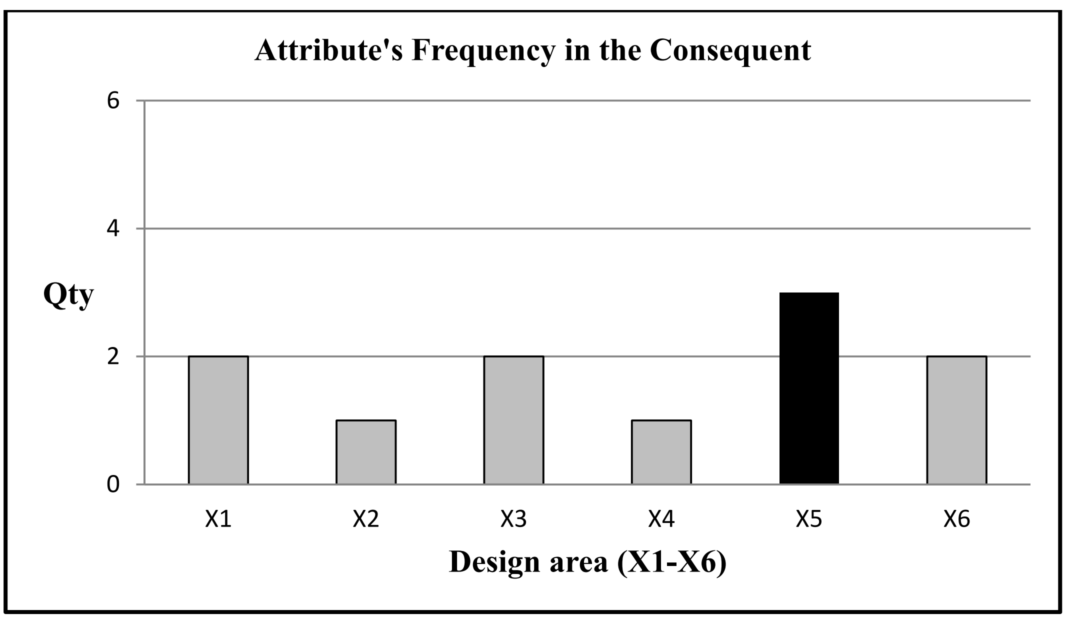

Figure 3.18 Most frequent attribute placed in the Consequent (Design)

The teacher frequently placed attributes X3 and X4 in the antecedent to construct context association rules, as illustrated in Figure 3.19. The descriptions of these attributes were:

X3: Evaluates students' participation periodically in class
X4: Evaluation fits the themes developed in class

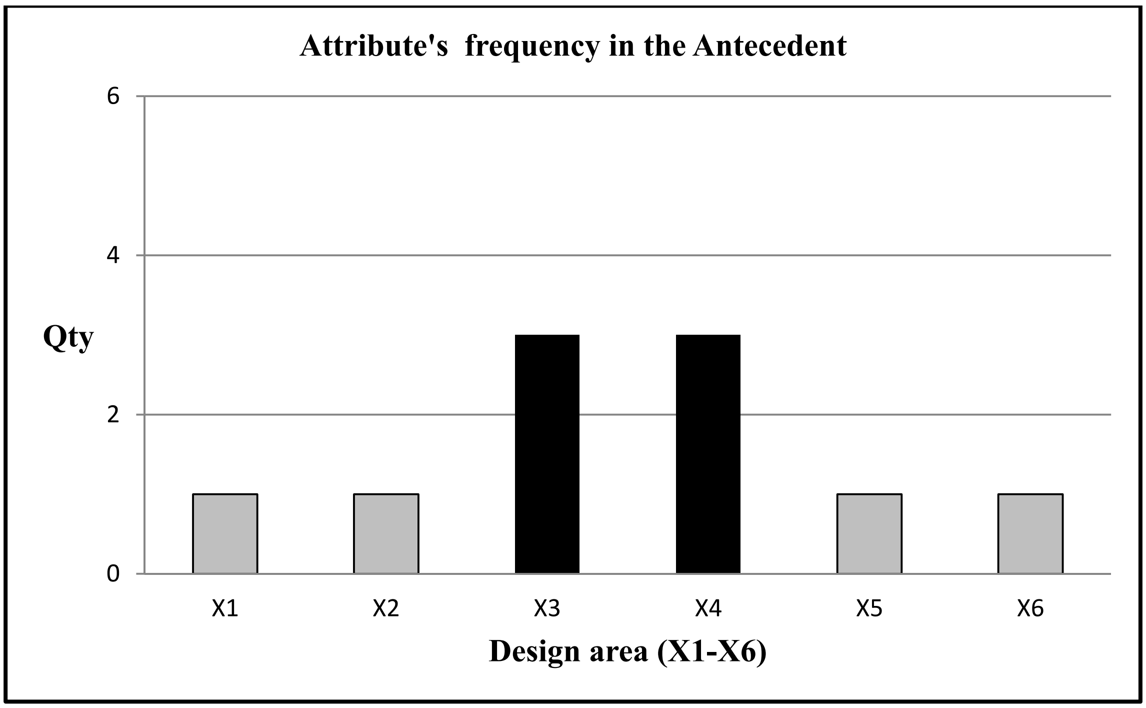

Figure 3.19 Most frequent attributes placed in the antecedent (Design)

The teacher eliminated one attribute from the antecedent in the association rule. The most frequently eliminated attribute from the antecedent was X2, as illustrated in Figure 3.20. The description of this attribute was: X2: Fulfills the program proposed at the beginning of the course

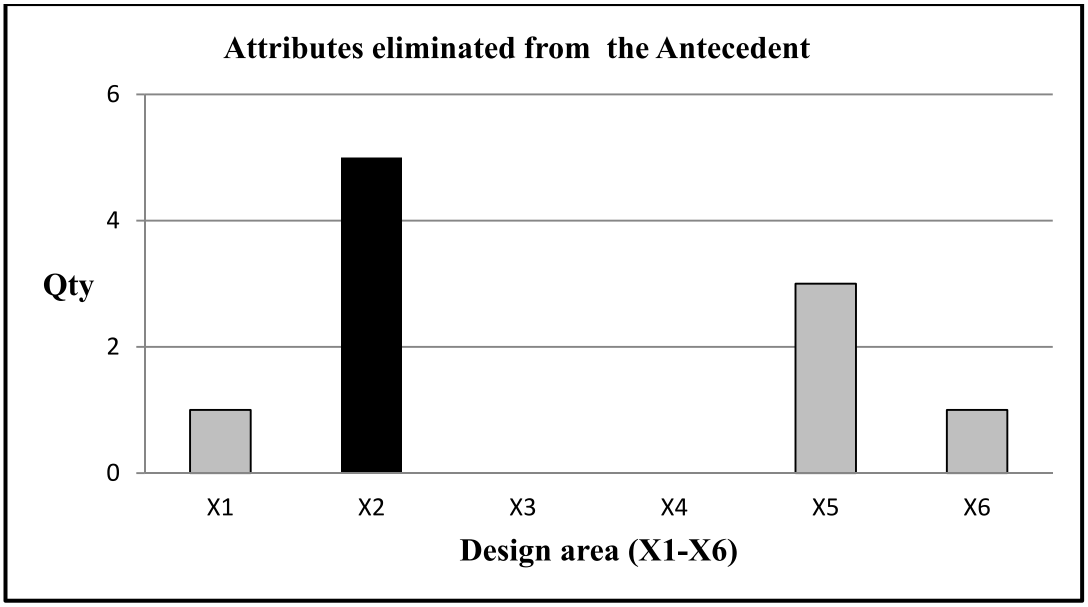

Figure 3.20 Most frequent attributes eliminated from antecedent (Design)

In the Learning Promotion area, the teacher constantly placed attribute X12 in the consequent as shown in Figure 3.21. The description of this attribute was:
X12: Motivates learning of the course material

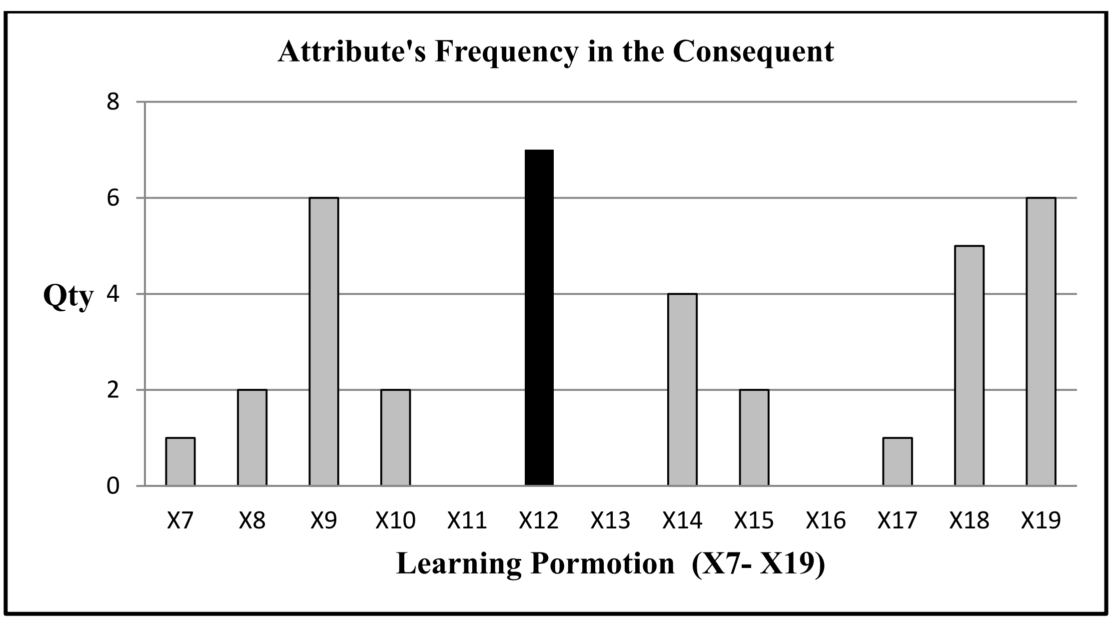

Figure 3.21 Most frequent attributes placed in the consequent (Learning Promotion)

Teachers never used attributes X11, X13, and X16 in the consequent.

The attribute most frequently placed in the antecedent to construct context association rules was X10 as illustrated in Figure 3.22. The description of X10 was: "Summarizes key ideas discussed before moving to a new unit or topic". Teachers never used attributes X8 and X19 in the antecedent.

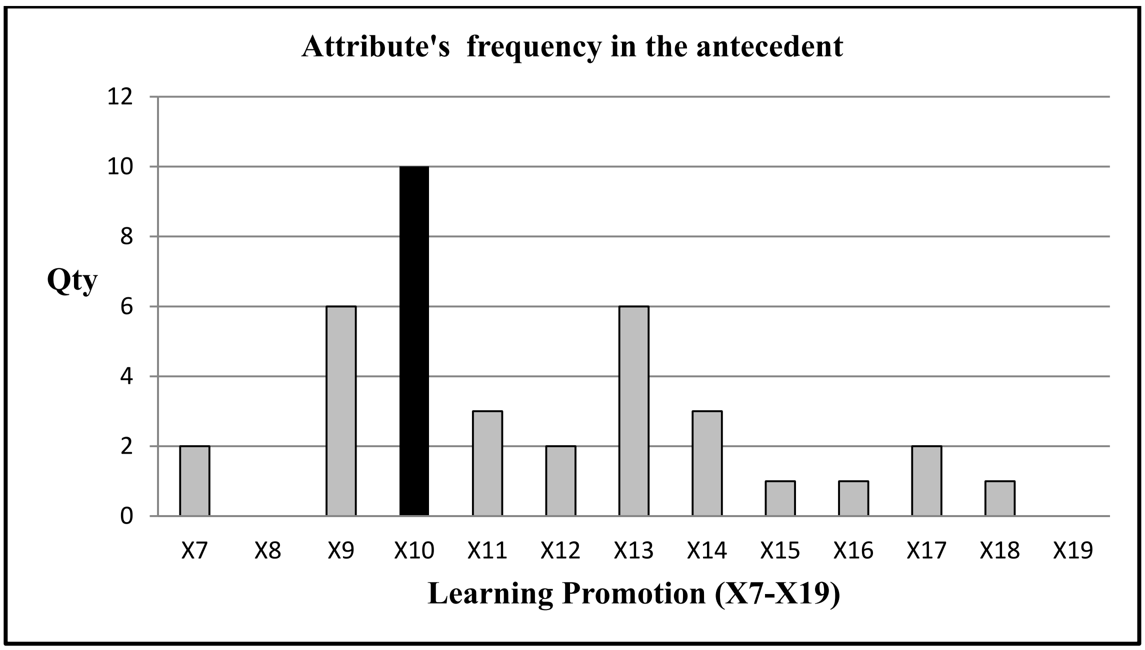

Figure 3.22 Most frequent attributes placed in the antecedent (Learning Promotion)

Teachers eliminated one attribute form the antecedent in the association rule. The attributes most frequently eliminated from the antecedent side of the rule were X8, X10, X14, X15, and X17 as can be seen in Figure 3.23. The descriptions of these attributes were:

X8: Explains class policies at the beginning of the course
X10: Summarizes key ideas discussed before moving to a new unit or topic
X14: Promotes reflection on topics covered
X15: Maintains fluid communication with students
X17: Responds to attributes in class about subjects related to the field

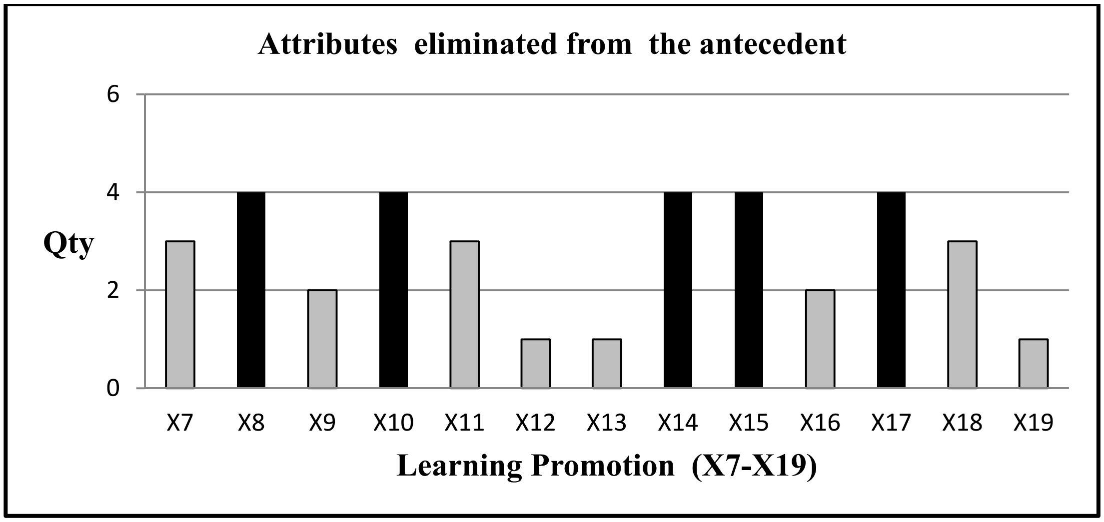

Figure 3.23 Most frequent attributes eliminated (Learning Promotion)

For Production and teaching materials & Education management, teacher constantly place in the consequent the attributes X20 as can be seen in Figure 3.24. The description of this attribute was: X20: Prepares instructional bibliographic or other resources to facilitate learning. Teachers never used attribute X23 in the consequent.

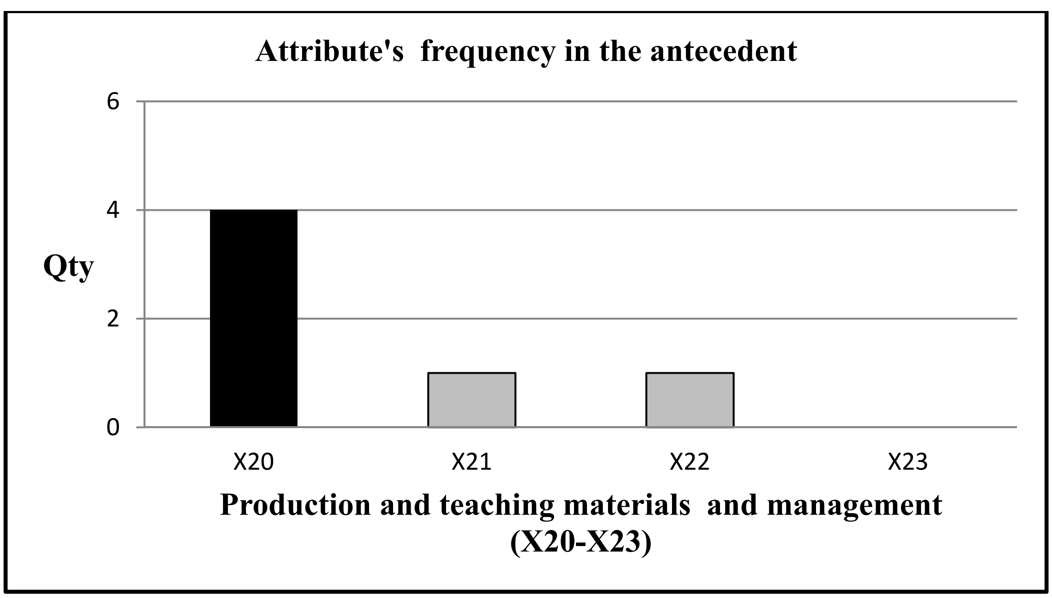

Figure 3.24 Most frequent attributes in the consequent (Production-Management)

The most frequent attribute placed in the antecedent side of the rule to construct context association rules was X21 as can be seen in Figure 3.25; where X21: Frequently uses schemes and graphics to support his or her explanations

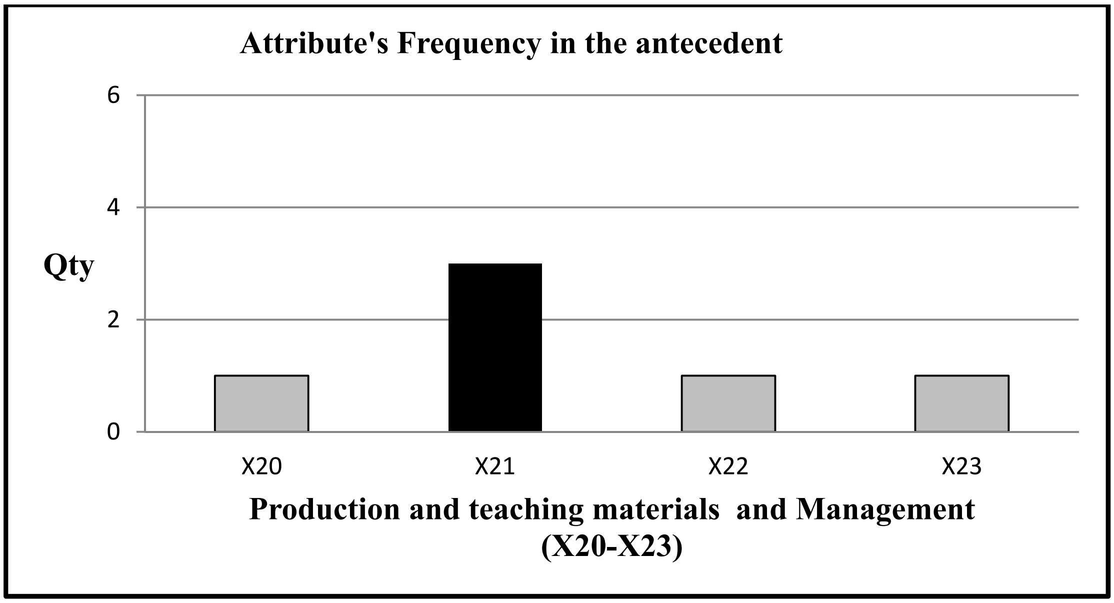

Figure 3.25 Most frequent attributes in the antecedent (Production-Management)

As you remember, teacher had to eliminate one attribute from the antecedent in the association rule. The most frequent attribute eliminated from the antecedent side of the rule, in *"Production and management area*" was X21 as can be seen in Figure 3.26**;** where X21 description is "Frequently uses schemes and graphics to support his or her explanations"

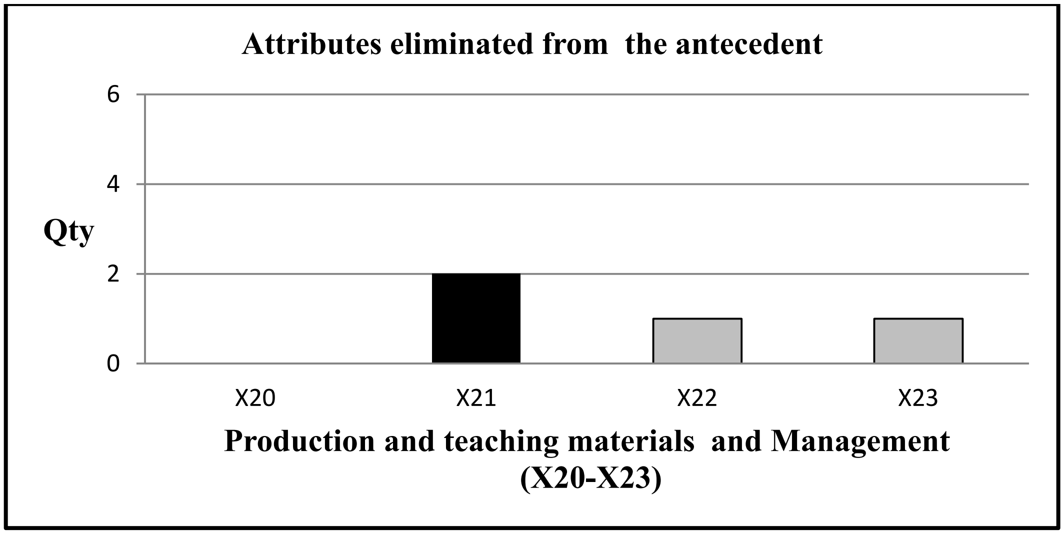

Figure 3.26 Most frequent attributes eliminated (Production-Management)

Finally, we obtained the frequency of the rules teachers created per area. Figure 3.27 shows the frequencies of each rule. This figure shows the rule (i.e. 10T19) and the rule's quantity constructed in context. The rule was expressed as 10T19, meaning that in the antecedent was the attribute X10 and in the consequent was the attribute X19. "T" between the attributes represents the "then" expression.

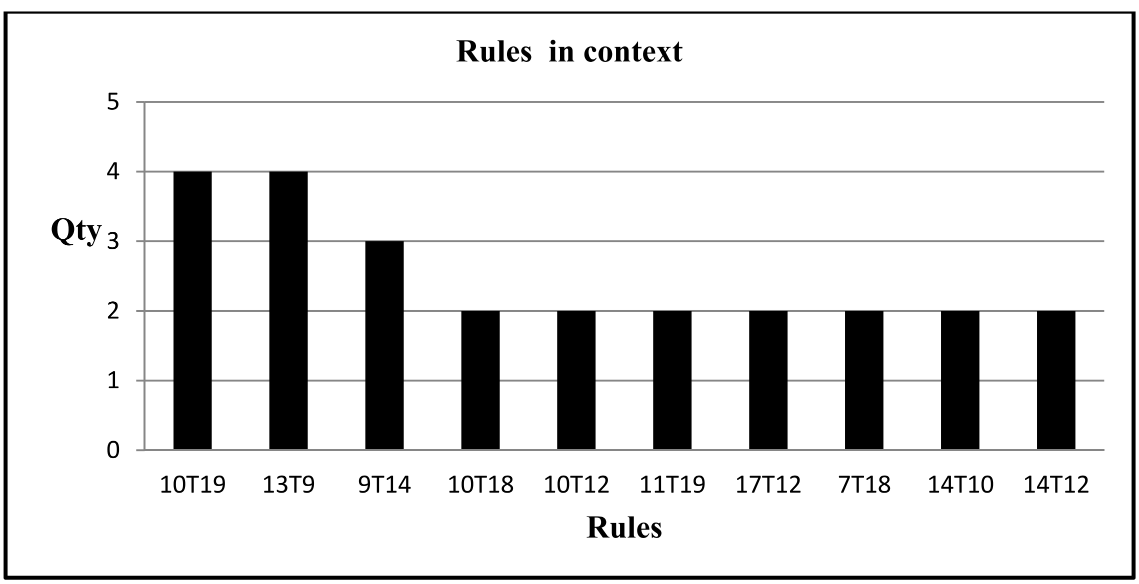

Figure 3.27 Frequency of the rules in context association rules

The descriptions of the three more common created rules in context are as follow:

X10	→	X19
Summarize key ideas discussed before moving to a new unit or topic	→	Develops class content in an understandable way
X13	→	X9
The teacher is willing to answer attributes and offer advice within and outside of the classroom	→	Encourage active students' participation in class.
X9	→	X14
Encourage active students' participation in class	→	Promotes reflection on topics covered

3.3.6 Conclusions from teachers' survey

From the survey we obtained Teachers' preferences about the attributes, here we present some conclusions obtained from the data analysed.

- In the general evaluation of the attributes, section A, the teacher selected a group of important attributes. However, in the evaluation per context areas, section B, the selection of important attributes changed. In other words, the results from section A where the teacher evaluated all the 23 attributes were different from the results from section B where the teacher chose favorite attributes in three areas (context). Attributes that were very important in a global way are less important within the three areas.

- There were no universal rules valid to all teachers, but within each context area, teachers tended to favor few key rules. When the teacher constructed rules freely, using the section A with high evaluated attributes, the rules generated by different teachers didn't agree, only one rule was similar among all teachers. This was not the case when we asked the teachers to construct rules using only the attributes inside each area. Figure 3.26 shows the rules and the number of teachers that created the same rule inside a given context.

- In section C, when we evaluated the reaction to low evaluation from teachers, we identified that teacher was not comfortable with the understanding of attributes X9, X10, X11, X14, X15, X17. In other hand, attributes X6, X16 and X19 appear as useful information not easy to change. In general these two categories presented incomprehension of the results and doubts about how to proceed in order to improve these attributes.

- Referring to the selected attributes to construct open rules and the selected attributes to construct context rules, we observed that a group of attributes appeared in both rules construction, in other words, these attributes were choosing to be in an association rule, independently if they were selected for open association rules or context association rules in section A or B. This is the case of attributes X9, X10, X11, X12, X13, X14 and X19. Besides the already mentioned attributes, teachers chose attributes X6, X16 and X21 frequently; when he created open association rules, and X2, X3, X4, X5 and X8 when he created context association rules.

- The most relevant areas that were more important for teachers were Design and Learning Promotion of the course. This was very noticeably when they freely chose attributes to construct rules. In Figure 3.19 and 3.26 we observe that the teachers constructed few rules using the attributes below X7 or above X19.

3.4 Teacher’s interview

Interviews have been used in the past to create knowledge in research (Kvale and Brinkmann 2009). Qualitative interviewing is a valuable instrument to capture consumer experiences, teachers’ experience or students’ experiences. In fact, qualitative interviewing is also a very powerful technique to understand and predict the behaviour of the interviewee subject. For these reasons, the teachers were questioned on the basis of replies to the three previous instruments. Each teacher expressed his opinions about the attributes, the evaluation results and the rules that are possible patterns. The duration of the interview was approximately one hour.

3.4.1 The interviewing model

Kvale and Brinkmann mentioned seven steps to construct an interview. These seven step were taken into account by Rubin and Rubin (Rubin and Rubin 2012) with a modification in the order of these steps with the possibility of returning to previous steps to improve the interview. We followed the model from Kvale and Brinkmann in the exact order. This model contained seven steps from designing the study to the report of the results.

We applied one to one interview because we were looking for personal opinions without the intervention of additional persons. We wanted the interviewees to feel free to speak without the pressure of someone else giving opinions or giving judgmental comments. Creswell suggest **the design and use of an interview protocol** for the interview task (Creswell 2012). We used an interview protocol with 14 questions. The number of questions was related to the attributes we intended to understand. First, we defined a **pilot testing** and applied it to three teachers. We didn’t consider this pilot as part of our case study because it was the first time we were using the interview and we needed to explain to the pilot interviewees how it worked, make sure the instructions were good and comprehensible, and provide enough clues to the people who have to answer the questions. The observations from the pilot improved the final version of the interview.

In our research, we selected six teachers to participate in the study. Each of our participants had to sign **a consent form for human relation as a requisite for the Ethical committee from ETS** (Review form, purpose of the study, amount of time, plans of using the results), and we kept a copy of these documents. Our research obtained an ethical certification from the Ethical committee from ETS; the certification is presented in the appendix 4. The **interviews started on time** and **ended on time.** Our participants were treated **in the best possible way.**

Creswell presents four types of participant formats for the interviewer. We, as interviewers, are placed in the non-participant/observer, because we started not knowing what was going on inside teacher class activities. Then, after some time in the interview, we carefully looked for commonalities among the teachers' answers, or similar situations experienced in the classroom, which could help the interviewee expands his ideas.

3.4.2 Selection of the interviewees

We were interested in selecting teachers with more than 5 years of experience teaching in the university, specifically for the Computing and Electrical Faculty where the study took place. We looked for teachers with high evaluation scores to identify the good teaching practices that assured high evaluations. From the group of faculty, we contacted, we chose six of them to cover technical and theoretical courses. The six participants teach at the university in the same school and continued to get high evaluations. They teach computing courses of four hours per week and two of them have lab time included in their classes. Each participant was very cooperative during the interview and provided valuable perspectives on the questions and on the association rules from the objective perspective.

3.4.3 Sampling strategy

Our sampling strategy was based on the **homogeneous strategy** because we were interested in the difference between teachers from Computing and Electrical Faculty. We selected for our illustrative case study six different courses, some with technical and some with theoretical emphasis

3.4.4 Size of the sample

In our specific case, we constructed a case study to generalize the information (Creswell 2012) but we wanted to emphasize on the specificities of each of our participants. In this sense, we aimed to find a lot of similarities among teachers' opinions and examples. However, we also focused on finding differences related to the type of the course teachers were developing with special interest in the tools, activities, even jokes and methods that teachers used to draw the students attention.
Creswell found that a common number of case studies interviewees would be four or five; we covered technical and theoretical courses as Programming Fundamentals, Research methods applied to computing, Software engineering I, Entrepreneurship, Web application development and Digital communications.

3.4.5 Interviews using Skype

We did six interviews. One interview was face to face, while for the rest we used Skype software. We did three of the interviews during working hours. In contrast, we did the other three interviews from our home at night.

The interviewer had previously met all the interviewees. Therefore, there was good communication, trust and familiarity to assure a friendly environment that would promote/ foster the interviewees' confidence. All the interviewees finished their interviews successfully. In other words, no one withdrew from it. The teachers preferred to have audio interview simply talking rather than video ones. Four of the interviewees preferred not to see the interviewer. We let the interviewee set the ambience, if he turned on the video, we understood that he was comfortable with it and then we did the same with our video. If they didn't turn video on, we asked them if they felt comfortable with it and work with audio only. Our interviews were straightforward and accomplished the desired objectives. They were planned ahead of time; we considered time difference between the participants and the interviewer and offered alternative times for scheduling their appointment(Deakin and Wakefield 2013) . At the moment of the interview, all the participants had access to internet without any technical problems.

The application of these three tools to the six teachers provided us with information related to teacher preferred attributes in a general and contextual way. Additionally, the teacher created his own rules and expressed his beliefs in the form of association rules. Finally, the teacher talked about his preferences and constructed rules.

CHAPTER 4

ARTICLE I : STUDENTS' AND TEACHERS' PERCEPTION OF ACADEMIC EVALUATION : METHODOLOGY TO CONSTRUCT A REPRESENTATION BASED ON ACTIONABLE KNOWLEDGE DISCOVERY (AKD) FRAMEWORK

Otilia María ALEJANDRO MOLINA
Software Engineering and IT, École de Technologie Supérieure (ÉTS),
1100, rue Notre-Dame Ouest, Montreal-Canada. H3C 1K3

Computing Faculty, Escuela Superior Politécnica del Litoral (ESPOL),
Km 30.5 Via Perimetral. Guayaquil-Ecuador

Sylvie Ratté
Software Engineering and IT, École de Technologie Supérieure (ÉTS),
1100, rue Notre-Dame Ouest, Montreal-Canada. H3C 1K3

Paper published on February 21, 2016 at Springer Information and Education Technology
DOI: 10.1007/s10639-016-9471-3

EAIT-D-15-00082R1

Abstract: This research introduces a method to construct a unified representation of teachers and students' perspectives based on the actionable knowledge discovery (AKD) and delivery framework. The representation is constructed using two models: one obtained from student evaluations and the other obtained from teachers' reflections about their teaching practice. We integrate both models into one that incorporates students' opinions and teachers' knowledge and meta-knowledge. This method provides a representation of a teacher's best teaching practices where student perceptions are presented as patterns in the form of association rules. The representation adds actionability to association rules by demonstrating how students' association rules are related between themselves and how they are related to teacher's meta-knowledge.

Keywords: AKD, Teacher's evaluation; Students perspectives; Teacher perspective; Objective measures; Subjective measures; Knowledge and meta-knowledge

4.1 Introduction

Teaching skills are commonly assessed in universities using evaluation questionnaires. Most of the time, the results obtained from these questionnaires identify solely the student’s perspective of the highs and lows in specific teaching areas. As we can guess, the students' point of view frequently differs from the teacher's view of their own work and consequently, the evaluation results may come as a surprise. Teachers may have a very poor understanding of the shortcomings perceived by students. Due to lack of precise information, they may end up applying considerable effort to improve non-problematic aspects of their teaching style without correcting the real source of the teaching problem.

The interpretation is therefore not straightforward and cannot be based solely on data obtained from surveys filled out by students; the analysis should rely on various levels of knowledge that encompass a broader perspective. For such complex problems that cannot be solved using data alone, the AKD framework proposes combining technical and business measurements, domain knowledge and meta-knowledge as part of the solution (Longbing 2010).

AKD is part of the new generation of Domain Driven Data mining (D3M) frameworks. The main objective of AKD frameworks is to discover actionable patterns that bring immediate application focusing on domain knowledge (increase in profits, better efficiency, etc.)(Longbing 2008); A pattern actionable capacity (Longbing 2007, Yang 2009) must be the result of a good balance between technical and business interests. AKD uses technical and business interestingness measures that help identify the interest of a pattern from the objective and subjective point of view. In other words, each pattern is analysed both objectively, using data, and subjectively, taking into considerations the opinions of the expert who work with the data.

In this work, we evaluate and analyze both students’ and teachers’ perspectives using the AKD framework from D3M. The students’ perspective is evaluated using association rules with objective measures, while the teacher's perspective is evaluated using subjective measures, domain knowledge and meta-knowledge. We then analyze these results to integrate them into a single representation model.

The end result is a global model representing the best teaching practices in the form of an interesting set of association rules. This representation adds actionability to association rules by including the real actions taken by teachers when they applied these rules. The model helps one understand how patterns and actions are related as well as showing what meta-knowledge is related to each rule.

This paper is organized as follows: the next section presents DK, MK, measurement concepts and some frameworks that have been proposed for AKD; **section 4.3** presents the dataset and the questionnaire structure used; the methodology is presented in **section 4.4** and its application to a particular case is detailed in **section 4.5**; **sections 4.6** and **4.7** present conclusions and future work.

4.2 Literature review

One of the characteristics of AKD is to combine DK and MK. In the first section of our literature review, we will circumscribe the definition of these concepts and relate them to our specific context. The second section defines the various measures that can be found under the objective and subjective labels. Finally, we present the mainstream AKD frameworks and delimit the one that we are using.

4.2.1 Knowledge, domain knowledge and meta-knowledge

Knowledge is a mix of constructed experiences, values, information or contextual data and expert insight that belongs to any enterprise in an implicit way and that transforms itself in assets for the organization over time (Davenport T. H 1999). These assets belong on the one hand, to people who manage the knowledge and learn in everyday experiences, and on the other hand, to the organization, which saves it in different forms including documents, processes and practices. People and organizations struggle to recognize and organize knowledge because of the large number and variety of business situations and solutions.

Domain knowledge (DK) is related to the experiences, values, and user insights implicit in user knowledge (Davenport T. H 1999). A simple definition about domain knowledge was given as “the knowledge of the subject area (domain), what you know about a subject or topic” (Paquette Gilbert 2011). DK is therefore related to the user experiences in a specific domain. For example, domain knowledge for a teacher might include academic problems that students face during the discussion of a specific topic (including insecurities and fears), feedback about the difficulties with a topic and specific abilities that the teacher has to apply to reduce fears in order to reach academic objectives.

In any industry, *domain knowledge experts* are individuals who have reached high levels of expertise in a particular domain. They specialize in specific problem resolution. They gain expertise doing similar tasks or resolving the same problem in different contexts, and store and apply these rules of thumb depending on the situation they need to solve.

To elicit domain knowledge from an expert, it is necessary to use techniques to retrieve the tacit knowledge and transform it into explicit knowledge in the form of rules (Flavell 1979, Yi-Dong, Zhong et al. 2002). This knowledge can then be standardized and applied in future similar situations.

Meta-knowledge is knowledge about knowledge. It can have different levels. One example of meta-knowledge is meta-cognition–what we know about our cognitive abilities or how we learn. Some authors mention that meta-knowledge and meta-cognition are related (Herrmann, Kienle et al. 2003), while others say that they form a part of each other (Valot and Amalberti 1992). In this paper, meta-knowledge will refer to both meta-knowledge and meta-cognition; we will focus on a teacher’s knowledge, cognitive tools, abilities, limitations and the use of strategies (Valot and Amalberti 1992).

Meta-knowledge (MK) is knowledge about the knowledge, i.e. the range of what has been learned, and mostly concerns the context of an activity(Valot and Amalberti 1992). It identifies persons and variables that intervene in the activity as well as strategies, steps and actions needed to accomplish the activity. If we consider the academic context, MK might include how students learn, how a specific topic is taught, what is easiest and what is hardest for students, etc.

To identify meta-knowledge, three questions have to be answered (Flavell 1979, Yang, Yu et al. 2009): *why*, *what* and *how*. The *why* identifies the reasons associated with performing a specific activity; the *what* defines the objective to be reached in the performing an activity; finally, the *how* refers to the strategy or steps required to accomplish the activity. Let's look at an example to show how to apply these questions in relation to the academic evaluation. If the activity is "*to make a student evaluate a teacher fairly by filing out a questionnaire*" to obtain the relevant meta-knowledge, we will ask the following questions: *why* is this evaluation done?; *what* concept do people interested in the results of this questionnaire learn?; *how* are learned concepts going to be used. In this process of identifying meta-knowledge, teachers have to talk about why they make students do specific activities, what activities and concepts they do in class and how the teacher tries to have those activities done in the best way. These activities are related to each of the survey questions. In this paper, we will call the survey questions *attributes* to help standardize the terminology.

We add the questions *when* and *where* to this traditional approach because they are necessary in the context of our research. The *when* question will focus on the time a specific teaching technique is applied; for example, in the case of students having to summarize key ideas, we need to ask if they must do that at the beginning or at the end of the class. The *where* question will be used if the class takes place in different locations such as labs, classrooms, outdoors, etc.

In this paper, teachers are considered "domain knowledge experts". The domain knowledge will be related to what teachers specifically know about the activities they perform during the class. For the meta–knowledge component, we will focus on the five questions mentioned above to extract the teacher's personal variables, the activities needed to be performed to reach an objective in class, and the strategies they apply to reach these objectives. The result will be a model aimed at expressing a teacher's best teaching practices in the form of an improved set of association rules. The resulting representation adds actionability to association rules and helps understand how the association rules are inter-related.

4.2.2 Technical and business interestingness measures

Technical experts in data mining (DM), are in charge of improving methods, algorithms and creating new measures to find and evaluate interestingness patterns. Business experts are interested in these technical expert improvement methods because these patterns should reveal new knowledge from within the business data. However, the patterns discovered by technical experts don't always fulfill the expectations of business experts mainly because they are not immediately applicable, they are not interesting to business experts or they don't matter to the business.

Table 4.1 shows a comparison between technical patterns and business patterns. Each of them shows a specific focus and none of them show relation between them. One focuses in the objective side and the other focus completely to the subjective side. That is why authors mentioned that a gap exists between the technical and the business interestingness patterns (Longbing 2008), (Longbing 2007).

Both patterns are considered relevant and complementary but they don't mix. Therefore, it is necessary to find a middle point; AKD suggest to integrate technical and business interestingness measures in its framework in the way that a pattern could be evaluated objectively with the technical considerations and evaluated subjectively with the expert considerations.

Table 4-1 Differences between technical interestingness and business interestingness patterns

	Technical Interestingness Patterns	**Business Interestingness Patterns**
Measures used to	Objective or statistical measures	Subjective, semantics, business measures

evaluate patterns		
Understood by	Technical people	Business people
Based on	Data	Business domain knowledge, user experience
Focused on	Efficient data mining techniques	Business concerns as profit, client satisfaction and improve business
Performed by	Academic world	Business world
Discover	Interesting patterns from data	Interesting and actionable patterns from the mix of data and the knowledge of the user
Driven by	Data driven	Domain driven
Satisfy	Satisfying expected technical significance	Satisfying business expectations
Aimed towards	Academic objectives	Business goals
Concerned with	Academia outputs	Business expectations

In AKD, two groups of measures are used for evaluating the interestingness of a pattern: *objective* and *subjective measures*.

Objective measures concern technical interestingness (Bing, Wynne et al. 2000)–rule structure, predictive performance, and statistical significance–of the data. Many objective measures have been proposed and some authors have studied their characteristics, properties (Tan, Kumar et al. 2002),(Geng and Hamilton 2007) and the suitability of a measure with respect to a certain domain(Xuan-Hiep, Guillet et al. 2006). Among these metrics we find: *support*, *confidence*, *lift* (or interest factor), *correlation*, *entropy*, *conviction*, *specificity*, *added value*, *Piatetsky–Shapiro*, *certainty factor* and others. Several authors (Yi-Dong, Zhong et al. 2002, Tan, Kumar et al. 2004, Geng and Hamilton 2006, Sandhu, Dhaliwal et al. 2010) present a complete list of these measures and discuss their significance.

In this study, we will be using association rules with *support*, *confidence* and *lift* measures. *Support* is the measurement resulting when we divide the number of occurrences of a specific itemset by the number of transactions. *Confidence* measures the probability of the appearance of item “b” in the consequent after item “a” appears in the antecedent. Depending on the context of the problem, *support* and *confidence* are commonly used in association rules to distinguish between strong and weak rules. The *lift* measure compares whether the proportion of transactions containing item “b” and item “a” is greater than the proportion of transactions that only contain item “b” among all transactions (Merceron and Yacef 2008). The *lift* measures the strength between two items.

Subjective measures are therefore related to the needs and interests of the user and the domain (Bing et al. 2000, Oliviera et al. 2009). Many authors have proposed subjective measures (Bing, Wynne et al. 2000, Geng and Hamilton 2006), (Silberschatz and Tuzhilin 1996); in particular, Geng and Hamilton (Geng and Hamilton 2006) propose nine criteria to determine whether a pattern is interesting: *conciseness*, *coverage*, *reliability*, *peculiarity*, *diversity*, *novelty*, *unexpectedness* (also called *surprisingness*), *utility* and *actionability* (also called *applicability*). We retained three of these nine criteria to apply to teacher's dataset: *novelty*, *unexpectedness* and *actionability.* They were selected because we are interested in patterns that correspond to accepted beliefs, patterns that contradict beliefs, and patterns that lead to the actions that can be taken to obtain an advantage.

Novelty (Geng and Hamilton 2006) can be recognized directly by the user as something new that does not contradict his own beliefs. A novelty pattern is an unknown pattern that cannot be inferred from other patterns (Geng and Hamilton 2007) (Xin Chen 2006). A related measure to the novelty criteria is usefulness. Usefulness conveys the sense of action, immediate application, doesn't contradict any belief, conveying a sense of action. For these reasons, our research will focus on the *usefulness* as part of the *novelty* criteria.

Unexpectedness is a pattern that contradicts a user's beliefs. Three different approaches have been proposed to cover this measure (Geng and Hamilton 2006). The first one consists of choosing unexpected patterns based on the user knowledge specifications. In the second approach, uninteresting patterns are eliminated using the feedback from the user. For the third approach, the user gives specific constraints to narrow the search space (Geng and Hamilton 2007). Our research focuses on the first approach because it provides an entry point for user's knowledge.

With respect to the *actionability* measure, if the pattern is immediately applicable then it is actionable. The use of *actionability* to measure the applicability of a pattern presents a difficulty because the actions taken after knowing a pattern is actionable could vary depending on the background of the decision makers, their personality and their decision making style (Bing, Wynne et al. 2000). To avoid this, it is necessary to reinforce the use of *unexpectedness* as a way to reach the *actionability*.

In summary, we retained three subjective criteria to help us find interesting patterns: *usefulness, unexpectedness*, and *actionability*.

4.2.3 Subjective criteria evaluation form

A subjective criteria based on the knowledge was proposed (Oliviera Rezende Solange 2009). Knowledge can be divided into five categories: *unexpected*, *useful*, *obvious*, *previous* and *irrelevant*. *Unexpected* knowledge is a pattern completely new to the user; he or she is surprised when it happens. *Useful* knowledge is a pattern that is not outstanding but could help the user in decision making. *Obvious* knowledge is a pattern that is already known and users are aware of it. *Previous* knowledge is a pattern that represents some old knowledge. Finally, *irrelevant* knowledge is a pattern without any importance. We extended this list (Oliviera Rezende Solange 2009) by adding the *interesting* classification. This *interesting* classification allows the teacher to evaluate his own rules and include two of the subjective measures mentioned before, usefulness and unexpectedness. Teacher completes the evaluation within a group of patterns pertaining to him and chooses the *interesting* one within a group. We propose both an individual and a group evaluation of rules in this step. In this way, the teacher suggests which unexpected patterns or useful patterns are more interesting to them. This will be explained in detail in section 4.4.3.2.

4.2.4 D3M and PA-AKD framework

The fusion of DM results with DK is the fundamental trait of the domain driven data mining methodology (D3M). D3M focuses not only on the data but also on the peripheral domain knowledge. D3M presents four layers (Cao, Zhang et al. 2010): the domain layer, the knowledge management and ubiquitous intelligence layer, the theoretical foundations layer, and the specialized techniques layer. Cao et al. (Cao, Zhang et al. 2010) describe these four layers in detail, while suggesting to complement them with new tools, methodologies and frameworks that will contribute to both D3M and AKD.

Four different frameworks have been proposed for D3M (Cao, Zhao et al. 2010): *post-analysis based AKD* (PA-AKD), *unified interestingness metrics based AKD* (UI-AKD), *combined mining based AKD* (CM AKD) and *multisource + combined mining based AKD* (MSCM-AKD). Each framework has relevant characteristics. For this work, we choose to apply the PA-AKD framework because it treats technical and subjective aspects of the problem, and also because it focuses on only one dataset. The framework works in two steps as is described in Figure 4.1.

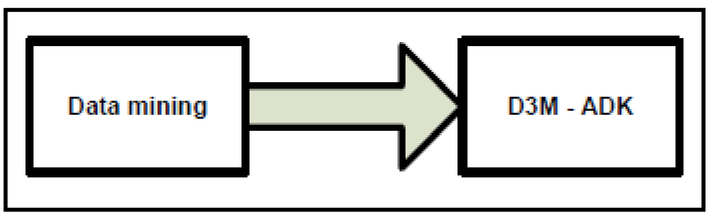

Figure 4.1Post Analysis based

The DM step consists in finding patterns and evaluating them using technical interestingness measures. Figure 4.2 shows the complete data mining step. General patterns are obtained using DM tools. This output is used as the input in the D3M-AKD step.

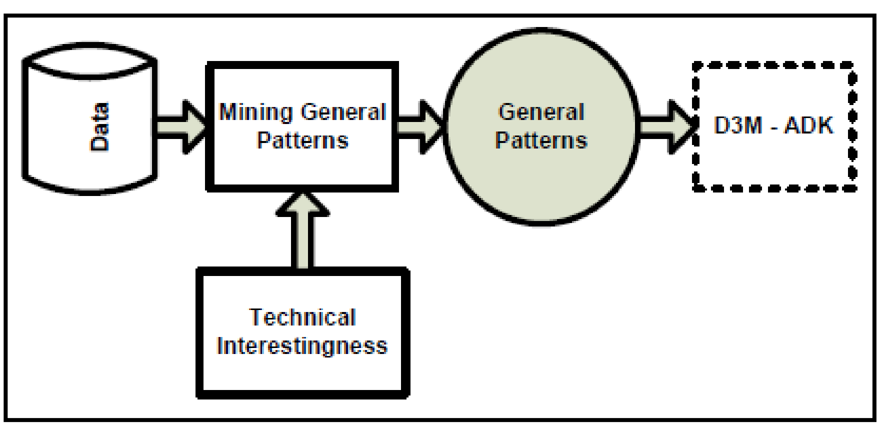

Figure 4.2 Data mining steps

The D3M-AKD step takes the results from DM and with the help of business interestingness measures as well as domain knowledge and meta-knowledge, identifies the most applicable interesting patterns in the form of *deliverables* or applicable patterns. Figure 4.3 shows the D3M-AKD process

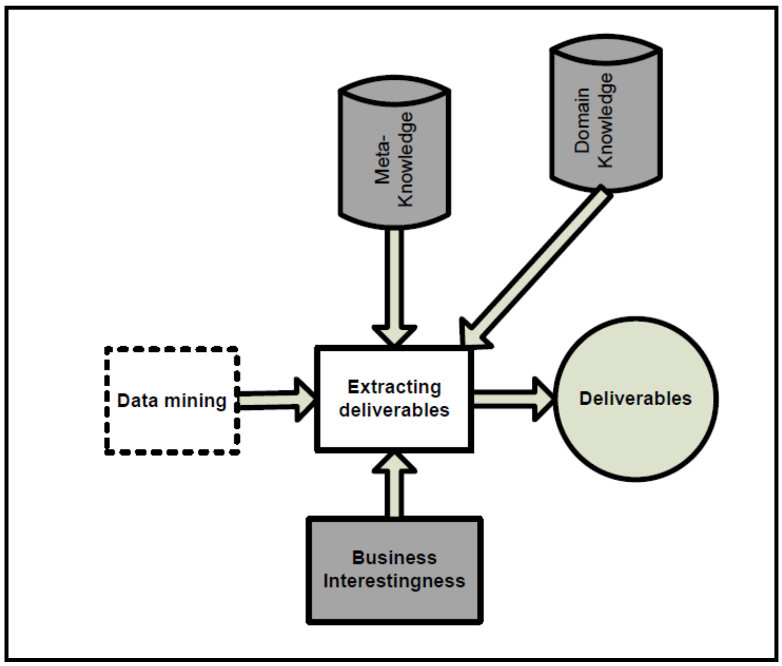

Figure 4.3 Post Analysis-AKD's approach
Taken from (Cao, Zhao et al. 2010)

4.3 Dataset

Our study uses a survey database from a Latin-American university. The database holds 64,138 survey questionnaires answered anonymously for the year 2009 (43 Mbytes in csv format). It contains information about 798 teachers who, as a whole, have given courses to 13,000 students. The university is composed of twelve schools (faculties or schools) and institutes. Each of them provides services to between 218 and 2,300 students per year.

We chose to work with the questionnaires for the faculty of electricity and computing (FIEC) because it has a high percentage of surveys as seen in Figure 4. Moreover, this faculty has a mean evaluation of over 8/10; we therefore have a better chance of finding teachers with well established, well defined and well evaluated teaching abilities.

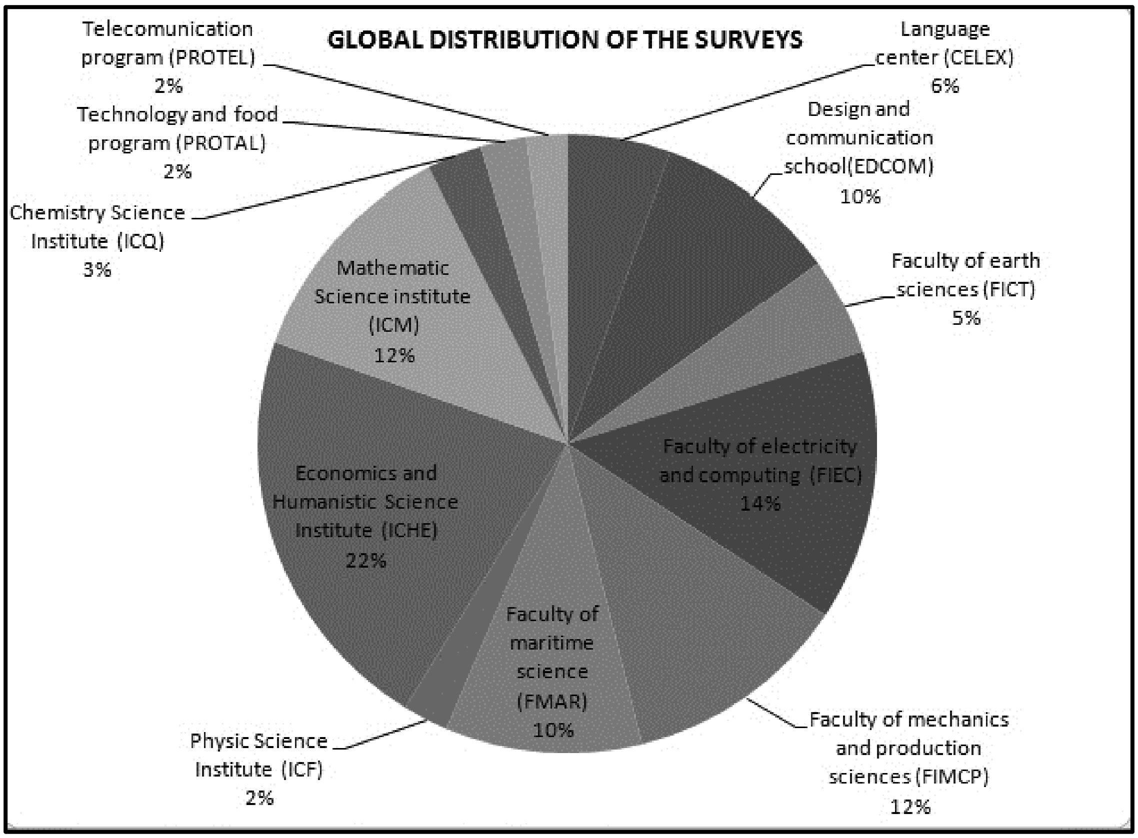

Figure 4.4 Surveys per University Faculty and Institute

Students usually take four to six courses each semester, resulting in eight to twelve courses per academic year. At the end of each course, each student has to fill out a survey.

Table 4-2 Areas and questions / items per area

Area	Attributes

Area	Attributes
Design (D)	X_1Uses audiovisual help to support the content of the class X_2Fulfills the program proposed at the beginning of course X_3Evaluates student participation periodically in class X_4Evaluations fit the themes developed in class X_5Provides clear instructions for learning assessment (tests, quizzes, presentations, simulations, dramatic representation, role playing, etc.) X_6Motivates students to do additional research
Learning promotion (LP)	X_7Explains the course schedule at the beginning of the course X_8Explains class policies at the beginning of the course X_9Encourages active student participation in class X_{10} Summarizes key ideas discussed before moving to a new unit or topic X_{11}Establishes relationships between new concepts and those already known whenever possible X_{12}Motivates learning of the course material X_{13}The teacher is willing to answer questions and offer advice within and outside of the classroom X_{14}Promotes reflection on topics covered X_{15} Maintains fluid communication with students X_{16}He/she is respectful towards students X_{17}Responds to questions in class about subjects related to the field X_{18}Delivers class content in an organized way X_{19}Develops class content in an understandable way
Production and teaching materials (PM)	X_{20}Prepares instructional, bibliographic or other resources to facilitate learning X_{21}Frequently uses schemes and graphics to support his/her explanations
Education managem ent (EM)	X_{22}Provides the results of the assessments on time X_{23}Attends classes on time
General evaluation	X_{24}Considering all the features, choose a score between 1 and 10 to evaluate teacher's overall performance

The data collected in these questionnaires includes information about the year, the semester, the teacher's name and answers to the 24 questions that evaluate his teaching practices. Table 4.2 describes the questions that help evaluate each area.

These 24 questions evaluate four areas of the class development: *design* (D), *learning promotion* (LP), *production of learning materials* (PM) and *education management* (EM).

The Research Center and Education Services (CISE) has been using this survey every semester for now 20 years. The center uses the data from the survey and then constructs and delivers the results to the teachers. Each area has specific questions that capture students' satisfaction. Up till now, each question is presented with a Likert scale(Marshall 2005) between 1 (strong disagreement) and 10 (strong agreement). For the moment, the university do not intend to change the type of scale used. Future work will show if it might be possible to use a wider scale in order to capture shadows of opinion.

For ease of exposition, we use the short form "$X_{N\text{-}DA}$" to refer to each question; the prefix "X" to identify it as a variable, "N" to specify the question number, and "DA" to supply a short description. At the end of the semester, students evaluate the course using a general score for attribute X24. This attribute is not part of those that we are going to evaluate. We use the X24 attribute as the dependent variable in the teacher's regression process.

We are interested in the antecedents and consequents of the association rules. We will look for 3-itemsets with 2 attributes in the antecedent side and identify 1 attribute in the consequent side.

4.4 Methodology

Our methodology consists of three steps: first, we apply DM to obtain general patterns from students' dataset; second, we construct two questionnaires and one interview to elicit DK and MK from the teacher; finally, we construct a model with the rules, the knowledge and meta-knowledge retrieved. Figure 4.5 presents a synthetic view of the three processes of the methodology: DM, DK and MK. In the following subsections, we explain each step in detail.

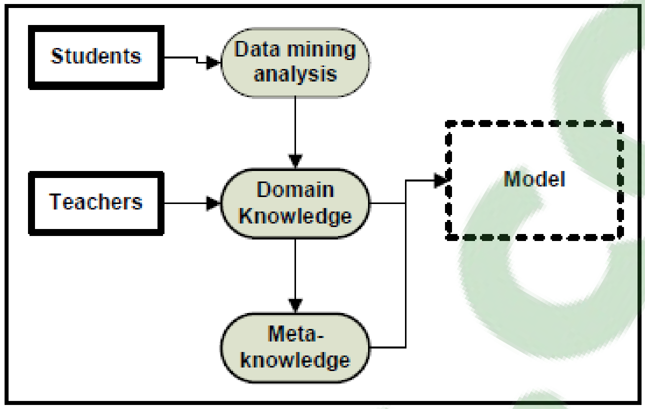

Figure 4.5 General view of the methodology

4.4.1 Objective evaluation

We will refer to the student evaluations at the end of each course as the *objective questionnaire* (or *obj-Q*). The *obj-Q* is the dataset containing the teacher evaluation surveys filled out by students. The attributes are listed in the Introduction Chapter. The *obj-Q* questionnaire conveys the students' perspective about the teacher. In section 5, we illustrate our proposed methodology by analyzing *obj-Q* for a specific teacher and course.

The dataset *obj-Q* is cleaned and converted to a csv format file. Evaluations with values of 0 or 1 in all questions as well as evaluations with text in the fields instead of numbers are eliminated. Using SPSS (Foundation 2000), we apply linear regression to control the dimensionality of the data and obtain a set of variables that have a strong correlation between them. We refer to this regression result as $obj\text{-}Q_{reg}$.

We construct another dataset based on the teacher's preferred attributes–where they choose their own attributes from the same questionnaire *obj-Q*. This second dataset is called *sub-Q1-Tatt,* and is part of the *sub-Q1* questionnaire. The *sub-Q1* is a questionnaire that retrieves the domain knowledge from the teacher. We now provide a brief explanation of this dataset with more detail in section 4.4.3

We apply regression to the *sub-Q1-Tatt* dataset, where the teacher selected the attributes of the *obj-Q* that were most representative for him. We obtain a reduced model based on the attributes selected by each teacher. We call this result *sub-Q1*$_{reg}$.

We compare the two regressions: *obj-Q*$_{reg}$ and *sub-Q1*$_{reg}$. If the attributes are the same in the student's model (resulting from the *obj-Q*$_{reg}$) and in the teacher's model (resulting from the *sub-Q1*$_{reg}$), then no further treatment is necessary because students` perspective variables are the same as teachers perspective variables. If they are not the same, we apply the data mining procedure to *obj-Q*$_{reg}$ variables as explained in the next sub-section. Figure 4.6 shows the comparison between these two datasets.

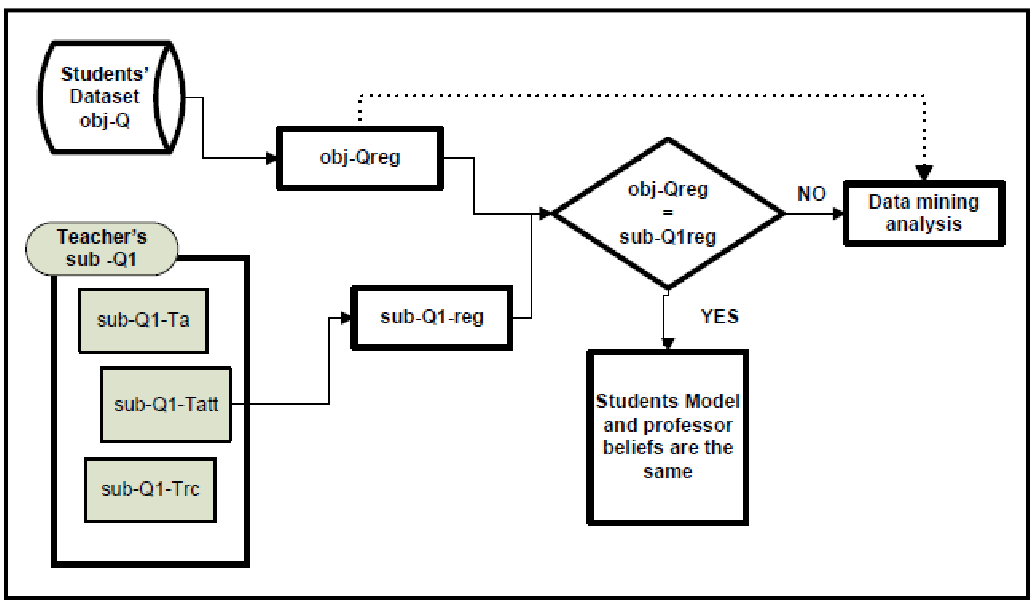

Figure 4.6 Comparison between obj-Qreg and sub-Q1reg

4.4.2 Data mining analysis

Starting with *obj-Q*$_{reg}$ variables, we first apply the *Apriori* Algorithm of the Arules package in R to generate the rules and we use the item frequency plot from the same package to identify the frequencies of the variables. We search for association rules with specific *support*, *confidence*, and *lift* boundaries. Rules with *lift* values higher than 1 are selected. A high lift value tends to express stronger relationships between attributes. These rules are called *general rules*. We obtain association rules in the form of X, Y->Z. The association rule is formed by antecedents and consequent. X and Y are in the antecedent side and Z is the consequent side. The teacher is then asked to evaluate *general rules* in *sub-Q2;* this is explained in section 4.4.3.2

4.4.3 Subjective evaluation

For the subjective evaluation, we construct two questionnaires (*sub-Q1*, *sub-Q2*) plus an interview (*meta-I*). The *sub-Q1* is a questionnaire that retrieves the domain knowledge from the teacher. The *sub-Q2* is the second questionnaire that identifies the interesting rules and rules that are part of the domain knowledge. The *meta-I* will help to obtain the meta-knowledge from teachers that is not available in *obj-Q*.

4.4.3.1 Domain Knowledge: *sub-Q1*

We start the subjective evaluation with *sub-Q1*. Figure 4.7 summarizes the process of this questionnaire and what it achieves.

The *sub-Q1* gives the teacher the opportunity to evaluate the same *obj-Q* filled out by the students, but in this case with the teacher's preferences. The s*ub-Q1* is composed of three components that capture the domain knowledge from the teacher. First, teachers select the most important areas from the *obj-Q* (we call this the *teacher area selection* or *sub-Q1-Ta*). These areas were mentioned previously in the Introduction Chapter: Design(D), Learning Promotion (LP), Production of Teaching materials (PM) & Education Management (EM). Second, the teachers choose the most important attributes per area that help them attain high evaluations (we call this *teacher attribute selection* or *sub-Q1-Tatt*). Each area has a group of attributes. Each attribute focuses on an aspect of teaching in that area. The attributes that can be chosen are those presented in the Introduction Chapter. We construct a new dataset with the selected attributes. As was mentioned in section 4.4.1, s*ub-Q1-Tatt* was the teachers' dataset to which we applied regression and obtained *sub-Q1*$_{reg}$.

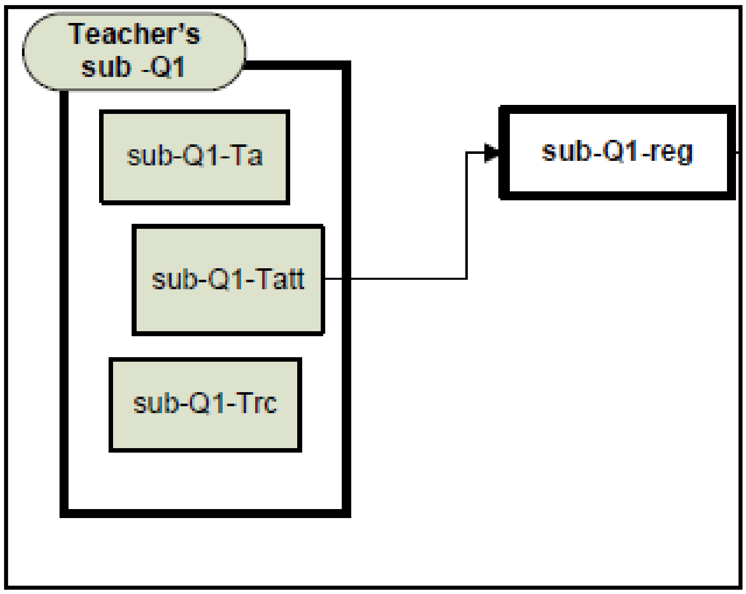

Figure 4.7 Subjective endorsement sub-Q1

In the third part of *sub-Q1*, teachers are invited to construct rules based on what they believe has worked in their classes (we call this *teacher rule construction* or *sub-Q1-Trc)*. Teachers construct their own rules, suggest attributes and place them in the antecedent or consequent areas from an association rule, as they deem relevant. These manually constructed rules are the rules of thumb that they perceive work in their class. Teachers suggest which attributes placed in the antecedent could produce a specific attribute placed in the consequent. This process is repeated for each important area identified for the teacher. *The sub-Q1-Trc* is part of the domain knowledge. The rules suggested by the teacher are incorporated into the domain knowledge.

4.4.3.2 Interesting rules and domain knowledge: sub-Q2

General rules are the input for the s*ub-Q2* questionnaire. The task is to identify those rules that, according to a teacher's experience, help them improve their teaching abilities. To accomplish this task, the questionnaire *sub-Q2* proposes one set of *interestingness* criteria along with five knowledge categories: *unexpected*, *useful*, *obvious*, *previous*, and *irrelevant* (as indicated in section 4.2.3).

Each rule is evaluated individually and within a group of rules with the same consequent but with different antecedents. For the individual analysis, each rule has to be evaluated using the five knowledge categories. The teachers select only one of the knowledge categories per rule. In the case that a teacher finds a rule classified as *irrelevant*, he/she explains why this association rule is irrelevant.

During the group evaluation, teachers select the most *interesting* rule from the group (interestingness classification). Each group has one interesting rule, because only one rule associated with a specific consequent should attract their attention the most; we encourage teachers to select those that really suggest something different from what they were doing in class. An example of the individual and global evaluation is presented in Table 4.3**.**

Table 4-3 Sub-Q2 Questionnaire

	Interesting	**Unexpected**	**Useful**	**Obvious**	**Previous**	**Irrelevant**
X1-X3->X5	x	x				
X1, X4->X5			x			
X2, X4->X5					x	

In Table 4.3, there are three rules that evaluate the same consequent X5. Each rule is evaluated using the five categories, *unexpected*, *useful*, *obvious*, *previous*, and *irrelevant* and only one rule from the group is evaluated in the interesting classification.

All the rules evaluated as *interesting*, *unexpected* and *useful* constitute the *interesting* rules. These rules show aspects that are noticed by students and not by the teacher; we use them to generate the “new and improved rules model”. This model should provide suggestions to improve a teacher’s performance (what students think the teacher is doing but the teacher is not aware of).

All the rules classified as *obvious*, *previous* and *irrelevant* constitute the common rules and show aspects already known by the teacher and by students. We called this the “real and actual rules model”. This model represents the domain knowledge in class and is referred to by students through the questionnaire *obj-Q* and confirmed by teachers through the s*ub-Q2*. Figure 4.8 summarizes this process.

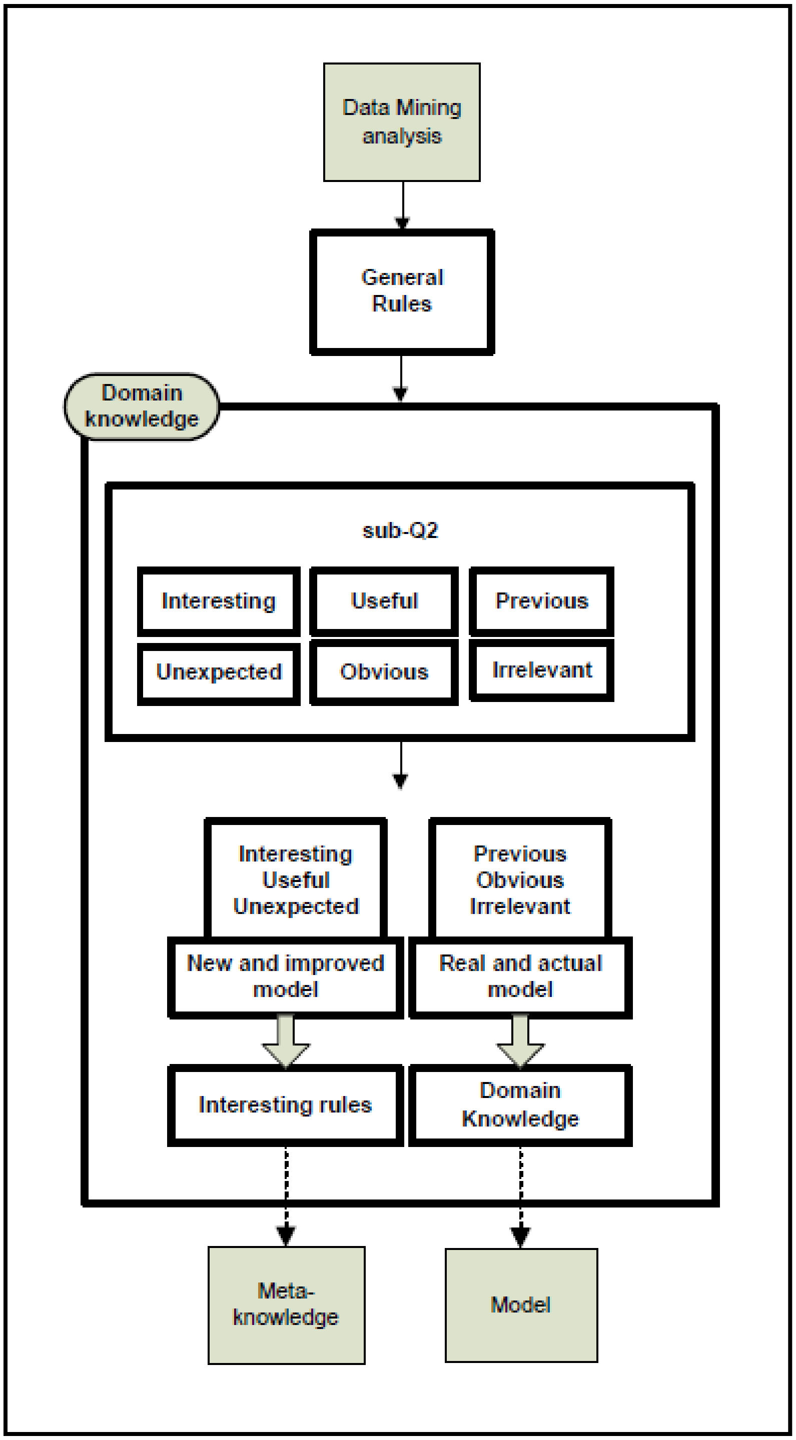

Figure 4.8 Domain knowledge retrieving process

This analysis helps gather domain knowledge in the form of common rules from the teacher. It identifies the interesting and new rules as well. Next, teachers are interviewed using a questionnaire called meta interview or *meta-I* to examine the selected interesting rules and understand the meta-knowledge associated with their course activities as explained in the next section.

Meta-knowledge: meta-I interview

The interview (*meta-I*) is based on the interesting rules classified in the last step. The interview covers a teacher's practices, activities, organization of the course, and the hidden knowledge about how they teach in class. The purpose of the *meta-I* is to extract meta-knowledge about each interesting rule. The final model can then be constructed. Figure 4.9 summarizes the process.

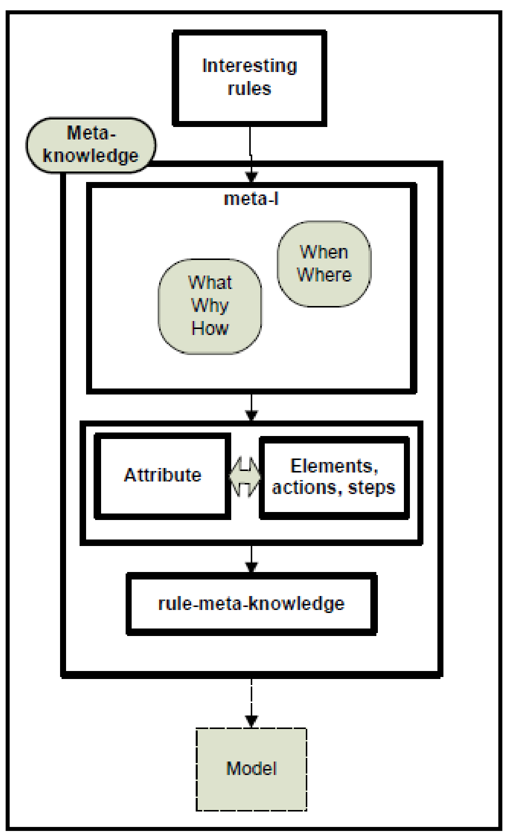

Figure 4.9 Meta-knowledge retrieving process

We asked questions about each interesting rule evaluated. The questions explore the three traditional questions we need to ask to identify meta-knowledge: *why*, *what* and *how*, and in addition we considered *where* and *when* questions to be pertinent. Each rule is composed of one or two antecedents and only one consequent. During the interview, we asked teachers questions such as: *Why* is attribute X10, in the antecedent of this rule, interesting?; *How* do you reach the objective of the attribute X10?; *What* activities do you perform to reach a well developed X10 attribute?; *When* do you apply this rule with the attribute X10 in the timeline of the class, at the starting point of the course (first class) or at the beginning of each class?; *Where* do you make students work in regards to attribute X10–in the lab or in class?

Answers to these questions provide descriptions of teaching processes that include different elements not explicit in the objective questionnaire and that are related to each rule. These elements include: the actions needed to complete the activity, the elements taken into account during the activity and the steps students need to perform to accomplish the activity. We call this new knowledge *rule-meta-knowledge*. This knowledge is going to fuse the rule attribute descriptions with the acquired meta-knowledge. As mentioned earlier, the attribute description is the *DA* component of the *"$X_{N\text{-}\boldsymbol{DA}}$"* attribute short form, used to refer to each question in the questionnaire *obj-Q1* and *sub-Q1-Tatt*.

In the following example, X9 is the attribute and "Encourages active student participation in class" is the description of the attribute, or *DA*:

- *(DA1): X9* Encourages active student participation in class

When we adapt the attribute description to the *rule-meta-knowledge*, it will describe more elements associated with the *DA*. This includes all the different elements teachers use to express an attribute during the interview. (DA1) is an example of an attribute description, (DA2) , (DA3), and (DA4) are attribute descriptions where the meta-knowledge has been incorporated.

Let’s say that the teacher used the following expression during the interview: “I encourage students to ask questions–not to agree with me–it is better this way, students then need to find stronger arguments to defend their points”. This information will then be incorporated into the model as follows (DA2, DA3, DA4):

- *(DA2) X9*: The teacher encourages students to ask questions.
- *(DA3) X9*: The teacher encourages students not to agree with his/her ideas.
- *(DA4) X9*: The teacher reinforces student’s self-confidence to make judgments conducting a questions session, eliciting an opinion and contrasting it with other student opinions.

Each attribute is enriched with more details obtained from the meta-knowledge. It is improved in an explicit way and made more comprehensible to the teacher. The *rule-meta-knowledge* per attribute is not the same for every teacher; it changes from teacher to teacher.

Figure 4.10 presents the complete methodology with DM, KD, MK and the participants (teachers and students).

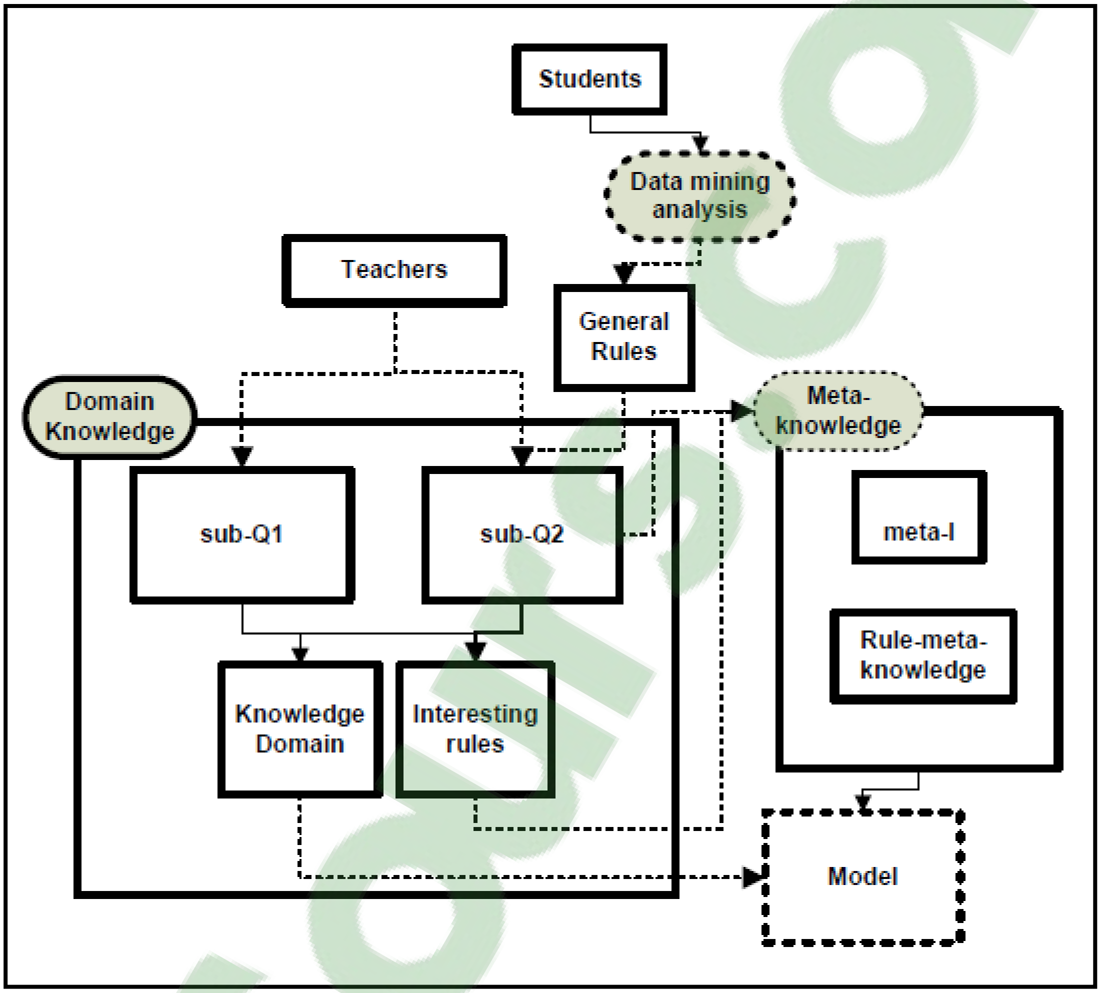

Figure 4.10 Complete methodology

Then, the final model includes the *sub-Q1-Trc* rules that are part of our domain knowledge (see Figure 4.11), as well as the *interesting rules* and the domain knowledge in the form of rules classified as the "new and improved rules model" and " the real and actual rules model", respectively (see Figure 4.8) which are obtained from *sub-Q2*. Lastly the *rule-meta-knowledge* from meta-I (see Figure 4.12) is incorporated.

4.5 Results for the Entrepreneurship course example

In this section, we present the results of applying our methodology to an entrepreneurship course. Section 4.5.1 refers to the objective evaluation phase. Section 4.5.2 presents the results of the subjective evaluation with *sub-Q1* questionnaire. Section 4.5.3 focuses on the results of the questionnaire *sub-Q2* and the rules from the improved and actual rules model. Section 4.5.4 focuses on the interview *meta I* and the meta-knowledge obtained from the teacher himself. Finally, section 4.5.5 presents the complete model for this course.

4.5.1 Objective evaluation phase

The *obj-Q* is the dataset for the entrepreneurship course. At the beginning, we had 23 attributes. We applied regression to this dataset using SPSS (Foundation 2000). After the regression, we obtained five significant attributes. The attributes are: LP(X7, X9, X12, X16) and EM(X22) with high correlation. *Obj-*Q_{reg} is represented with these five attributes. The significant attributes belong to the LP and EM questionnaire areas.

4.5.2 Results from Sub-Q1

The teacher completed the s*ub-Q1* questionnaire. In the first part, he began by selecting all the areas of interest and then the attributes for each area that he thought allowed him to achieve high evaluations. The selected areas represented by s*ub-Q1-Ta* are: *design*, *learning promotion*, *production of materials* and *education management*. The selected group of attributes per area represented by *sub-Q1-Tatt* are: D(X1, X3, X5, X6); LP(X9, X10, X11, X12, X13, X14, X16); PM(X20, X21); and EM(X23).

We then applied regression to *sub-Q1-Tatt* on the attributes selected by the teacher. We obtained a group of attributes that form the *sub-*$Q1_{reg}$: X12, X13, X16, and X23. We compared o*bj-*Q_{reg} with s*ub-*$Q1_{reg}$. Two variables were similar (X12, X16), and the remaining ones were different. Next we used the R script and applied *Apriori* on the results from o*bj-*Q_{reg}. We obtained a set of association rules with support of 0.2, confidence of 0.9, and a lift greater than 1. The rules obtained in this step are called *general rules* and are the input to the s*ub-Q2*.

Finally, the teacher generated his own rules. Some of the rules included in his domain knowledge are represented by s*ub-Q1-Trc* as:

X3, X14 ->X6 X5, X13 ->X12
X21, X1->X9 X21, X16->X23
X11, X20->X6

4.5.3 Results from Sub-Q2, actual model and improved model

We constructed s*ub-Q2* with the *general rules* obtained from the previous step. The teacher evaluated them using the five knowledge categories and the "interesting classification" mentioned earlier. From s*ub-Q2*, two rules were classified as part of the "new and improved model" (interesting rules). The rest of the rules were classified as "real and actual model" (knowledge domain). We added to these rules the ones created by the teacher in s*ub-Q1-Trc* as part of the actual model. The new model is built using the two interesting rules:

X9, X22->X7
X12, X22-> X16

In Figure 4.11, all white nodes represent association rules that are part of the DK. All black nodes represent interesting association rules and all nodes with black and white circles represent association rules that connect domain knowledge with the interesting rule. Antecedents and consequents in each association rule are connected with a connector (small black circle in Figure 4.11).

In part (a) of Figure 4.11**,** association rule attributes X3 and X14 are in the antecedent and attribute X6 is in the consequent. In part (b) of Figure 4.11**,** association rule attributes X9 and X22 are in the antecedent side of the rule X9, X22 -> X7, but X9 is also in the consequent for association rule X21, X1 -> X9. Nodes X9 and X12 have double circles (white and black) representing the fact that attribute X9 and X12 are part of both DK and interesting rules and help connect them. We have two interesting rules represented by black circles, X9, X22 ->

X7 and X22, X12-> X16 where attribute X9, X22 and X12 are in the antecedent of the interesting rules and X7 and X16 are in the consequent in the same interesting rules. These two rules complement the DK.

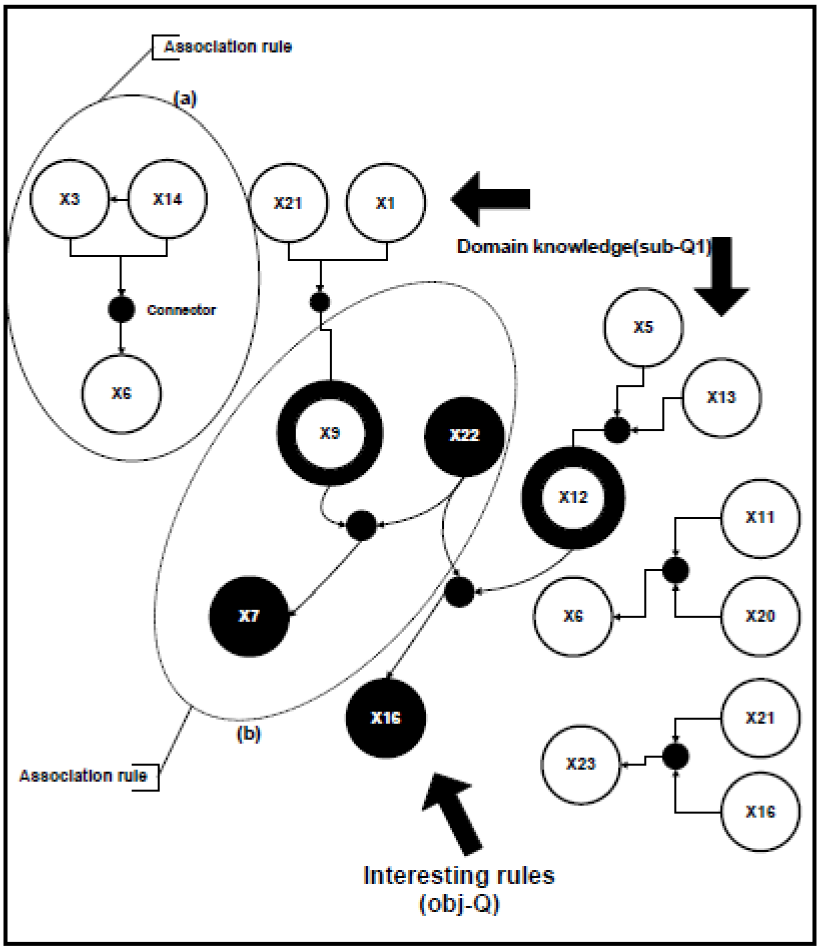

Figure 4.11 Domain knowledge and Interesting rules graphic

4.5.4 The interview and the meta-knowledge

With this diagram in hand, we interviewed the teacher. We applied the questions to obtain the meta-knowledge for each association rule and each attribute in the rule:

Why is attribute X12 important for your class? *What* objectives do you have to reach to ensure attribute X12 is well evaluated? *How* do you reach this attribute (strategies)? *Where* do you apply the actions to realize attribute X12? *When* do you try to develop this attribute X12?

We applied the same questions to the association rules presented in the domain knowledge. We asked how he reaches each association rule. In this case, rules X9, X22 and X7, and each of the attributes associated with these rules. In this step, the teacher applies the *why*, *what*, *how, when and where* question to each of the attributes and then makes an analysis of the attributes together. It is possible that during the interview a few more questions are asked to clarify the teacher's response. For example:

- When did you develop this rule in class, at the beginning of the course or at the end?
- What does this rule mean to you?
- Do you have specific examples of where this rule has been applied?
- Is this a frequent rule? Do you use it in every class, or every semester?

The teacher gave information related to each of the rules in the domain. He explained them in detail, and he provided different interpretations of the same rule. This constitutes the *rule-meta-knowledge* about this specific course and this particular teacher. For example:
Attribute X22: Provides the results of the assessments on time and is expressed in the attribute’s rule-meta-knowledge as follows:

> X22a: Provides the results of the assessments on time using all the communication tools he has on hand; email, web publications and paper.
> X22b: Provides the results of the assessments and informs students how these results are connected to the theory and the failed or erroneous parts of the exam.
> X22c: Provides the results of the assessments on time; and thus shows that the teacher is very organized with his own time.

In this way, we obtained different interpretations of the same attribute.

4.5.5 Construction of the graphical model

After the interview with the teacher, we are ready to develop a model with the meta-knowledge, domain knowledge, and the interesting rules. Because of space limitations, Figure 4.12 only shows one of the two best rules with the *rule-meta-knowledge* retrieved from the teacher's interview. It shows a rule with two antecedents and one consequent. The nodes X9 and X22 are the antecedents of the rule and its consequent is X7. X9 comes from domain knowledge; X22 and X7 come from the interesting rules. We represented the meta-knowledge with letters A, B, C, D, E, F, G and H and attached them to their corresponding attribute. The smallest black circle in the middle is a connector used to attach attributes in association rules.

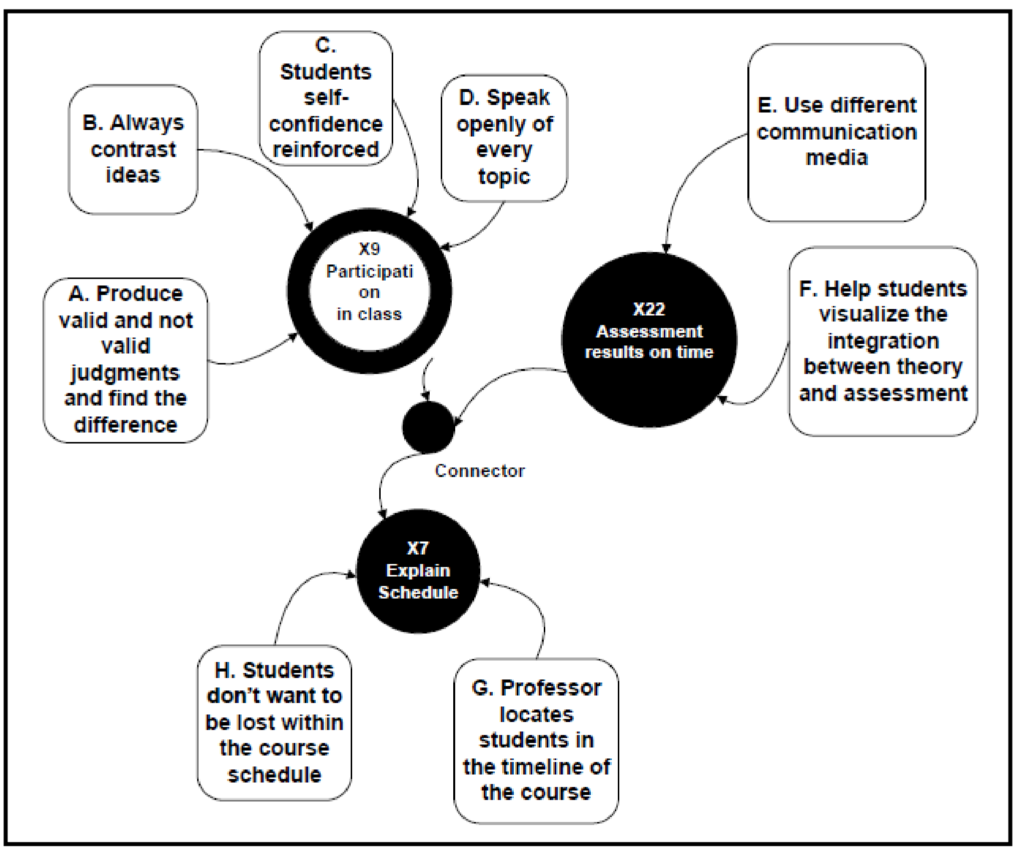

Figure 4.12 Improved model with the meta-knowledge included

We used two different datasets with two different perspectives to construct this model. Both perspectives have a representation in the final model. The student perspective is shown with the white nodes and a teacher's domain knowledge is shown with the black nodes. The double circle node X9 shows that its attributes are part of both, the old and the new knowledge, and that they are connected. The teacher of the entrepreneurship course found different elements interesting in this rule. For example, A, B, C and D rule-meta-knowledge show the different strategies he uses to stimulate student participation in class: freedom to express themselves; confidence expressing ideas; contrasting concepts; and generating judgments and decision making.

Antecedent X22 has the rule-meta-knowledge E and F. E expresses that the teacher uses all types of communication media (internet, SMS, content system management) to deliver their grades to students and to be available to them at all times. The F expresses that providing the assessment results on time helps students visualize the integration between theory and tests (see Figure 4.12). In the same rule, the consequent X7 has the rule-meta-knowledge H and G. H expresses the fact that students do not want to be confused in regards to the course schedule. G expresses how the teacher establishes the point students have reached within the timeline of the course.

The student dataset creates the teacher's domain knowledge; the teacher can select and analyze the association rules and select those that show something interesting. These interesting characteristics, combined with the usefulness and unexpected knowledge categories, produce actionable patterns. The teacher classifies the patterns as useful or unexpected. He recognizes specific patterns applied and defines them as part of his signature during class. This is more frequent with patterns classified as useful.

We said these patterns are actionable for different reasons: first, rules can be tested during the following semester using the different actions and strategies the teacher mentioned in the graphical model. The graphical representation shows different strategies and actions related to specific attributes; therefore, he can increase, improve or reduce the amount of activities he performs for each attribute. Finally, new relations between attributes, actions and strategies can also lead to discovering new relations or patterns between attributes.

For example, in this rule X9, X22 -> X7, these antecedents and consequent in and of themselves do not describe anything understandable for the teacher at first sight. When the rule-meta-knowledge is added, a relationship is found between antecedent X22 and the consequent X7 and the relation with attribute X9 gives more meaning to the relationship. The communication factor exists when the teacher returns the assessment to the students (attribute X22). Students and teacher work with this communication. Feedback develops between them based on the homework or course timeline (attribute X7). Finally, being informed about their grades and having homework feedback helps students reinforce their self-confidence in class.

The representation helps teachers further improve their effectiveness at work. A database of different descriptions about each attribute helps provide clarification to express and understand the questions of the survey. The same attribute could mean different things to different teachers. This methodology shares some common understandings for an attribute.

4.6 Discussion

These final models visually express teachers' experience or, what can be called, their "rules of thumb". In so doing, the university authorities are in a better position to advise new teachers in the field. For teachers, the graphical resulting models offer a comprehensive access to good practices giving them an integrated view of students' perception and teachers' reflections.

4.7 Conclusions

The methodology presented in this paper provides a bridge between association rules, knowledge domain and meta-knowledge. We start with association rules that once evaluated and filtered, form the basis for the construction of a model from the point of view of the student. We complement this student model with the teacher's model. The teacher's model is constructed with the interestingness rules, the knowledge domain and the meta-knowledge. An interview with the teacher creates the rule-meta-knowledge that further adapts and enriches attributes in the models.

4.8 Results

The proposed methodology provides the following results:

- The methodology complements the AKD framework because it proposes a new graphical representation of the meta-knowledge and the domain knowledge.
- The model provides a useful tool to integrate different perspectives that complement each other.
- Association rules about old knowledge and association rules about new knowledge are presented in a unified representation.
- The methodology allows for the creation of meta-knowledge and domain knowledge. The AKD framework assumes the pre-existence of a meta-knowledge base and a domain knowledge base and treats them as independent elements. In our methodology, we start by creating domain knowledge and rules of thumb which are then used, along with the user knowledge, to construct the meta-knowledge.
- We show how to effectively use *interestingness*, *unexpectedness* and *usefulness* as subjective measures to obtain actionable patterns.
- We use the usefulness measure to identify interesting rules. Rules that are classified as useful can be more easily explained by the user than those evaluated as unexpected. Users express more meta-knowledge from useful rules than from the unexpected rules. This provides much more actionability to the rules. It is clearer and more action-oriented for stakeholders.
- The evaluation of the rules through subjective measurements helps select specific rules and prevents us from testing rules of no interest to the user.

- We suggest the addition of *where* and *when* questions to any well-planned interview for the retrieving of meta-knowledge. These questions retrieve meta-knowledge about time and space where the knowledge process is generated.
- The meta-knowledge obtained from the interviews can be used to create a database of common understandings to different users of what an attribute expresses. The construction of the rule-meta-knowledge database helps reduce confusion with respect to the descriptions of attributes.
- The final representation with association rules, domain knowledge and meta-knowledge provides valuable insights of the models for the stakeholders, managers as well as authorities from the institution and shows how association rules are related.
- In case we applied a different scale than Likert scale, it is necessary to understand the new one first and then adapt it to the methodology. The methodology might work with a different one.

4.9 Future Work

Actually, there exist some platforms that deal with interviews and feedback from clients such as “Customer Insight and Action” platforms or “Enterprise Feedback Management” (surveys, contact centers conversations, customer feedback comments, phone interviews and behavioural interviews) (Hirschowitz 2001, Bailey, Baines et al. 2009).

These platforms are used in Human Resources, IT, or Marketing and Sales groups, where questionnaires and the indirect interaction with clients or with workers clarify different expressions related to what the client senses as good service or in the case of an employee, what he senses as a good work environment.

Our future work will be focused on the application of the same methodology to the customer relationship or the client service department to identify the meta-knowledge associated with an enterprise and demonstrate how this representation model can be applied to provide actionable patterns that lead to better customer service.

This kind of work, in conjunction with client service, represents a new way to analyse patterns within a business. The identification of meta-knowledge that is part of “day to day” activities and the effort to relate this meta-knowledge to known activities can transform the way people see simple patterns in data mining and applicable patterns in any business area.

4.10 Acknowledgments

“CISE, Education Service and Research Center” from Escuela Superior Politécnica del Litoral in Ecuador, for contributing with the data.

This research is sponsored by a scholarship for postgraduate students given by SENESCYT, Secretaría Nacional de Educación Superior, Ciencia y Tecnología e Innovación, at “Academia 2010” Ecuador.

CHAPTER 5

ARTICLE II : FUSION OF OBJECTIVE AND SUBJECTIVE MEASURES TO ASSESS THE UTILITY OF ASSOCIATION RULES

Otilia María ALEJANDRO MOLINA
Software Engineering and IT, École de Technologie Supérieure (ÉTS),
1100, rue Notre-Dame Ouest, Montreal-Canada. H3C 1K3

Computing Faculty, Escuela Superior Politécnica del Litoral (ESPOL),
Km 30.5 Via Perimetral. Guayaquil-Ecuador

Sylvie Ratté
Software Engineering and IT, École de Technologie Supérieure (ÉTS),
1100, rue Notre-Dame Ouest, Montreal-Canada. H3C 1K3

Paper submitted to Springer Data Mining and Knowledge Discovery on February 11, 2016

Abstract

In data mining, the applicability of a pattern is a very important aspect. When patterns are not relevant, doubts arrive about the utility of data mining techniques. This not only happens in business but also in the academia, where teachers want to improve their teaching practices but patterns are not applicable. The generation of patterns coming from a subjective evaluation dataset, and a utility measure with subjective evaluation datasets become relevant. The objective of the utility measure is to evaluate and encourage effort on a truly actionable pattern. Due to the lack of utility mining measures that focus on subjective attributes analysis and a methodology to calculate the utility based on these kinds of attributes, we proposed a semantic utility formula constructed using two perspectives: objective and subjective. This paper also provides a survey instrument to obtain subjective data for cases when subjective data is not readily available from objective datasets. For illustration purposes, we applied our approach to university courses where the utility calculation considers the objective data from courses and the semantic information retrieved from the teacher. In this setting, we aimed to

find association rules and discover the best teaching patterns that produce the highest utility based on teacher's abilities, i.e. best efforts, strong characteristics and best practice.

Keywords: Utility mining; Weight Association rules; Utility measure; Transactional Utility; External Utility.

5.1 Introduction

The utility is a quantitative measure of how useful a pattern is. It brings economic and social benefits to any industry or business. However, it is important to apply and prove the pattern (Jiang, Wang et al. 2005, Lee, Park et al. 2013). Researchers consider utility as a semantic measure because it depends on the usefulness an expert assigns to the associated attributes based on the semantic significance it has for him (Geng and Hamilton 2006). In data mining, we applied utility mining to obtain the highest utility patterns.

In fact, authors claimed that utility measure should include other components besides cost and quantity (Kleinberg, Papadimitriou et al. 1998). These components include subjective information such as: client's preference, supplier' objectives and the business improving component. They reinforce the selection and application of a pattern based on the environment around the measured items.

Moreover, Kleinber, et. al. worked with datasets and objective data (quantity and price values) but there was no analysis on subjective attributes where quantity and price are not part of the dataset. We proposed a utility formula using two perspectives the objective and subjective and enhanced it using the semantic perspective. We proposed the instrument to obtain information from the teacher too. We applied our methodology to a university context and calculated the teacher's attributes that could improve as a survey result. In this setting, we aimed to find association rules and discover the best teaching patterns that produce the highest utility based on teacher's abilities. Our results proved that utility can be measured from a subjective dataset and that is possible to identify the attributes with higher utility for teachers.

This paper is organized as follows: section 1, presents the introduction of the research; section 2, reviews the pertinent literature; section 3, presents the methodology to obtain the new utility formula that includes the objective and the subjective analysis, the predicted and objective attributes analysis and the utility measure; section 4, presents the results from the methodology; section 5, presents the results of the utility formula, and section 6 presents discussion and conclusions.

5.2 Literature Review

Utility mining is a new topic in data mining (Pillai and Vyas 2010). The aim of Utility mining is to identify itemsets or patterns with higher utilities that will improve business incomes. In this sense, Utility mining presents three different approaches depending on the way the utility is gathered and focused: the *technical* (itemset frequency), the *semantic* (itemset weight) and the *objective attribute aspects*.

In the technical approach, frequent itemset mining (FIM) (Lee, Park et al. 2013) found items with high correlations used in marketing, promotion and cross-selling, website analysis, credit evaluations and medical fields. FIM was not related to the semantic significance of the items, but to the objective sense of the data.

In the semantic approach, High utility itemset mining, HUIM, (Raymond, Qiang et al. 2003) captured the semantic significance of the itemsets, relating quantity and price. Neither FIM nor HUIM generated a *benefit* or profit because these approaches focused on itemset level. Weighted association rule mining, WARM, (Wang, Yang et al. 2000, Tao, Murtagh et al. 2003) was another well-known model that associated weights to the items in the database to reflect interest of the item in a transaction. The expert expressed interest on specific items in the transaction. Weighted Utility association rule mining, WUARM, (Khan, Muyeba et al.

2008) emphasized on items' weights as their significance and the items' frequency in transaction. WUARM identified the high selling items that increased profit in business. Sandhu et al. (Sandhu, Dhaliwal et al. 2010), presented an approach based on WUARM, where they worked with Apriori Algorithm, weight factor and utility. This approach gave weight and utility elements to the attributes and presented a combined utility weight score.

Yi et al. used the objective attribute approach to obtain Objective Oriented Utility based association mining, OOA, (Yi-Dong, Zhong et al. 2002). They presented an approach based in objective attributes that are relevant to the user and focus the search of profitable attributes only on the semantic sense of the attributes. They presented the concept of *objective attributes* and placed them in the consequent side of the association rules.

On the other hand, Utility mining considers some components. Yao et al. (Yao, Hamilton et al. 2003) presented a theoretical model for a utility mining approach that referred to the general components a utility mining measure should include. They mentioned two components, the *transaction utility* and the *external utility*. The *transaction utility* retrieved from the dataset (transaction's information). The *external utility* is the additional information not found in the transaction, but rather in the elements around the transaction and possibly included in the database. Wang et al. presented an adaptation of Yao et al. model and called it "general utility mining" given importance to the frequency of itemsets and the partial utilities of the attributes in the itemset. (Wang, Liu et al. 2007).

All the referred approaches worked with objective datasets in the application of utility mining. These datasets allow the calculation of profit, quantity or cost and the application of weights. In case datasets contain subjective information, profit, quantity or cost are difficult elements to define. None of these utilities mining approaches evaluate subjective attributes. In these sense, the utility of a subjective attribute should consider people's judgement and opinion. Oliviera et al. used five knowledge categories to classify subjective patterns (Oliviera Rezende Solange 2009): interesting, unexpected useful, obvious, previous and irrelevant. The~~se~~ authors classified association rules using these categories. A pattern is interesting if the rule is new and different; it is unexpected if it contradicts specific beliefs; useful if the pattern complements knowledge, about what is already known; obvious if the pattern confirms a common action; previous if it is an old pattern already known; and irrelevant if it is a pattern that is not interesting at all.

To create a utility measure that evaluates objective and subjective attributes, we designed a measure that integrated objective and subjective utilities. We used Yao et al.'s and Yi et al's. approaches. In addition, we included the "effort" required to improve the attribute utility, in case it is very low. We applied this measure to a group of teachers' patterns. These patterns are association rules generated from an objective students' dataset. Additionally, we retrieved subjective information from teachers about the objective dataset. Our measure considered the importance teacher gave to each survey question, the importance students' gave to the survey questions, and the effort to improve attributes when they were low evaluated.

5.3 Methodology

Our methodology consisted in four steps. The first step was an objective analysis that applied linear regression (**section 5.3.1**) to the student dataset. The second step was a subjective analysis to obtain open rules, context rules and evaluation rules (**section 5.3.2**). The third step derived the predictive and objective attributes to form a new dataset that would be part of the Apriori algorithm (**section 5.3.3**). Finally we analysed the resulting association rules in step 4, with the help of our utility measure (**section 5.3.4**). Figure 5.1. presents the methodology illustrated.

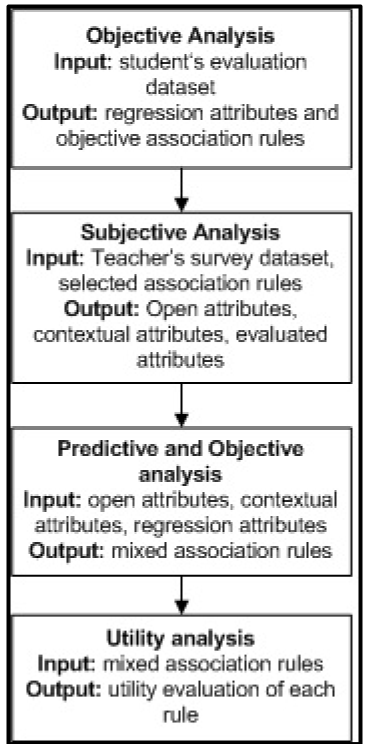

Figure 5.1 Methodology

5.3.1 Objective analysis

We illustrated the methodology with a dataset called the *student's dataset*. The first step consisted of applying linear regression to the *student's dataset*. We referred to the variables resulting from this process as the *regression attributes*. (See Figure 5.2)

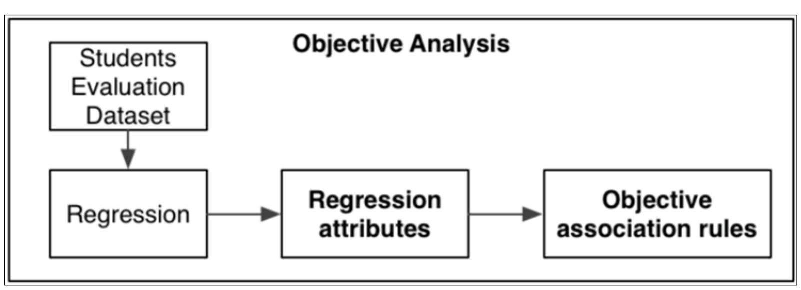

Figure 5.2 Objective analysis

After applying linear regression, we analysed some elements: the independence of the observations using the Durbin-Watson statistic, the no presence of multicollinearity with the

Tolerance/VIF values, R^2, and the R^2 adjusted; this last one described the model with the obtained *regression attributes*. Durbin Watson statistic detected the presence of autocorrelation; when the value was below 1.5, it represented a positive correlation of the variable and if it was above 2.5, the variables were negatively correlated. Moreover, *multicollinearity* appeared when two independent variables are correlated, where the tolerance = 1- R^2 and the VIF = 1/tolerance; we preferred a VIF lower than 10. *Adjusted R^2* and the *coefficient of determination R^2* expressed the model using these group of variables, in other words, how well the data fit the statistical model; the coefficient of determination has a range between 0 and 1. This statistic gave the goodness of the fit of a model, indicating how much the linear regression line approximated to the real points of the data; if the R^2 was 1 it means that the regression line fit the data perfectly, if it was close to 0 it didn't.

After the analysis of these statistics, we fed Apriori with the *regression attributes.* We obtained association rules containing these attributes. We evaluated these association rules in section E as part of the *subjective analysis*.

5.3.2 Subjective analysis

We created a teachers' survey instrument to ask teacher the importance of the questions in the students' dataset. We constructed survey S1 that contained five sections. The aim was to obtain the teacher's opinion on these 23 students' dataset questions in different contexts and identify, in so doing, if there was a relation between students and teacher's opinion. We did this comparing the objective analysis and the subjective analysis.

Survey S1 instrument contained five sections: *Section A (question evaluation),* retrieved teacher personal opinion about each question in the survey; *Section B (evaluation per area),* teacher evaluated questions inside each of the 3 areas (Design, Learning promotion, Product and teaching materials & Education Management); S*ection C (low evaluation reaction),* teacher react to situations where questions present a low rated value; S*ection D_a and D_b (creation of open and context rules respectively)* teachers created *open rules* using variables from *Section A*, and *context rules* using variables from S*ection B*. Finally, in *Section E, (Analysis using knowledge categories),* teacher analysed the rules generated from the regression analysis on *student dataset* (objective analysis), and evaluated the results using 6 knowledge categories: *interesting, unexpected (in the left hand side, LHS; in the right hand side, RHS; in both sides, BS), useful, obvious, previous and irrelevant*. Figure 5.3 presents the six sections of this interview.

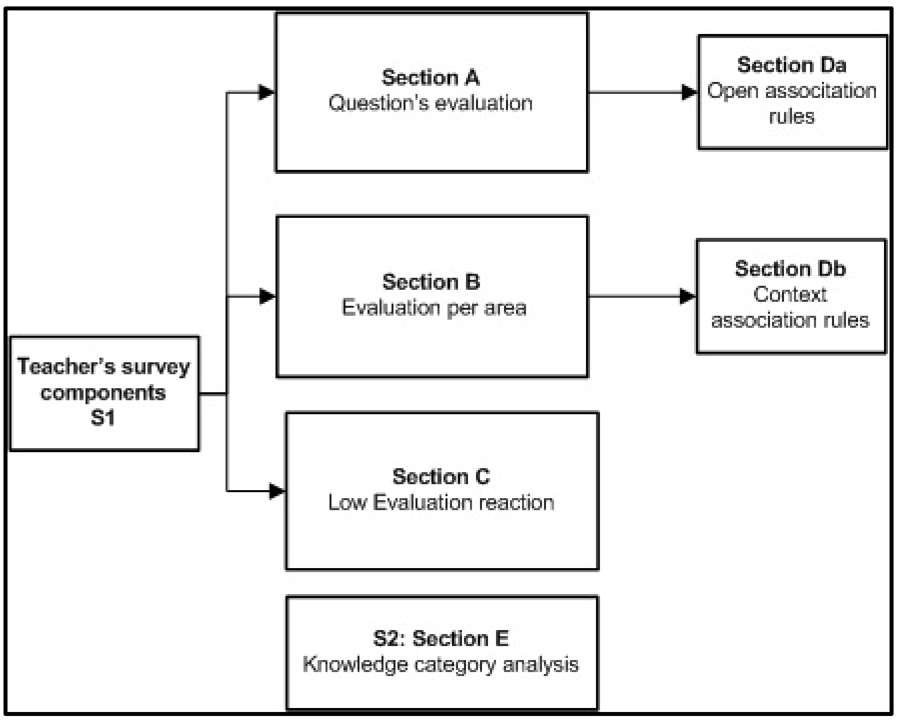

Figure 5.3 Teacher survey's components

In the following section we explained the content of each section of this survey instrument in more details.

Section A: Question evaluation

The goal of this section was to obtain the teacher's more precise understanding of each survey question. In this section, experienced teachers expressed their opinion about each survey question assigning a value between 0 (when he considered that a statement is irrelevant and 10 (when he considered that it is extremely important). This value depends only on the subjective thinking of the teachers and their experience about each survey question.

Section B: Evaluation per area

Section B evaluated the three areas that appear in the student survey (*design*, *learning promotion,* and *production of learning materials & education management).* Table 5.1 showed the number of questions per area and the maximum numbers of questions we asked the teacher to choose as the most relevant per area. For example, for the Design area comprising six questions, the teacher could only choose up to three questions with higher relevance.

Table 5-1 Questions per area and question chosen by teacher

Area in Student Survey	Questions per area	Max. # of questions chosen by the teacher
Design	6	3
Learning promotion of the course	13	6
Production of learning material & education management	4	2

As we could see, *section A* and *section B* evaluated the same survey questions but in different conditions. *Section A* did it in a general way, leaving freedom to the teacher to choose among all questions. In *section B,* the teacher focused on the evaluation of one specific area. He chose which questions were more interesting within this area, and left out other questions that he considered irrelevant or not useful. At this point, the teacher assigned values to each question in section A from the students' survey. These values changed when we asked the teacher to evaluate again the same question focusing on the area (Section B). Therefore, if a question had a value of 8 (very important) in *section A*, the teacher could assign another value to the same question in *section B (zero if teacher didn't choose it, ten if he chose it)*.

Section C: Low evaluation reaction

In this section, we confronted the teacher to the hypothetical situation of receiving a low evaluation in the students' survey questions. Teacher determined too, how hard it was to improve these results for next semester. The aim was to understand what the teacher knew about the questions' results from the survey and if he could interpret and improve the results he retrieved from this section. Thus, we asked the teacher to quantify what he could do with these results using four different categories. These value categories went from 10 to 1, where ten means lack of understanding of the results, and 1 means total comprehension of what to improve. Table 5.2 shows the categories for this *low evaluation reaction*:

Table 5-2 Categories of low evaluation reaction

Categories	**evaluations**
a) "I don't know how to use these results"	10
b) "Results are useful but not easy to change"	7
c) "Maybe the results do not depend only on the teacher"	4
d) "Results are useful and straightforward".	1

When the teacher identified low evaluated question as difficult to improve, he assigned a value of 10. This means, the teacher didn't know why this question was low evaluated. When the teacher knew why he received a low evaluation but he realized that fixing it would require more resources as time, money, knowledge, attitude, etc. he assigned a 7. In other words, he would need to improve something in class or in his teaching style. When the teacher knew that the low evaluation occurred due to causes that were above the course domain, he assigned a value of 4. For example, a cause could be the students' personal situation. Finally, if the teacher knew exactly where the problem was and what he needed to modify in his class, he assigned a value of 1. This means he could apply the changes in order to get a higher evaluation next time.

Section D_a: Creation of *open association rules* using variables from section A

In this section, we asked teachers to creat open association rules. Hereafter, we will refer to the survey questions as attributes. To construct the open association rules, we used the 23 attributes from the survey. We defined *open association rules* (henceforth *open rules*) as the rules constructed using attributes from all areas; these open rules were of the form a →b (where “a” and “b” were attributes or a group of attributes that belong to the attributes group X1 to X23. Thus, attributes used to create an *open rule* were called *open attributes*.

In section D_a, we invited the teacher to create up to five rules with the attributes that were evaluated high, in *section A,* and placed them in the consequent side of the association rule. Then, from the entire group of attributes, he chose two attributes and placed them in the antecedent side of the open rule. In other words, the teacher created association rules of 3-itemset (two attributes in the antecedent and one attribute in the consequent). We instructed teacher not to use in the antecedent the same attribute that he had placed in the consequent. The antecedent → consequent open rule construction denoted which attribute placed in the antecedent appeared at the same time with an attribute place in the consequent based on general considerations. The teacher could try different options of attributes in the antecedent for the selected consequent. Once teacher constructed each rule, he selected only one of the two attributes in the antecedent of each rule. The final rule had one attribute in the antecedent and one attribute in the consequent. The reason for eliminating one of the antecedents was the teacher weighted implicitly the relative importance of the attributes in the antecedent and chose only one. The remaining attributes, those in the antecedent or in the consequent of these rules, were called *open attributes*. Figure 5.4 shows Section D_a and Section D_b open and context attributes.

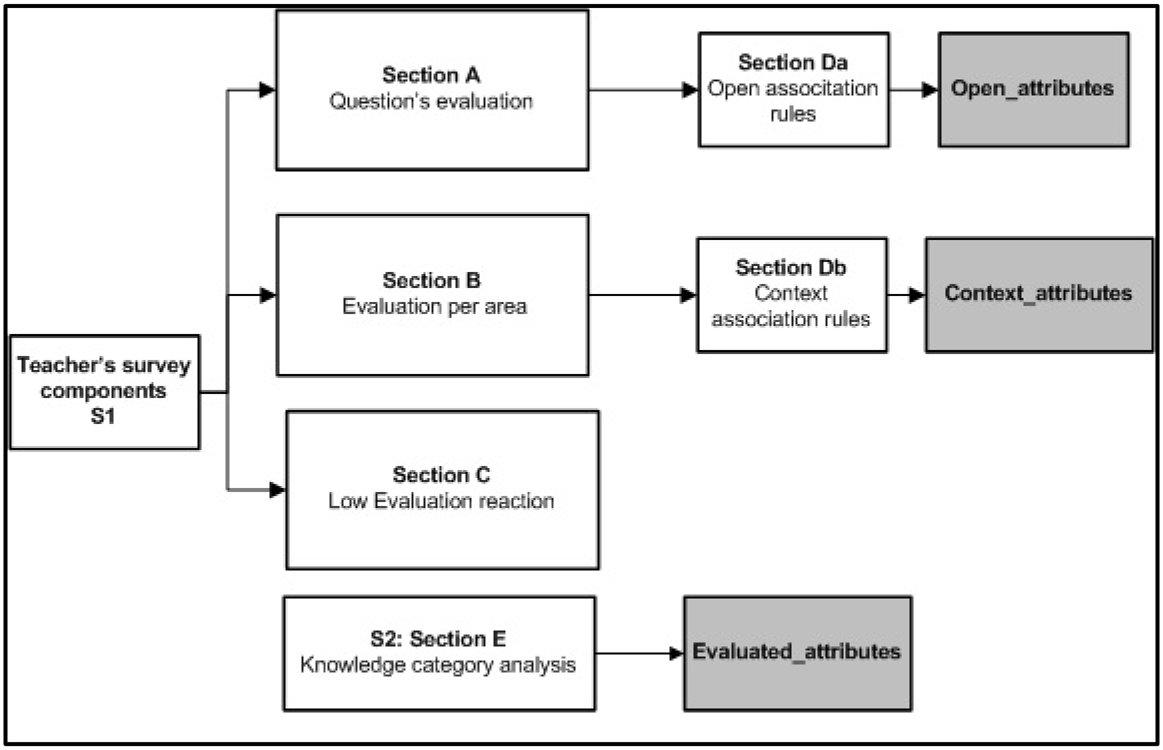

Figure 5.4 Open, context and evaluated attributes

Section Db: Creation of *contextual association rules* using variables from section B

We defined *Context association rules* (henceforth *context rules*) as the rules constructed using attributes that belong to specific areas. These context rules are of the form c → d, where c and d are an attribute or group of attributes that belong to the *design area* (X1 to X6); f → g , where f and g are an attribute or a group of attributes that belong to the *learning promotion of the course area* (X7 to X19), and r →t , where r and t are an attribute or a group of attributes that belong to the *production of learning materials & education management area* (X20 to X23). We called *context attribute* each attribute that was part of a *context rule*.

In section D_b, the teacher created *context rules* for each area, using only the attributes selected in *section B*. In *Section B - evaluation per area,* the teacher selected a group of variables per area that were the most interesting for him. The teacher placed one of the selected attributes in the consequent side of the rule and two attributes in the antecedent side of the rule. For example, if the teacher selected in the *design area* attributes X2, X3, and X5 as the most important attributes for that area, he could choose one of these three as consequent of a *context rule*, while the two other attributes for the antecedent were chosen from the complete set of attributes of the *design area (*X1, X2, X3, X4, X5 and X6). We instructed the teacher not to use in the antecedent the same attribute that he had already placed in the consequent. These rules were 3-itemset (two items in the antecedent and one item in the consequent). The antecedent → consequent context rule construction denoted which attribute placed in the antecedent appeared at the same time with an attribute place in the consequent based in the area consideration.

The teacher could try different combinations of attributes in the antecedent side of the rule for the consequent attribute selected. Next, the teacher chose only one of the two attributes in the antecedent. We used only rules with one attribute in the antecedent and one attribute in the consequent. Teacher repeated this step for the three areas. The reason for eliminating one of the attributes in the antecedent side was to encourage the teacher to weight implicitly the relative importance of the attributes that were in the antecedent and chose only one. In summary, during this step, the teacher weighted the attributes and identified which were more important per area, which ones were more interesting for him as results (consequent), and which ones were more interesting/useful to appear in the antecedent of the same rule.

Section E: Questionnaire for Teacher's evaluation rules

In this section, we constructed survey S2. We used rules with higher support, confidence and lift values. We obtained these rules from the linear regression analysis applied to students' dataset and the Apriori algorithm. Then, the teacher analysed the association rules obtained from the regression attributes using knowledge categories. For this purpose, the teacher classified these association rules using six knowledge categories: interesting, unexpected, useful, obvious, previous and irrelevant.

After the teacher categorized the association rules, we used the rules classified as *interesting, unexpected* and *useful* and obtained the frequency of the attributes in the antecedent and in the consequent. The frequency results from the attributes represented teacher's preferred attributes. Then, we chose the attributes with higher frequencies in the antecedents and in the consequents.

5.3.3 Predictive and Objective Analysis

We are now at the third step of the methodology, consisting in deriving the *predictive* and *objective attributes* to form a dataset to feed the Apriori algorithm. We used two methods for the construction of this dataset. We could use each of the methods to find predictive attributes and objective attributes; both methods focused on either side of the association rule. The first method looked for similar attributes between the open and context rules among all teachers. This method considered all teachers' opinions. With this method, we found the high frequency similar attributes among all teachers. When no similar attributes appeared between teachers, we considered a second method where the analysis is done only on teacher's opinion. In this case, we considered only the open and context attributes that belonged to the teacher and obtained the similar attributes between open and context attributes (either in antecedent and consequent). All the antecedent's similar attributes obtained with any of the two methods are called predictive attributes. All the similar attributes that were in the consequent obtained with either of the two methods were called objective attributes.

We used the attributes in the antecedent as the *predictive attributes* and we used the group of attributes that appeared in the consequent in high frequency as the *objective attributes*. We included the *regression attributes* into the objective attributes (if they were not included in the objective attributes). These would be part of the dataset to feed Apriori. Figure 5.5 presents the procedure in details.

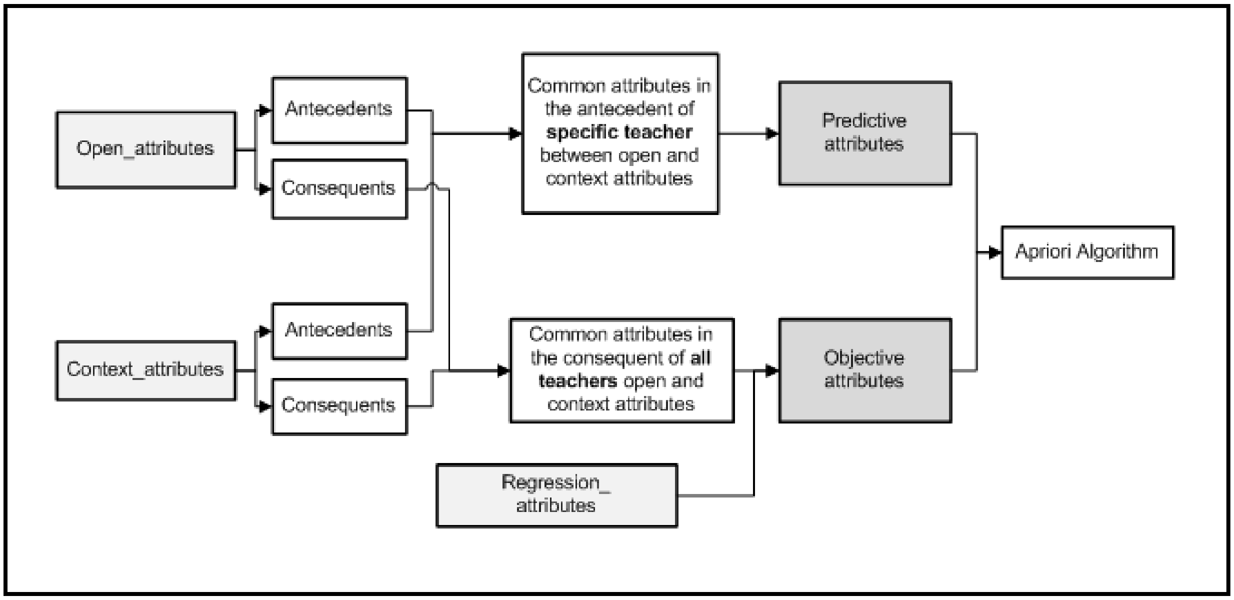

Figure 5.5 Predictive and objective constructed dataset

We used R package "arules" and constructed a script for searching specific attributes in the antecedent and specific attributes in the consequent of the association rules. This package works with Apriori Algorithm. The *predictive* and *objective attribute* were specific attributes obtained during the analysis of attributes among teachers. This allowed us to obtain association rules that had the predictive attributes in the antecedent and the objective attributes in the consequent. We obtained association rules and evaluated them using support, confidence and lift. We applied the utility measure to the association rules with high support confidence and lift values.

5.3.4 Utility analysis

The fourth step consisted in applying the utility formula to the association rules with high support, confidence and lift. To do this, we constructed the utility formula.

The utility formula considered two components: the transaction utility and the external utility. The transactional utility considered the interest reflected by the student. The external utility was represented by the teacher interest and the effort the teacher needed to apply in order to improve an attribute evaluated lower (Yao, Hamilton et al. 2003). This was represented with the following function.

$$F\ (T,\ E) \tag{5.1}$$

Thus, T represented the transaction utility and E was the external utility. These two components represented the Utility function.

The transactional utility T was a function of the students' evaluation (Is), while the external utility E was a function of the teachers' evaluation (It) and the teachers' improving ability (Eft). In other words:

$$F(T(Is),E(It,\ Eft)) \tag{5.2}$$

We explained how to obtain the Is, It and Eft for the utility formula in the next section. It is important to emphasize that each student evaluation had twenty-three questions, from now on attributes and the questions used for evaluation a Likert scale from 0 to 10. A "mark" was the value a student assigned to an attribute in the survey question respect to teacher's evaluation. First, we specified some definitions:

Definition 1: Weight of a variable (Wxi) was the total sum of all the marks for an attribute in a dataset.

$$\sum_{i=1}^{|DB|} Wxi \tag{5.3}$$

Definition 2: Actually evaluation weight (Aew) was the relation between the sum of all the marks from a specific attribute inside the database and the sum of all the attributes' weight inside the evaluation

$$Aew = \sum_{i=1}^{|DB|} Wxi \Big/ \sum_{j=1}^{m} \sum_{i=1}^{n} Wxixj \tag{5.4}$$

Definition 3: Excellent evaluation weight (Eew) was a constant, and the maximum value an attribute can reach if it has the higher mark. It depended on the amount of transactions in the data set and the amount of attributes to be evaluated.

$$Eew = \sum_{i=1}^{|DB|} Wxi\, h \Big/ \sum_{j=1}^{m} \sum_{i=1}^{n} Wxih\, xjh \tag{5.5}$$

Definition 4: Actually – Excellent factor (AEF) was the relation between the actually evaluation weight and the excellent evaluation weight (Eew).

$$AEF = \frac{Aew}{Eew} \tag{5.6}$$

Definition 5: Variable frequency was the frequency of a specific variable appearing in the database. If a transaction has zeros or no answers, then the variable frequency is lower than one.

Definition 6: Interest of students (Is) was the first component of the utility formula. It represented the utility the students' evaluations gave in the form of weights to the teacher performance. This was a component that only depends on students' perspective. It was represented by the mean of the attribute Xi, the Actually-Excellent factor (AEF) and the Variable frequency. The interest of the student (*Is*) had values between 0 and 10 where cero was the minimum value and ten was the maximum.

$$Is = Mean\ Xi * AEF * Variable.frequency \tag{5.7}$$

Definition 7: Question evaluation (QE); we obtained QE from section A in the subjective analysis. Qe represented the evaluation teacher gave to every question that was in the questionnaire.

Definition 8: Evaluation per area (EA); we obtained EA from section B in the subjective analysis. It was the evaluation teacher gave to every question but inside a context

Definition 9: Frequency of antecedent's attributes in the evaluated rules (FAe_r), we found FAe_r in *section E* where teacher analysed the rules coming from the objective analysis using six knowledge categories. We selected only those rules classified as *interesting*, *unexpected* or *useful*; the attributes from these rules were called evaluated attributes; we obtained the frequency of the attributes place in the antecedent of the *evaluated rules*.

Definition 10: Frequency of consequent's attributes in the evaluated rules(FCe_r), we found FCe_r in *section E* where teacher analysed the rules coming from the objective analysis using six knowledge categories. We selected only those rules classified as *interesting*, *unexpected* or *useful*; we called the attributes from these rules, *evaluated rules* and obtained the frequency of the attributes place in the consequent of the *evaluated rules.*

Definition 11: General Question evaluation (GQE) was the mean of all the teachers *General evaluation (GE)* of an attribute. From section A, we obtained a mean of all the teachers' evaluations per attribute.

Definition 12: Interest of teacher (It) was the second component of the utility formula. This component represented the teachers' perspective, his insights and experience about attributes' questionnaires. To calculate this factor we used question evaluation (QE), evaluation per area (EA), the frequency of the attributes in the antecedent of *evaluated rules, (FAe_r)*, frequency of the attributes in the consequent of the *evaluated rules (FCe_r)* and the general mean questionnaire evaluation (GQE). The interest of the teacher had values between 0 and 10.

$$It = \frac{QE + EA + FAe\_r + FCe\_r + GQE}{5} \tag{5.8}$$

Definition 13: Recuperation effort (Re) is the effort teacher put in an attribute to improve it; we determined the recuperation effort based on the knowledge teacher had about the attribute and the complexity of the actions. Using *Section C* of the survey, the teacher classified the usefulness and actionability of the answers. The ranges went from very useful and straightforward to not useful and not straightforward. Depending on these categories, we defined our actual knowledge of the attributes. Figure 5.6 determines the recuperation effort (Re)

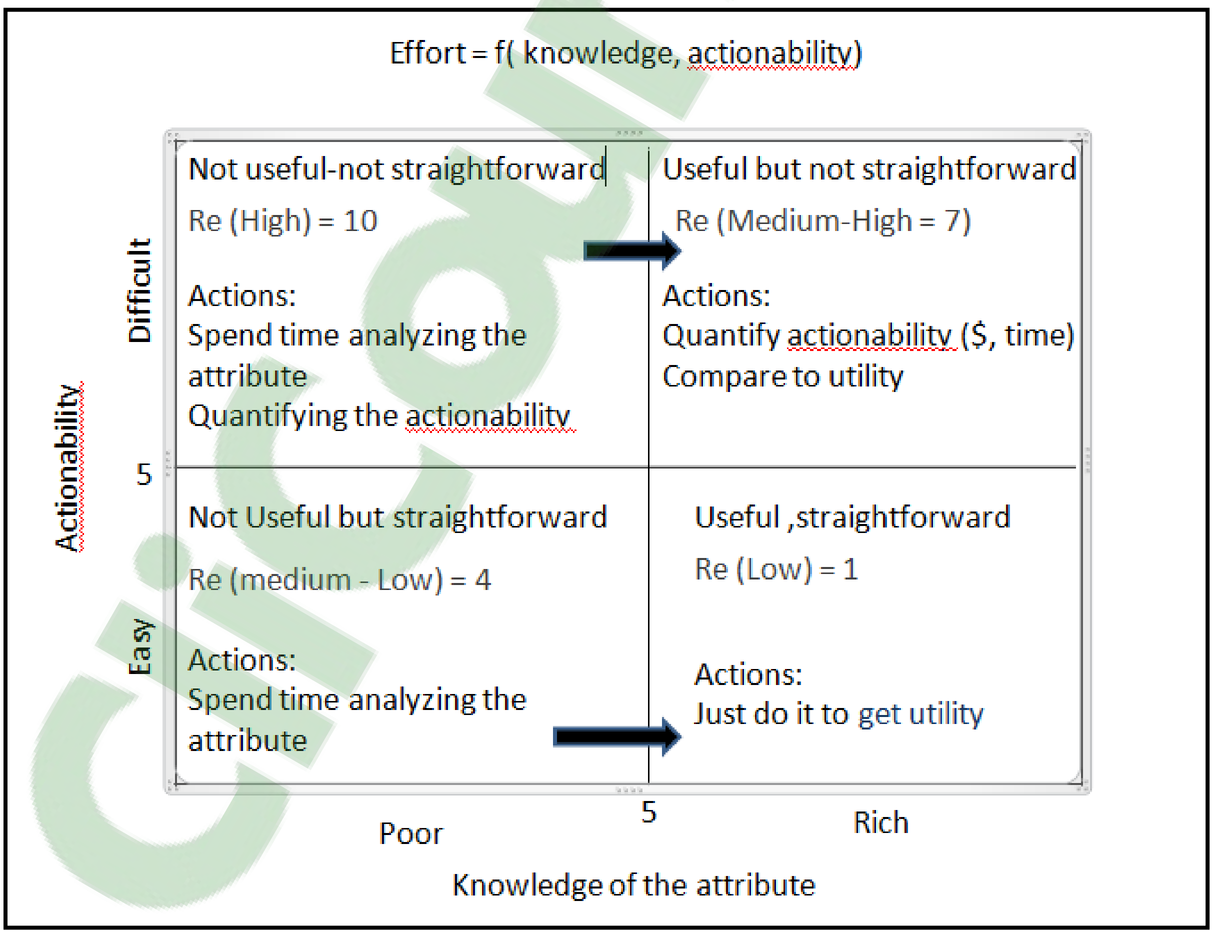

Figure 5.6 Effort function (Actionability vs. Knowledge)

Definition 14: Actual Knowledge (AcK), was how much knowledge a professor had about an attribute. If the attribute had a very high Recuperation Effort, the Actual knowledge (AcK) it had about the attribute, was very low, because teacher knew very little about this attribute. The actual knowledge was in the ranges of 0 to 9. We obtained the AcK from the following formula:

$$AcK = 10 - Re \tag{5.9}$$

Definition 15: Effort (Eft), was the ability to improve a specific attribute which was related to the actual knowledge of the attribute and the relation between teacher (It) auto-evaluation and students (Is) evaluation. Eft measured the ability of a teacher to improve an attribute. We presented Eft as a relation of efforts between teacher's efforts and students' importance. The relation was as follows:.

$$Eft = \frac{AcK * It}{Is} \tag{5.10}$$

Definition 16: Utility (U) was a profit measure that evaluated the importance of an attribute in terms of three elements: Interest of the students (*Is*), Interest of teachers (*It*) and the effort(*Eft*).

$$U = Eft * (Is + It) \tag{5.11}$$

We represented the same formula using the actual knowledge (*Ack*), teacher Interest(*It*) and student Interest(*Is*):

$$U = AcK * It + \left[\frac{Ack * \ It^2}{Is}\right] \tag{5.12}$$

5.4 Illustration of the methodology

To illustrate the methodology presented in the preceding section, we used the student evaluation dataset concerning two teachers, T1 and T7. We chose these two teachers because they have widely different evaluations

5.4.1 Objective analysis

For the objective analysis, we applied linear regression to *students' dataset* from two different teachers: teacher T1 and teacher T7; during the analysis of teacher T1, we obtained a regression with an R^2 of 1 and a Durbin Watson value of 2.03; during the analysis of teacher T7, the R^2 value was 0.95 and the Durbin Watson value was 1.92. The variables obtained from the linear regression were the *regression attributes.* Table 5.3 presents the *regression attributes* for T1 and T7.

Table 5-3 Regression attributes T1, T7

Teacher	**Regression attributes**
T1	X4, X9, X11, X19, X21,X23
T7	X5, X13, X18

5.4.2 Subjective analysis

We tabulated results from *Section A, B* and *C* from teacher's survey. In Table 5.4, we presented section A**,** where teachers evaluated the importance of attributes ranking them from 0 to 10. T1 to T9 represented nine different teachers and X1 to X23 represented the 23 attributes; we used only two teachers. Each T_n represented a different teacher's opinion about a Xn attribute. We obtained the General mean between the nine surveyed teachers for each attribute in the survey. *GQE* represented the General mean of the questions evaluation per attribute between teachers; i.e. attribute X2 has a *GQE* or mean = 9,11. We presented individual teachers' evaluations in Table 5.4

Table 5-4 Question's evaluation, section A

Attributes	T1	T2	T3	T4	T5	T6	T7	T8	T9	GQE
X1	8	6	6	6	10	8	8	8	6	7.33
X2	8	10	10	10	10	10	8	10	6	9.11
X3	6	8	8	8	10	8	8	10	10	8.44
X4	10	8	10	6	10	8	10	10	8	8.89
X5	8	8	10	10	10	8	8	10	6	8.67

We analysed the evaluation per area only for T1 and T7, *section B*. Table 5.5 presents the rankings for teachers` attribute inside the context analysis. Comparing tables 5.4 and 5.5, for teacher T1, we see attribute X4 had a rank of 10 in section A (Table 5.4) but it had a 0 in the evaluation per area, section B (Table 5.5).

Table 5-5 Evaluation per area, section B

Area	Question	T1	T7
Design Area	X1	10	10
	X2	0	10
	X3	10	10
	X4	0	0
	X5	10	0
	X6	0	0
Learning Promotion Area	X7	0	10
	X8	10	10
	X9	0	0
	X10	10	10
	X11	0	0
	X12	10	0
	X13	0	0
	X14	10	10
	X15	0	0
	X16	0	10
	X17	10	0
	X18	10	10
	X19	10	10
Production of learning materials & education management area	X20	0	0
	X21	10	0
	X22	0	0
	X23	10	0

We analysed the *low evaluation reaction*, *section C*. We used four numbers to evaluate each attributes; these numbers are on *Table 5.2, Categories of low evaluation reaction*. A value of ten meant teacher found this attribute very difficult to fix and understand. A value of one meant an attribute could be very easy to fix and teacher knew a lot about it. Table 5.6 presents some of the attributes and teachers' opinions; i.e. teacher T1 evaluated attribute X1 as very hard to improve if he was low graded, but teacher T7 evaluated the same attribute as a very easy attribute to fix.

Table 5-6 Low evaluation reaction section C

Question	**T1**	**T7**
X1	10	1
X2	4	4
X3	1	4
X4	10	4
X5	4	7
X...	...	...
X23	10	4

5.4.3 Open association rules creation

In *section Da*, teacher constructed *open rules* choosing high evaluated attributes from *section A*. He placed the high evaluated attribute on the consequent side of the rule. Then, he chose and placed two attributes in the antecedent that he suggested might appear together with that selected consequent.

In Table 5.7, *Rule A*, the teacher placed attribute X2 in the consequent and suggested that X2 could appeared if attributes X7 and X8 were in the rule' antecedent too. Teacher constructed these rules using all the attributes.

Table 5-7 Open rules T7

Rule	Antecedent Attribute 1	Antecedent Attribute 2	Consequent
Rule A	X7	X8	X2
Rule B	X9	X15	X3
Rule C	X3	X19	X5
Rule D	X2	X14	X10
Rule E	X10	X13	X17

Table 5-8 Final open rules T7

Rule	Antecedent Attribute	Consequent Attribute
Rule A	X7	X2
Rule B	X15	X3
Rule C	X19	X5
Rule D	X14	X10
Rule E	X10	X17

In Table 5.8, we present teacher's prioritization between the two attributes in the antecedent. Table 5.8 shows the final *open rules*.

5.4.4 Context association rules creation

Teacher constructed *context rules* in *section Db*. Teacher used the attributes selected in *section B*. Here is an example in the *design area*. This area only had six variables, from X1 to X6. Teacher selected attributes X5 and X6 as the most important in *section Db*. Teacher placed attributes in the consequent side of the Table 5.9; then from the whole group of attributes from that area (X1 to X6) he chose attributes to put them in the antecedent side and complemented the rule.

Table 5-9 Context rule construction, T7

Rule	Antecedent 1	Antecedent 2	Consequent
Rule A	X2	X4	X5
Rule B	X2	X3	X5
Rule C	X1	X3	X6

The next step was to prioritize between antecedents. In Table 5.9, to construct Rule A, teacher chose between X2 and X4. He chose X4 as the most possible attribute appearing with X5 in the consequent. Table 4.10 shows the final prioritized attribute per rule.

Table 5-10 Final context rules, T7

Rule	Antecedent	Consequent
Rule A	X4	X5
Rule B	X3	X5
Rule C	X3	X6

We repeated this process to all the three areas of the survey and obtained *context rules* for each area.

5.4.5 Evaluated association rules creation

In the objective analysis, we created rules using *regression attributes* and selected them using support = 0.7, confidence = 0.9 and lift >0.9. Teacher evaluated these rules using six knowledge categories. We obtained the frequency of attributes in the antecedent from the rules before and after the teacher classified them with knowledge categories and obtained the relation between these two values. Teacher T1 had 44 rules to be analysed at the beginning. After the analysis, he evaluated as interesting, useful or unexpected 24 rules. Table 5.11 presents the attributes' frequencies for T1; Table 5.12 presents the attributes' frequencies for T7.

Table 5-11 Evaluated attributes T1

Attribute	X4	X9	X11	X19	X21	X23
Frequency	0.34	0.89	0.89	0.449	0.614	0.36

Table 5-12 Evaluated attributes T7

Attribute	X4	X5	X13	X18
Frequency	0.024	0.44	0.11	0.11

5.5 Predictive and Objective analysis

To obtain the predictive attributes, we applied the methods from section 5.3.

5.5.1 Predictive and Objective Analysis: Predictive Attributes

We used the first method to compare open and context attributes. We searched for common attributes in the antecedent of teachers' open rules and in the antecedent of teachers' context rules. We found that the only common attribute place in the antecedent between nine professors was X20. Since the first method didn't retrieve enough information, we applied the second method to find similar attributes focusing on the teacher particular analysis. Therefore, we identified *open*, *context*, *evaluated* and *regression attributes* for teacher T1 and T7 for this particular analysis. The result was the *predictive attributes.*

In Table 5.13, we present *open attributes* and *context attributes* for T1. The high frequency attributes between o*pen attributes* and *context attributes* are X14, X16 and X17. On the other hand, we compared the *evaluated attributes* with the *regression attributes* on Table 5.14. The *regression attributes* were X4, X9, X11, X19, X21 and X23; and the *evaluated attributes* more selected during the knowledge categorization were X9, X11 and X21. We integrated the subjective selection (*open* and *context attributes*) to the objective selection (*regression attributes*). We saved the *evaluated attributes* for future analysis. We focused on these attributes *(evaluated attributes*) when we evaluated all the rules with the utility measure. The predictive attributes were X4, X9, X11, X14, X16, X17, X19, X21 and X23.

Table 5-13 Open /context attributes comparison for teacher T1

Attributes	X4	X9	X11	X14	X16	X17	X19	X21	X23
Contextual attribute				1	1	1			
Open attributes	1		1	1	1	1			
	0.5	0	0.5	1	1	1	0	0	0

Table 5-14 Evaluated/regression attributes comparison for teacher T1

Attributes	X4	X9	X11	X14	X16	X17	X19	X21	X23
Regression attributes	1	1	1				1	1	1
Evaluated attributes		1	1					1	
	0.5	1	1	0	0	0	0.5	1	0.5

For professor T7, Table 5.15 presents the *open attributes and context attributes*, where X11, X13 and X21 were common attributes. We integrated to the subjective selection the *regression attributes*; in Table 5.16, attributes X5, X13 and X18. Then, for *predictive attributes,* we used X5, X11, X13, X18 and X21. The teacher preferred the evaluated Attribute X5. We focused on it when we applied the utility measure.

Table 5-15 Open and context attributes comparison or teacher T7

Attributes	X4	X5	X11	X13	X18	X21
Open attributes	1		1	1	1	1
Context attributes			1	1		1
	0.5	0	1	1	0.5	1

Table 5-16 Evaluated and regression attributes comparison for teacher T7

Attributes	X4	X5	X11	X13	X18	X21
Regression attributes		1		1	1	
Evaluated attributes		1				
	0.	1	0	0.5	0.5	0

After this step, we needed to find the *objective attributes* that appeared in the consequent side of the rule.

5.5.2 Objective attributes

We placed objective attributes in the consequent side of the rule. We made the same search for common attributes in the consequent side for the group of the nine teachers. We found more common attributes between the *open* and *context rules*. We presented this analysis below.

We applied the first method to find similar attributes placed in the consequent of the *open* and *context rules*. The attributes with the higher frequencies were X12, X14, X19, and X23.We considered high frequency if more than 40% of the teachers had the same opinion. We called these attributes *open attributes*. Table 5.17 shows the attributes frequencies.

Table 5-17 Frequency of attributes in open rules (consequent)

	X12	**X13**	**X14**	**X15**	**X16**	**X17**	**X18**	**X19**	**X20**	**X21**	**X22**	**X23**
S%	0.4	0.0	0.4	0.0	0.1	0.2	0.1	0.7	0.0	0.0	0.3	0.6

We obtained the frequency of the attributes in the consequent of the *context rules* created by teachers. The attributes with the higher frequencies were X9, X12, X14, X18 and X19. These attributes were called *context attributes*. We present frequencies in Table 5.18.

Table 5-18 Frequency of attributes in context rules (consequent)

	X9	X10	X11	X12	X13	X14	X15	X16	X17	X18	X19	X20	X21	X22	X23
S%	0.6	0.2	0.2	0.8	0.1	0.6	0.3	0.0	0.1	0.6	0.7	0.4	0.4	0.1	0.1

Three attributes were common among teachers in the consequent side of the *open* and *context rules,* X12, X14 and X19 in Table 5.17 and Table 5.18. We integrated the similar attributes from open attributes and the similar attributes from context attributes (consequent) and considered them as *objective attributes*. In this way, our final *objective attributes* for the dataset were the union of similar open attributes and similar context attributes: X9, X12, X14, X18, X19, and X23. Hence, we used these variables as the *objective attributes* for all the teachers and used them to obtain the rules using the Apriori algorithm.

Then, we applied the Apriori algorithm using the predictive and objective attributes; we obtained association rules with two itemsets in the left hand side (LHS) of the association rule for each of the objective attributes and one itemset in the right hand side (RHS) of the association rule. If some of the *predictive attributes* were in the *objective attributes too*, we recommended to leave them in the *predictive attributes* and eliminated them from the *objective attribute* for one reason; predictive attributes contained the regression attributes, the attributes that explain a percentage of the model of the teacher and that is part of the objective analysis. Objective attributes were obtained from a subjective analysis and present different opinions from different professors, because of these, we gave more weight to the regression attributes placed in the antecedent than in the consequent.

We obtained an evaluated group of rules using support, confidence and lift and apply the utility formula. From T1 we obtained 20 association rules with support 0.48, confidence 0.93 and lift > 1.6. From T7 we obtained 54 association rules with support 0.62, confidence 0.94 and lift > 1.09.

5.5.3 Evaluation with utility measure

From the previous three steps, 5.1, 5.2 and 5.3 we obtained *Open, context*, *regression* and *evaluated attributes*, for each teacher T1 and T7. Figure 5.19 shows all these attributes by teacher T1 and T7.

Table 5-19 Predictive and objective attributes T1 and T7

Attributes	**T1** **Student dataset**	**T7** **Student dataset**
Open attributes	X14,X16,X17	X11,X13,X21
Context attributes	X4,X11,X14,X16,X17	X4,X11,X13,X18,X21
Subjective= Open ∩ Context	X14,X16,X17	X11,X13,X21
Regression attributes	X4,X9,X11,X19,X21, X23	X5,X13,X18
Evaluated attributes	X9,X11,X21	X5
Predictive attributes = (Open ∩ Context)-Subjective ∪ Regression	X4,X9,X11,X14,X16, X17,X19,X21,X23	X5,X11,X13,X18,X21
Objective attributes	X9,X12,X14,X18,X19,X23	X9,X12,X14,X18,X19,X23
New dataset	X4,X9,X11,X12,X14, X16,X17,X18,X19,X21, X23	X5,X9,X11,X12,X13, X14,X18,X19,X21,X23

The new dataset for T1, for example, contained the predicted attributes; this is the common attributes between the open and the context attributes plus all the regression attributes found through the regression analysis. In this case X4, X9, X11, X14, X16, X17, X19, X21 and X23 are the predictive attributes; we found the objective attributes looking for common attributes among the nine teachers. We saw that the predictive attributes are common for these two teachers.

5.5.4 Utility rule evaluation using utility measure for T1 and T7

We fed the Apriori algorithm with the new dataset for T1 and T7 respectively. Apriori algorithm returned association rules. First, we searched for association rules with the predictive attributes in the antecedent, then, we searched for rules that presented objective attribute in the consequent. We analysed the rules using lift measure.

For T1, in the student dataset, we searched first for rules that contained predicted attributes. We found 20 rules with this characteristic and a lift value between 1.6 and 2.06, then we analysed the objective attributes, only objective attribute X12 was presented. We chose the association rules that accomplished these two conditions and that have higher lift values, in these case the lift values were between 1.9 and 2.06. These found rules for T1 were:

a) X9 then X12
b) X14 then X12
c) X9, X11 then X12
d) X9,X14 then X12
e) X9,X16 then X12
f) X11,X14 then X12
g) X11, X16 then X12
h) X14, X16 then X12

We repeated the same procedure for T7. We found 54 rules with predicted attributes with a lift value between 1.14 and 1.45. Then, we looked for the objective attributes inside the last group. We found 21 association rules with the predicted attributes and some objective attributes (X9, X12 and X19). These 21 association rules presented a lift value between 1.37 and 1.45. These are the rules:

a) X5, X11 then X9
b) X5, X13 then X9
c) X5, X18 then X9
d) X5, X21 then X9
e) X13, X11 then X9
f) X13, X21 then X9
g) X5, X11 then X12
h) X5, X13 then X12
l) X5, X11 then X19
m) X5, X13 then X19
n) X5 X18 then X19
o) X5, X21 then X19
p) X11 then X19
q) X11, X13 then X19
r) X11 X18 then X19
s) X11, X21 then X19

i) X5, X18 then X12
j) X11, X13 then X12
k) X13, X21 then X12
t) X13 X21 then X19
u) X18, X21 then X19

To obtain the utility, we used the equations (5.10) and (5.11). We calculated the actual knowledge of the variable *(AcK),* with equation (5.9), the teacher's interest *(It),* with equation (5.8), and the students' interest *(Is),* with equation (5.7).

First, we needed to calculate the mean weight of each attribute with formula (5.3), $\sum_{i=1}^{|DB|} Wxi$, and the sum of all the attributes' weights for the whole dataset, $\sum_{j=1}^{m}\sum_{i=1}^{n} Wxixj$. Then to calculate the *actually evaluation weight*, for each attribute, we used formula [5.4], $Aew = {\sum_{i=1}^{|DB|} Wxi} \Big/ {\sum_{j=1}^{m}\sum_{i=1}^{n} Wxixj}$. In Table 5.20, we have the student dataset where each Id_n represents a student' survey and X_n represents the survey' attribute. For attribute X1, the mean was 8.36, the sum of all students' weight for this attribute (X1) was 276, the sum of all the attributes' weight was 6138 and the *actually evaluation weight* for X1 was 0.0449 (see Table 5.21). This value was different for all the attributes because it depended on the total weight of each attribute.

Table 5-20 Weight calculation

	X1	X2	X3	X4	X5	...	X23	
Id1	8	8	8	8	8	...	8	
Id2	6	7	7	7	7	...	7	
Id3	3	4	4	6	4	...	5	
Id4	10	8	8	8	8	...	8	
Id5	10	9	9	0	10	...	8	
Id6	2	2	2	2	2	...	2	
Id7	9	9	9	9	9	...	9	

	X1	X2	X3	X4	X5	...	X23	
Id8	10	10	10	10	10	...	10	
...	...	...	...	...	...	...	...	
Id33	6	6	6	6	6	...	7	
Mean	8.36	8.09	8.12	7.73	7.85	...	8.12	
Weight	276	267	268	255	259	...	268	6138

Table 5-21Aew calculation

	Mean	$\sum_{i=1}^{\lvert DB \rvert} Wxi$	$\sum_{j=1}^{7} \sum_{i=1}^{5}$	**Aew**
X1	8.36	276.0	6138.0	0.045
X2	8.09	267.0	6138.0	0.043
X3	8.12	268.0	6138.0	0.044
X4	7.73	255.0	6138.0	0.042
X5	7.85	259.0	6138.0	0.042

To calculate the *Excellent Evaluation Weight*, *Eew,* we used formula (5.5). *Eew* was a constant that expressed the higher mark of an attribute compared to the total sum of all the attributes in the survey evaluated with the highest mark. We assumed the highest attribute's mark for a dataset with 33 surveys (Id33) and 23 attributes; therefore, we had an attribute weight of 330 for each attribute and a total weight of 7.590 for all the survey.

Is calculation. During the application of the utility formula, we obtained *Actually-Excellent Factor (Aef)* from the relation Aew/Eew for T1. This factor increased the value of the students' evaluation when teacher did better and decreased it when teacher did worse. Table 5.22 presents the calculations to obtain the *Aef* factor. In the same table, we included the *attribute mean* and the *variable.frequency*. In this case, because all the attributes were presented in all transactions, the *variable frequency* had a value of 1.

Table 5-22 Actually Excellent Factor (Aef)

	Mean	$\sum_{i=1}^{\lvert DB\rvert} Wxi$	$\sum_{i=1}^{\lvert DB\rvert} WxiT$	$\sum_{j=1}^{7} \sum_{i=1}^{5} WxixjT$	**Eew**	**Aef**	**Variable. Frequency**	**Is**
X1	8.36	276.0	330.0	7590.0	0.043	1.034	1.00	8.65
X2	8.09	267.0	330.0	7590.0	0.043	1.000	1.00	8.09
X3	8.12	268.0	330.0	7590.0	0.043	1.004	1.00	8.15
X4	7.73	255.0	330.0	7590.0	0.043	0.955	1.00	7.39
X5	7.85	259.0	330.0	7590.0	0.043	0.970	1.00	7.62

Looking at the *Aef* factor column, we saw that students perceived teacher gave more attention to attributes X1, X2 and X3, and attributes X4 and X5 had less attention in the students' evaluations.

It calculation. We calculated the It utility formula component. For this, we needed to calculate *QE, EA, FAe_r, FCe_r* and *GQE.* The definitions of each of these components were in section 5.3.4 (definitions 7 to 11)

QE was the value teacher gave to the attribute; *EA* was the evaluation per area for each attribute assigned by the teacher. To obtain the *FAe_r* and the *FCe_r,* from the evaluated rules for T1 in section E, we chose only the rules selected as unexpected, interesting and useful; we calculated their attributes' frequencies in the antecedent and in the consequent. In Table 5.23, we presented the attributes in the evaluated rules. GQE was the mean between all the teacher general evaluations of an attribute. T1 evaluated rules presented the following attributes X4, X9, X11, X19, X21 and X23:

Table 5-23 Frequency of antecedents and consequents in evaluation rules

	Antecedent frequency attributes	**Consequent frequency attributes**
X4	12/24.	6/24
X9	10/24.	10/24
X11	14/24.	2/24
X19	3/24	3/24
X21	11/24	5/24
X23	15/24	7/24

Then, we calculated the It component of the utility formula using the equation (5.8). Table 5.24 presents all the values to calculate the It component for the utility formula.

Table 5-24 Teacher interest calculations

	QE	**Re**	**FAe_r**	**FCe_r**	**GQE**	**It**
X1	8	10.0	0.0	0.0	7.3	5.1
X2	8	0.0	0.0	0.0	9.1	3.4
X3	6	0.0	0.0	0.0	8.4	4.9
X4	10	10.0	5.0	2.5	8.9	5.3
X5	8	0.0	0.0	0.0	8.7	5.3

Comparing the students' mean evaluation in Table 5.22 *(Mean)* and the importance teacher gave to the attributes' evaluation in Table 5.24 (*QE),* for attribute X1, students and teacher opinions were similar, but for attribute X4 teacher assigned a ten value ranking while students assigned it with a 7,73.

Comparing *It* and *QE* inside Table 5.24, teacher opinion (It) was focused on five different variables as can be seen; if a teacher considered important an attribute in a general way (QE) and inside a context (EA), the importance of the attribute rose; the same happened if the attributes appeared in the antecedent *(FAe_r)* and in consequent (FCe_r) of the evaluated rules. From student perspective *(Is)* and teacher perspective *(It),* we found that the global ranking teacher had for these attributes was very low, different from what students say about teacher.

Eft calculation. To calculate the Effort (Eft), we used equation (5.10); AcK was the quantified value of knowledge teacher has about the variable.

In Table 5.25, for attributes X1 and X4, the recuperation effort (RE) to improve was 10, which meant that these were very hard attributes. Therefore, the teacher didn't have enough *actual knowledge* (Ack) about it. This produced a multiplication by 0, meaning that this was not a strong variable at all for him to improve. But in the case of X2, X3 and X5, its *RE* was different. The *RE* to improve X3 was 1, and this means that teacher understood at least 90% of this attribute. The Actual knowledge of these attributes were equal or higher than 6. We calculated the *Eft* using the equation (5.10) for each attribute. Here from the five attributes, attribute X3 had a high utility value. The utility formula results for the first five attributes were:

Table 5-25 Utility Calculation T1

	RE	AcK	Eft	Is	It	Utility
X1	10	0.00	0.0	8.65	5.07	0.0
X2	4	6.00	2.54	8.09	3.42	29.22
X3	1	9.00	5.41	8.15	4.89	70.44
X4	10	0.00	0.00	7.39	5.28	0.0
X5	4	6.00	4.14	7.62	5.33	54.07

Finally, we evaluated the association rules for T1 and T7 using the utility values per attribute. The rule with higher utility for T1 was *X14 then X12, it* contained an open-context attribute; but it didn't contain evaluated attributes. Table 5.26 shows the association rules evaluated for T1

Table 5-26 T1 utility evaluation association rules

Students Dataset T1	
Rules	**Utility**
X9 then X12	74.99
X14 then X12	**86.72**
X9, X11 then X12	65.25
X9,X14 then X12	77.1
X9,X16 then X12	49.99
X11,X14 then X12	73.07
X11, X16 then X12	45.97
X14, X16 then X12	57.81

In the case of T7, the higher utility contained X5 (evaluated attribute), X13(open context attribute) and X19, as can be seen in Table 5.27

Table 5-27 Utility evaluation association rules

Students Dataset T7	
Rules	**Utility**
X5, X11 then X9	21.36
X5, X13 then X9	35.23
X5, X18 then X9	31.56
X5, X21 then X9	23.13
X13, X11 then X9	29.2
X13, X21 then X9	30.96
X5, X11 then X12	25.43
X5, X13 then X12	39.3
X5, X18 then X12	35.63
X11, X13 then X12	33.26
X13, X21 then X12	35.03

Students Dataset T7	
Rules	**Utility**
X5, X11 then X19	27.7
X5, X13 then X19	**41.56**
X5 X18 then X19	37.9
X5, X21 then X19	29.46
X11 then X19	23.9
X11, X13 then X19	35.53
X11 X18 then X19	31.86
X11, X21 then X19	23.43
X13 X21 then X19	37.3
X18, X21 then X19	33.63

5.6 Experiment with extreme values

5.6.1 Objective and Subjective analysis (step 1 and 2)

We generated a *synthetic positive dataset* that only contained excellent values, this means 7, 8 9 and 10. This positive synthetic dataset imitated students surveys ranked high. We made linear regression to it and evaluate R^2 = .645, Durbin-Watson value = 2.327 and VIF lower than 10.

We also generated a *synthetic negative dataset* that only contained regular values; this means from 0 to 6. Since the synthetic negative dataset had only values equal or below six, it is a synthetic dataset that imitated students' surveys ranked low. We made linear regression to it and evaluated the adjusted $R^{2=}$ 0.83, Durbin Watson value = 2.149 and VIF lower than 10.

We analysed these two synthetic dataset using the objective analysis and subjective analysis from teachers T1 and T7. It is important to emphasize that the open and context attributes didn't depend on the dataset. It was information teacher gave based on his knowledge; therefore, the results for the open and context attributes were the same for every dataset, students' dataset and synthetic datasets. Regression and evaluated attributes depended on the dataset. Therefore, we obtained new regression results from these new two synthetic datasets and obtained the predictive attributes. In this case, we used only the regression results. Table 5.28 and Table 5.29 present *open* and *context attributes* and the common attributes among these two teachers. Tables contain the *regression attributes* for each dataset. Note that the *evaluation attributes* were the attributes obtained after teacher's rule evaluation using the knowledge categories; because teacher didn't evaluate again rules generated with the synthetic dataset, we used the same evaluated attributes from the students' dataset.

Finally, the *objective attributes* were the same for all the datasets because they were the selection of the nine teachers.

Table 5-28 Predictive and Objective attribute construction T1

Attributes	**T1 Student dataset**	**Synthetic positive dataset**	**Synthetic Negative dataset**
Open attributes	X14,X16,X17		
Context attributes	X4,X11,X14,X16,X17		
Subjective= Open ∩ Context	X14,X16,X17		
Regression attributes	X4,X9,X11,X19, X21, X23	X9,X21,X22	X5,X17,X19
Evaluated attributes	X9,X11,X21		
Predictive attributes = (Open ∩ Context) ∪ Regression	X4,X9,X11,X14, X16, X17,X19,X21,X23	X9,X14,X16,X17,X21, X22	X5,X14,X16,X17,X19
Objective attributes	X9,X12,X14,X18,X19,X23		
New dataset	X4,X9,X11,X12,	X9,X12,X14,X16,X17,	X5,X9,X12,X14,X16,X17,

Attributes	T1 Student dataset	Synthetic positive dataset	Synthetic Negative dataset
	X14, X16,X17,X18,X19,X21, X23	X18,X19,X21,X23	X18,X19,X23

Table 5-29 Predictive and Objective attribute construction T7

Attributes	T7 Student dataset	Synthetic positive	Synthetic negative
Open attributes	X11,X13,X21		
Context attributes	X4,X11,X13,X18,X21		
Subjective= Open ∩ Context	X11,X13,X21		
Regression attributes	X5,X13,X18	X9,X22,X21	X5,X17,X19
Evaluated attributes	X5		
Predictive attributes = (Open ∩ Context) ∪ Regression	X5,X11,X13,X18,X21	X9,X11,X13,X21,X22	X5,X11,X13,X17,X19,X21
Objective attributes	X9,X12,X14,X18,X19,X23		
New dataset	X5,X9,X11,X12,X13, X14,X18,X19,X21,X23	X9,X11,X12,X13,X14, X18,X19,X21,X22,X23	X5,X9,X11,X12,X13,X14,X17, X18,X19, X21,X23

The new dataset is the input for the next association rule creation.

5.6.2 Generated Association rules and utility formula (step 3 and 4)

From the constructed datasets we obtain the association rules. We analyse them based on the predictive attributes and the objective attributes of each rule. We first analyse the predictive attributes focusing on each component: open & context, regression and evaluated attributes. Then, we analyse the predictive attributes with the objective attributes. We do the same thing for the synthetic positive dataset searching for rules with only open & context and regression attributes in the antecedent.
We found 23 rules with a lift value between 0.99 and 1.06, then we analysed the objective attributes and only the objective attribute X19 presented lift values between 1.02 and 1.06. These are the rules presented in Table 5.28 in the column *Synthetic positive.* Finally, *Synthetic negative* column in Table 5.28 presents the rules corresponding to the same analysis, with a lift value of 1.029.

Every rule presented for T1 has an open & context attribute and a regression attribute in the antecedent and the consequent is one of the objective attributes presented in a rule with a high lift. Additionally, the rules in the first column (student dataset) have evaluated attributes. Table 5.30 and 5.31 present the best association rules for T1 and T7 respectively.

Table 5-30 T1 rules for the three datasets

students Dataset T1	Synthetic positive	Synthetic negative
X9 then X12	X17 then X19	X16 thenX18
X14 then X12	X21 then X19	X17 then X18
X9, X11 then X12	X22 then X19	X19 then X18
X9,X14 then X12	X16, X17 then X19	X16 X17 then X18
X9,X16 then X12	X16, X21 then X19	X16 X19 then X18
X11,X14 then X12	X17, X21 then X19	X17 X19 then X18
X11, X16 then X12	X21 X22 then X19	
X14, X16 then X12		

We repeated the same process for T7 for the synthetic positive dataset. We searched for rules with only predictive attributes in the antecedent. We found 42 rules with lift values between 0.98 and 1.15. Then, we analysed the objective attributes within the rules with predictive attributes and we found objective attributes X12. X18 and X19 presenting lift values between 1.02 and 1.06; these were the rules presented in Table 5.29 in the column *Synthetic positive.* Finally, *Synthetic negative* column in Table 5.29 presents the rules corresponding to the same analysis, we obtained 79 rules with lift values between 1.02 and 1.06, and only rules with the objective attribute X12 had lift values of 1.06

Table 5-31 T7 rules for the three datasets

students Dataset T7	**Synthetic positive**	**Synthetic negative**
X5, X11 then X9	X9,X13 then X12	X13 then X12
X5, X13 then X9	X9,X21 then X12	X11 X13 then X12
X5, X18 then X9	X11then X18	X13, X17 then X12
X5, X21 then X9	X11,X13 then X18	X13 X19 then X12
X13, X11 then X9	X9 then X19	X13, X21 then X12
X13, X21 then X9	X11 X19	
X5, X11 then X12	X9, X11 then X19	
X5, X13 then X12	X9, X13 then X19	
X5, X18 then X12	X21 then X19	
X11, X13 then X12	X22 then X19	
X13, X21 then X12	X11 X13 then X19	
X5, X11 then X19	X11 X22 then X19	
X5, X13 then X19	X13, X22 then X19	
X5 X18 then X19	X21 X22 then X19	
X5, X21 then X19		
X11 then X19		
X11, X13 then X19		
X11 X18 then X19		

students Dataset T7	Synthetic positive	Synthetic negative
X11, X21 then X19		
X13 X21 then X19		
X18, X21 then X19		

Once all the rules were selected, we applied the utility measure to them and evaluated the results. Table 5.32 presents the utilities for T1 using the regular student dataset, the synthetic positive dataset and the synthetic negative dataset.

The *Is* (the interest of the student) was different for the three datasets. On the other hand, the *It* (interest of the teacher) for the synthetic datasets was the same for both synthetic but different for the students' dataset because the *It* for these datasets didn't consider the evaluated rules. The *Eft* for the synthetic negative dataset was very high; the effort to improve from a lower grade dataset was bigger than from the student dataset or from a positive synthetic dataset.

Variables with the higher utilities for T1 were X12, X14 and X19. We also had to consider some attributes that have 0 utility values. This happened because the recuperation effort for these attributes was very high, then these attributes were not considered as advantageous for the teacher. This was the case of attributes X1, X4, X7, X8, X16, X17, X21.

Table 5-32 Utilities T1

	Re	**AcK**	**Eft -**	**utility -**	**Eft**	**utility**	**Eft +**	**Utility +**
X1	10	0	0.00	0.00	0.00	0.00	0.00	0.00
X2	4	6	6.94	73.80	2.54	29.22	3.57	54.60
X3	1	9	16.06	204.19	5.41	70.44	7.62	135.44
X4	10	0	0.00	0.00	0.00	0.00	0.00	0.00

	Re	AcK	Eft -	utility -	Eft	utility	Eft +	Utility +
X5	4	6	14.04	178.18	4.14	54.07	5.88	105.61
X6	4	6	7.83	76.82	2.78	29.26	3.59	53.29
X7	10	0	0.00	0.00	0.00	0.00	0.00	0.00
X8	10	0	0.00	0.00	0.00	0.00	0.00	0.00
X9	1	9	12.19	116.16	4.76	57.86	5.29	78.34
X10	1	9	18.94	250.46	5.92	80.21	8.73	158.93
X11	4	6	7.89	80.26	3.54	45.79	3.72	56.15
X12	1	9	18.50	260.74	7.19	92.12	9.33	173.79
X13	4	6	6.77	59.60	2.05	22.55	3.05	42.22
X14	1	9	18.32	246.92	6.00	81.32	8.81	160.93
X15	1	9	10.70	105.08	3.41	39.70	5.15	75.46
X16	10	0	0.00	0.00	0.00	0.00	0.00	0.00
X17	10	0	0.00	0.00	0.00	0.00	0.00	0.00
X18	1	9	19.65	250.44	5.55	76.31	8.07	149.22
X19	1	9	23.00	310.50	7.58	104.80	8.96	174.29
X20	1	9	13.44	123.01	3.79	42.05	5.25	78.10
X21	10	0	0.00	0.00	0.00	0.00	0.00	0.00
X22	4	6	8.62	77.97	2.44	27.01	3.36	49.90
X23	4	6	15.29	190.81	5.32	81.61	5.37	101.95

In table 5.33, for teacher T7, we had the same situation for *Is* and *It*. *Is* was different for the three different datasets. *It* for the student dataset was different from the two synthetic datasets because the synthetics datasets didn't use evaluated rules, and the students' dataset used evaluated rules. The *AcK* for teacher T7 was more interesting than T1. T7 was more aware of the variable knowledge, which was demonstrated in the RE and the *AcK* columns from table 5.33

Comparing the *Effort* to improve in the synthetic negative dataset, the effort was higher than the effort in the student dataset. In this case, teacher recognized at least 30% of the attributes knowledge. Then, this teacher had different levels of attributes, those that were better known and those that were weakly known. Each attribute presented an utility.

Consequently, once each question had a utility, we applied the utility value to the association rules obtained from the different datasets (students' dataset, synthetic positive and synthetic negative).

Table 5.34 presents three types of rules: Rules obtained from the students dataset, rules obtained from the synthetic positive dataset and rules obtained from the negative dataset. Each rule was evaluated with the utility formula. We included the utilities for each of these rules.

Table 5-33 Utilities T7

	RE	**AcK**	**Eft -**	**Utility -**	**Eft**	**Utility**	**Eft +**	**Utility +**
X1	1	9	20.31	247.51	5.35	72.70	8.16	144.90
X2	4	6	10.39	148.16	3.71	52.66	5.75	106.21
X3	4	6	11.22	151.77	3.46	50.02	5.58	102.05
X4	4	6	8.10	88.78	4.62	61.34	4.34	65.10
X5	7	3	4.52	41.78	2.57	35.30	1.90	27.23

	RE	AcK	Eft -	Utility -	Eft	Utility	Eft +	Utility +
X6	4	6	7.76	76.42	2.45	28.17	3.69	53.83
X7	4	6	13.05	180.66	3.75	55.45	6.26	116.22
X8	4	6	14.50	198.93	3.75	56.75	6.14	117.80
X9	7	3	3.60	31.77	1.02	11.60	1.59	22.09
X10	4	6	13.44	187.22	4.07	58.16	6.43	119.73
X11	7	3	4.32	47.15	1.46	17.24	2.13	33.04
X12	4	6	6.05	58.00	2.24	23.81	3.26	44.57
X13	4	6	7.75	73.31	4.31	58.79	3.59	51.12
X14	7	3	6.41	91.33	2.26	30.62	3.24	60.58
X15	7	3	3.04	28.20	1.15	11.61	1.54	21.19
X16	7	3	7.26	101.04	1.98	29.42	3.03	59.36
X17	4	6	7.11	65.05	2.10	24.11	3.14	45.36
X18	7	3	7.13	95.30	2.87	47.78	2.92	55.72
X19	7	3	7.90	105.74	2.26	30.62	3.02	58.37
X20	4	6	9.34	84.06	2.62	28.34	3.57	52.48
X21	4	6	7.37	62.41	2.02	22.45	2.83	41.22
X22	4	6	7.83	64.53	2.35	23.37	2.99	41.94
X23	4	6	11.38	109.44	2.49	32.07	3.78	61.56

Table 5-34 Utilities from different datasets T1

Students Dataset T1		**Synthetic positive**		**Synthetic negative**	
Rules	**Utility**	**Rules**	**Utility**	**Rules**	**Utility**
X9 then X12	74.99	X17 then X19	87.145	X16 thenX18	125.22
X14 then X12	**86.72**	X21 then X19	87.145	X17 then X18	125.22
X9, X11 then X12	65.25	**X22 then X19**	**112.09**	**X19 then X18**	**280.47**
X9,X14 then X12	77.1	X16, X17 then X19	58.096	X16 X17 then X18	83.48
X9,X16 then X12	49.99	X16, X21 then X19	58.096	X16 X19 then X18	186.98
X11,X14 then X12	73.07	X17, X21 then X19	58.096	X17 X19 then X18	186.98
X11, X16 then X12	45.97	X21 X22 then X19	74.73		
X14, X16 then X12	57.81				

In the student dataset, the rule with higher utility was X14 then X12. It contained an open-context attribute; but the evaluated attributes were not part of this rule. In the synthetic positive dataset and in the synthetic negative dataset only regression attributes were inside the rules, no open and context attributes.

In the case of T7 rules, in Table 5.35, the student dataset presented a rule containing open-context attribute and evaluated attributes and had a utility of 41.56; the synthetic positive dataset presented a rule with two attributes (X13, X22), one corresponding to the open-context attribute and one corresponding to the regression attributes and this was the same case for the synthetic negative dataset (X13, X19).

Table 5-35 Utilities from different datasets T7

Students Dataset T7		Synthetic positive		Synthetic negative	
Rules	**Utility**	**Rules**	**Utility**	**Rules**	**Utility**
X5, X11 then X9	21.36	X9,X13 then X12	39.26	X13 then X12	65.65
X5, X13 then X9	35.23	X9,X21 then X12	35.96	X11 X13 then X12	59.48
X5, X18 then X9	31.56	X11then X18	44.38	X13, X17 then X12	65.45
X5, X21 then X9	23.13	X11,X13 then X18	46.62	**X13 X19 then X12**	**79.02**
X13, X11 then X9	29.2	X9 then X19	40.23	X13, X21 then X12	64.57
X13, X21 then X9	30.96	X11 X19	45.70		
X5, X11 then X12	25.43	X9, X11 then X19	37.83		
X5, X13 then X12	39.3	X9, X13 then X19	43.86		
X5, X18 then X12	35.63	X21 then X19	49.79		
X11, X13 then X12	33.26	X22 then X19	50.15		
X13, X21 then X12	35.03	X11 X13 then X19	47.51		
X5, X11 then X19	27.7	X11 X22 then X19	44.45		
X5, X13 then X19	**41.56**	**X13, X22 then X19**	**50.47**		
X5 X18 then X19	37.9	X21 X22 then X19	47.17		
X5, X21 then X19	29.46				

Students Dataset T7		Synthetic positive		Synthetic negative	
Rules	**Utility**	**Rules**	**Utility**	**Rules**	**Utility**
X11 then X19	23.9				
X11, X13 then X19	35.53				
X11 X18 then X19	31.86				
X11, X21 then X19	23.43				
X13 X21 then X19	37.3				
X18, X21 then X19	33.63				

5.7 Discussion

Students' and teacher evaluation could be tested using this methodology. The methodology allowed the mix of student objective evaluation, the teacher's subjective consideration and teachers' colleagues, to create a dataset that retrieve association rules more relevant for teacher and students.

During the subjective analysis, we reviewed some elements such as: the beliefs teacher had about what works or not, the teacher's reaction when the students' evaluation was very low and the common perception from a group of teachers about the same evaluation survey. This methodology helped to define quantitatively elements observed in a qualitative manner.

This methodology is applicable to any customer service evaluation that uses any kind of survey that retrieve objective information from a client, i.e. client service survey. The objective questionnaire using the Likert scale or any scale the evaluation of the objective questionnaire was possible with the use of the Likert scale but another scale could be use with the proper adaptation to the methodology; the application of linear regression to obtain the predictive and objective attributes helped to obtain the expected dataset. The subjective analysis was part of the methodology and defined the construction of a survey that helped to retrieve knowledge from people working in client service area.

In next section, we discuss some elements from the subjective analysis, the construction of the utility formula and the application of the utility formula.

5.7.1 Subjective analysis

The subjective analysis using open and context attributes showed attributes high evaluated within the general context and low evaluated within specific context; Attributes' importance changed in relation to teacher's opinion and the context area. This situation was common among teachers; none of the teachers assessed an attribute equally either in the general context or in the specific context.

During the evaluation of Section C, we noticed two cases about the low evaluated attributes: Teachers understood some attributes easily than others. When they needed to identify what to do with a low ranked attribute, sometimes they said they knew what to do, other teachers had no clue. This could represent the teacher needed to learn more about specific attributes.

We found a difference between objective and predictive attributes: for the *objective attributes*, we observed a group of attributes of common interest among teachers. On the other hand, for the *predictive attributes*; no common agreement existed among teachers. We thought that happened because each teacher had different beliefs about what attributes could predict specific results. In this case, objective attributes were the same because there was a common agreement about what are the attributes teachers want to reach as results.

Open and context attributes were part of the predictive attributes and they didn't depend on the students' dataset; therefore, Teacher's subjective fingerprint contained the open and context attributes

We can found the predicted attributes between open and context attributes plus the regression attributes. They were the subjective – objective teacher's fingerprint because they represented the attributes teacher worked with (subjective) and the attributes that represented teacher's model (objective); the students defined the latter.
When we compared the subjective attributes (open and context attributes) with the *regression attributes,* we found some matches. Therefore, there is a relation between teacher auto evaluation and students' perceptions.

Finally, we obtained rules from Apriori algorithm using the regression attributes. When we evaluated these rules using the knowledge categories, we noticed that almost half of the rules disappeared because they didn't belong to any of the three categories of unexpected, useful and interesting. Then, the knowledge categorization helped to differentiate rules.

5.7.2 Utility formula

The utility formula had three components: Student interest (Is), Teacher interest (It) and effort (Eft). *Is* and *It* presented students and teacher's perspective. When we analyzed these two perspectives, we found three big differences. First, attributes that students evaluated as important were not important for the teacher and vice versa. Secondly, the AEF factor reflected the amount of interest teacher applied to specific attributes during the course. Additionally, this indicator expressed where the teacher put more effort and where he did not put enough effort. Finally, *It* was very low compared to the *Is* value.

One possible reason was that the teacher was tougher with himself when he evaluated his attributes. Another possible reason is the appearance of specific attributes in the association rules the teacher evaluated. If the attribute didn't appear in the association rule, then this attribute was lower in the final utility value per attribute.

5.7.3 Application of the utility formula

By using the utility formula, we proposed a way to measure the utility in teachers' association patterns. This utility considered teachers` perception and students' perception and the opinion from teachers' colleagues. Additionally, the measure provided with the effort variable. The effort variable helped improve the attribute when the attribute utility was very low.

We evaluated three types of datasets for a teacher: a real dataset, a positive dataset with high ranked values and a lower dataset with lower ranked values. Each dataset presented different regression values; in other words, they presented three different models, the first defined by the students' dataset, and the other two created synthetically. The real data set provided a real situation with real attributes values, and the other two datasets showed the extremes models with high ranked values or low ranked values. The utility obtained from the datasets for each attribute showed three possibilities.

The low ranked dataset showed very high utility values per attribute, because it depends on the students' dataset. If the students` dataset had low values, the effort to improve was higher that a normal dataset. The formula showed that if teacher had low evaluations, he needs to work a lot to improve. The high ranked dataset showed utility values higher that the students dataset but lower than the low ranked dataset. The reason is the same, in this case the students' dataset had high evaluations then the utility was lower. We compared both datasets with the real students' dataset; these two always have higher utility values, because the datasets present extreme values. Additionally, the use of evaluated rules into the utility formula helps to eliminate attributes not important to the teacher.

The basic idea of the utility formula was that only attributes that the teacher understood well were improvable. Rules obtained in this way focus on what teacher should work on with more possibilities to improve because these attributes represent strengths for him.

This also meant that attributes that have utility of cero value, didn't have any value on the rules that contained them, in other words, rules with low utility attributes were not interesting for the teacher. The *It* value could have a value of cero. This depends on the amount of knowledge the teacher had about the attribute. If the *Eft* was cero, the utility for this attribute was cero. The knowledge was so low that it was not a good option for the teacher to improve. If the *Re* is too high the attribute also disappears from the rule.

The interest of the teacher (It), the interest of the student (*Is*) and the effort to improve (*Eft*) showed other interesting results. In the case of *It*, teachers expressed their real knowledge through these variables (*RE* and *AcK* variables);

The dataset influence on the values of the *Eft* factor because although the teachers interest *(It)* was constant, different datasets (the students' dataset and the synthetic datasets) produced different values for the *Eft* factor.

The interest of the students (*Is*) was focused on how much the teacher did, and how much he was supposed to do, based on the evaluation of the students. Here, we saw that teacher gave more attention to specific attributes.

The synthetic datasets for these teachers showed rules with specific combinations of attributes. The synthetic positive dataset and the synthetic negative dataset always presented one open-context attribute with one regression attribute in case of teacher T7. Additionally, for this teacher in the student dataset, he presented rules that included the evaluated attribute which was presented in the regression attributes.

The case was different for teacher T1, the highest utility rule presented an attribute from the open-context attributes and their synthetic datasets presented attributes that correspond to the regression attributes. In case of teacher T1, this could happen because this teacher had a lot of *Re* and *AcK* variables in cero values that means that this teacher didn't have enough knowledge about his best variables and these did not match students' opinion.

5.8 Conclusion and future work

During this research, we constructed a utility measure based on a created dataset containing predictive and objective attributes. We evaluated association rules retrieved from the students' dataset, using this measure. Very few rules remained with high utility. The high utility corresponded to the attributes with more teachers' knowledge. Teacher could use either the high utility pattern or the intermediate utility pattern. The first one to keep doing well during class, the latter to start improving new attributes with low utility.

As future work the inclusion of additional variables in the utility formula as "the grade and the excellence of the course" or "the teaching hours in a term (number of courses, number of groups)" should give another interesting perspective.

CHAPTER 6

ARTICLE III: TOPIC MODELING ANALYSIS OF TEACHERS INTERVIEWS: TOWARDS A COMPREHENSIVE VIEW OF EVALUATED TOPICS IN THE CLASSROOM

Otilia María ALEJANDRO MOLINA

Software Engineering and IT, École de Technologie Supérieure (ÉTS),
1100, rue Notre-Dame Ouest, Montreal-Canada. H3C 1K3

Computing Faculty, Escuela Superior Politécnica del Litoral (ESPOL),
Km 30.5 Via Perimetral. Guayaquil-Ecuador

Sylvie Ratté
Software Engineering and IT, École de Technologie Supérieure (ÉTS),
1100, rue Notre-Dame Ouest, Montreal-Canada. H3C 1K3

Paper submited to Journal Elsevier Computer & Education on February 14, 2016

Abstract

Teachers' evaluation results regularly provide a score for each question in the survey. No explanation is available on how to improve low scored survey questions. No answers readily offer an explanation on the low values even if teacher took into consideration all the questions mentioned in the survey. Some examples of these questions included *making class content understandable, increase motivation for research, and produce reflection of topic.* Teacher's evaluation results also do not consider whether he spent a lot of time and effort on activities to improve students' knowledge not included in the survey questions. We applied an interview to teachers asking about the practices they performed during classes that allowed them to obtain good evaluations in surveys. We are interested in discovering the hot topics teacher talked about during the interview and identifying if there was a relation with the topics evaluated by students in the survey. We performed our research in four steps: First, we used an unsupervised method, topic modeling, to create the categories involving these

interviews; second, using these categories, we created a dictionary. Third, we compared the interviews themes to the dictionary and classified interview's themes based on the resulting topics; finally, we used two human categorizers to classify the interviews manually and compared these results from the unsupervised method with the human categorization using kappa measurements. The results showed that human categorization and machine categorization had an interesting value of coincidence above 0 .41; the survey students used to evaluate the course only covers 50% of the activities done by the teacher. The other 50% is not included in the survey. This finding suggests improvement to the survey is necessary to evaluate teachers.

Keywords: Topic Modeling, Dictionary, Cosine similarity, Text Mining; Teacher Interview; Academic Categorization.

6.1 Introduction

The analysis of interviews and surveys using text mining tools is a wide research area. Most of the topics dealt with finding knowledge from surveys and interviews applied in different situations like post war military procedures, people's behaviour or students' learning. Ramesh created summaries from the interviews related to military situations that let them learnt and changed specific procedures in the battlefield supplies (Ramesh Sharda 2009). Minami et al. used surveys to understand the relation between learning styles and grades (Minami and Ohura 2013). Fuller et al. analysed deception or lies in criminal people statements through text mining (Fuller, Biros et al. 2011). These researches used a variety of techniques like association rules, decision trees, topic modeling and text mining to analyse interviews. Few researches focused on interviews applied to teachers to understand their insights and compare teachers' point of view with students' insights.

In the Latin American University where this study took place, the University has a survey that evaluates teacher's development. The survey evaluates three areas: Design, Learning Promotion, and Production of teaching materials & Management of education of the course.

The survey raises students’ awareness of the activities the teacher should accomplish during a course. Hence, the teacher prepares his materials every semester. He constructs the design, materials and administration of the course. However, misunderstandings could occur when students evaluate teachers considering only topics in the survey and do not consider others aspects in this evaluation tool.

Incidentally, we identified topics that teacher spoke about during classes that did not appear in the survey. For this purpose, we applied an interview to six teachers to get them to talk about the activities they did to draw students' attention to their classes. First, we set the conversation and let teachers speak freely about the related activities. Then, we applied topic modeling to the interviews to find topics and constructed a dictionary based on the students’ survey questions and related them to the newly discovered topics. After that, we divided the interviews into themes. Later, we classified each theme from the interviews using the topics found in the model classification. Then, we asked human experts to do the same classification. Finally, we evaluated these two classifications using the kappa measurement and compared the model classification with the experts’ classification, to test the accuracy of the model.

Thus, we found similarities and differences between what teachers said and what students evaluated. The students’ evaluations use questions that evaluate specific areas of the course, but the teachers provided key additional information regarding the activities of the survey. We identified new areas to classify interview themes that may suggest the need to improve the current survey.

6.2 Literature review

6.2.1 Probabilistic topic modeling

Probabilistic topic modeling are algorithms that work with statistical methods to identify the most relevant theme inside a group of documents (Blei 2012). Another name for them is mixed membership models in automatic content analysis methods (Grimmer and Stewart

2013). The topic found using topic modeling algorithm is a distribution of words over a fixed vocabulary (Blei 2012), where the topic is represented with the group of distributed words from the vocabulary along the documents. One statistical method topic modeling work with is LDA or Latent Dirichlet Allocation; the basic concept behind LDA is the use of joint and conditional distributions. This method is very useful when there is a group of documents with different themes or there is no categorization (Grimmer and Stewart 2013).
LDAvis (Sievert and Shirley 2014) is a web-based interactive visualization tool of topics models based on LDA. This visualization tool shows topics in two forms; the first one shows the topic referred inside the whole group of topics and the second one focuses in the words inside the specific topic. Additionally, this tool worked with a "lambda" value that identified the weight of the probability of a word in a topic; this lambda worked like a numeric boundary where only the most relevant words above lambda represented the topic, eliminating all those that are not relevant words. The identification of the proper "lambda" value in LDAvis identified specific words that label the topic. The literature review suggests a lambda value around 0.3 or 30% to identify the words that are more relevant inside one topic, but they are not relevant in the group of topics (Sievert and Shirley 2014). The prevalence or extension of the topic is another measure considered in LDAvis. The size of the circles that enclose a topic expresses the extension of the topic.

Three R libraries are useful in topic modeling analysis, the Natural language processing library (Kipper and Ruutmann), the LDA library and the LDAvis library.

6.2.2 Dictionary

The construction of dictionaries for different tasks is a common practice; e.g., in sentiment analysis (Jamoussi and Ameur 2013), and in the analysis of political opinions (Grimmer and Stewart 2013). These practices used the construction of a corpus to define the dictionary.
Gensim Python library (Řehůřek 2015) is a tool used to construct dictionaries. With this tool we converted documents to vectors and represented into the Vector Space Model (Deerwester, Dumais et al. 1990). This library, Gensim, uses LSA, Latent Semantic Analysis.

LSA reveals the semantic relation between words (Kuralenok and Nekrest'yanov 2000). We used LSI-LSA and actual corpora to train the model. LSI is an indexing technique that organizes words based on their semantic relationships. When using the Gensim library, it is necessary to select a corpora format. This corpora format serialized the corpus to the vector space.

6.2.3 Cohen’s Kappa statistic for measuring agreement

Cohen’s Kappa statistic is the standard measure for computing the inter-rater reliability coefficient, or the reliability between the opinion of two raters or human categorizers. Kappa statistics compares the classification of a subject between different human evaluators using different categories. It uses a contingency table and includes the observed and expected frequencies on the diagonal of the square contingency table. In Viera et al., they worked with two categories and two evaluators. (Viera and Garrett 2005). Table 6.1 shows the columns and rows from a square contingency table for two evaluators and more than two categories.

Table 6-1 Contingency table

Evaluator 1						
		Category c1	Category c2	...	Category cm	Total
Evaluator 2	Category c1	**a**	b	...	e	T2c1
	Category c2	f	**g**	...	k	T2c2
	...	...	...	**...**	...	
	Category cm	t	p	...	**o**	T2cm
Total		T1c1	T1c2		T1cm	n

The observed proportional agreement between two or more observers is:

$$Po = \frac{1}{n}\sum_{i=1}^{m} a + g + \cdots + o \qquad (6.1)$$

Where a, g,...o are the number of categories correctly classified, then divided by the total amount of observations (n). The expected agreement is:

$$Pe = \sum_{i=1}^{m}\left(\left(\frac{\text{T1c1}}{\text{n}}\right)\ \left(\frac{\text{T2c1}}{\text{n}}\right)\right) + \left(\frac{\text{T1c2}}{\text{n}}\right)\left(\frac{\text{T2c2}}{\text{n}}\right) + \cdots + \left(\frac{\text{T1cm}}{\text{n}}\right)\left(\frac{\text{T2cm}}{\text{n}}\right) \tag{6.2}$$

For example, T1c1 is the total amount of classifications for the evaluator 1 in the Category c1. (Where evaluator 1, in category c1, coincides with evaluator 2 in all his categories) and T2c2 is the total amount of classifications for the evaluator 2 in the Category c1 (where evaluator 2, in category c1, coincides with evaluator 1 in all his categories). Finally, the Kappa equation is:

$$\hat{k} = \frac{P0 - Pe}{1 - Pe} \tag{6.3}$$

In other words, Po is the summation of all the diagonal values (the correct classifications) divided by the total number of observations. Pe is the summation of the total classifications of the column and the total classification of the row of the same category divided by the square of the total number of observations.

We interpreted Kappa agreement using the following ranges (Viera and Garrett 2005), if kappa measure is less than cero, it means negative numbers, there is no chance of agreement; if the kappa measure is between 0.01 and 0.20 then there is a slight agreement between evaluators; if the kappa measure is between 0.41 and 0.60 there is a moderate agreement between evaluators; if the kappa measure is between 0.61 and 0.80 there is a substantial agreement; finally if the range is between 0.81 and 0.99 there is an almost perfect agreement.

6.3 Materials

To accomplish the objective of this research, we applied the four steps mentioned before: First, the application of the topic modeling technique to teachers' interviews, second the construction of a definition for each of the survey`s twenty-three questions, to build the academic dictionary; third the integration of the question`s definition to a specific topic that best suited them and finally the classification made by the model and by the experts`. We evaluated the agreement between model and experts using the kappa measure.

Next, we present the students survey questions as the base of our dictionary and the interviews used for the topic modeling analysis.

6.3.1 Survey question descriptions

At the end of each course, each student filled out a survey. This survey has 24 questions that evaluated teachers' teaching practices.

Twenty-three questions evaluated teacher in three areas: design, learning promotion and production & teaching materials and education management. Students evaluated the questions using the Likert Scale. Students graded the questions in the survey from 0 to 10. Question 24 represented an overall evaluation of the course. We presented survey questions per area and a short description in parentheses. This short description helped in the construction of the dictionary.

Table 6-2 Survey questions names

Area	Survey Questions and short descriptions
Design	X1. Uses audiovisual help to support the content of the class. *(Audiovisual)* X2 Fulfills the program proposed at the beginning of course. *(Fulfill program)* X3. Evaluates student participation periodically in class *(Participation evaluation)* X4 Evaluations fit the themes developed in class. *(Evaluation fit themes)* X5 Provides clear instructions for learning assessment like tests, quizzes, presentations, simulations, dramatic representation, role playing, etc. *(Assessment instructions)* X6 Motivates students to do additional research *(Motivation)*

Area	Survey Questions and short descriptions
Learning promotion	X7 Explains the course schedule at the beginning of the course. *(Course schedule)* X8. Explains class policies at the beginning of the course. *(Class policies)* X9. Encourages active student participation in class. *(Participation encourage)* X10. Summarizes key ideas discussed before moving to a new unit or topic. *(Summarization)* X11. Establishes relationships between new concepts and those already known whenever possible. *(Relation between concepts)* X12. Motivates learning of the course material. *(Motivate learning material)* X13. Is willing to answer questions and offer advice within and outside of the classroom. *(Advise)* X14. Promotes reflection on topics covered. *(Reflection)* X15. Maintains fluid communication with students. *(Communication)* X16. Is respectful towards students. *(Respect)* X17. Responds to questions in class about subjects related to the field. *(Answer questions)* X18. Delivers class content in an organized way. *(Class content organize)* X19. Develops class content in an understandable way. *(Class content understandable)*
Production and teaching materials & Education management	X20. Prepares instructional, bibliographic or other resources to facilitate learning. *(Material preparation)* X21. Frequently uses schemes and graphics to suport his/her explanations. *(Graphics and schemes)* X22. Provides the results of the assessments on time. *(Assessment results)* X23. Attends classes on time. *(Punctuality)*
General evaluation	X24. Considering all the features, choose a score between 1 and 10 to evaluate teacher's overall performance

6.3.2 Interviews

The interview consisted of a set of questions formulated to the teacher to capture his thoughts and activities during class; specifically, how he performed the different tasks described in the survey. Interview's questions referred to how teacher encouraged the research practice in class, the study of the material, how he made his classes more interesting and more. We interviewed the teacher in person or via internet, using Skype. We did 6 interviews. We digitalized all interviews and used them in the topic modeling step; we divided each interview as themes. Each theme was in the form of a complete sentence or group of sentences. Experts classified these ideas in the classification step.

6.4 Methods

The topic modeling and the dictionary generation consisted of four steps. The first step was to define the optimal number of topics generated from the group of documents. In the second step, we constructed the dictionary to describe students' survey questions. In the third step, using the dictionary and the topic classification, we created the topic - dictionary. Finally, we used Cohen Kappa statistics for measuring agreement between model's classifications that uses LSI against the experts' classification. Figure 6.1 shows the four different steps.

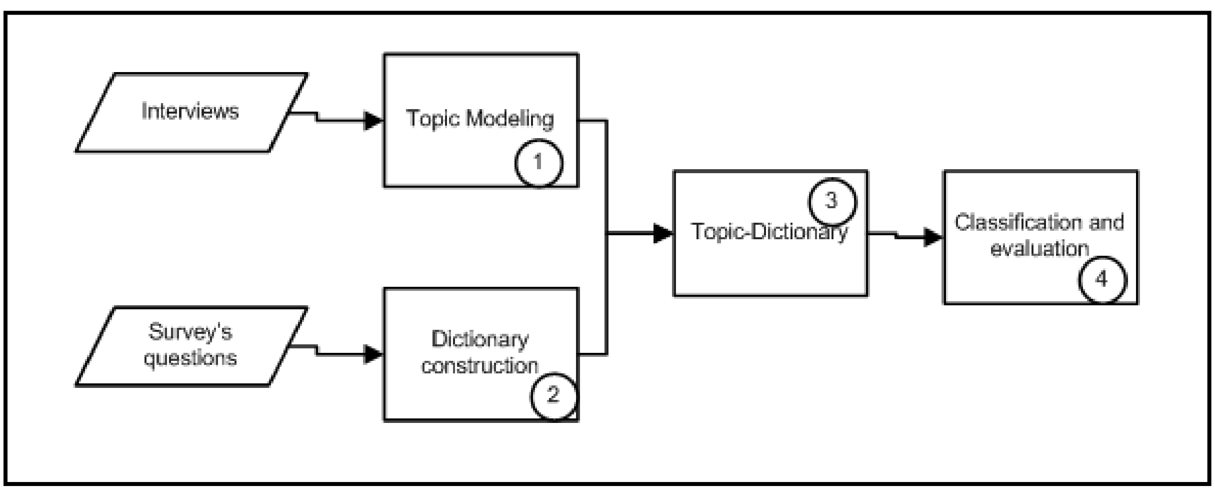

Figure 6.1 Topic modeling-dictionary construction and classification steps

6.4.1 Topic modeling

We used LDAvis application using a “lambda” value of 0.3. We applied topic modeling to the interviews. We made several tests to choose the best representation of the topics through the topic’s words. We started doing 3 through 20 topics. We found that the best representation used 6 topics. The lambda value allowed seeing more related terms among the topics. Figure 6.2 shows the topic modeling procedure.

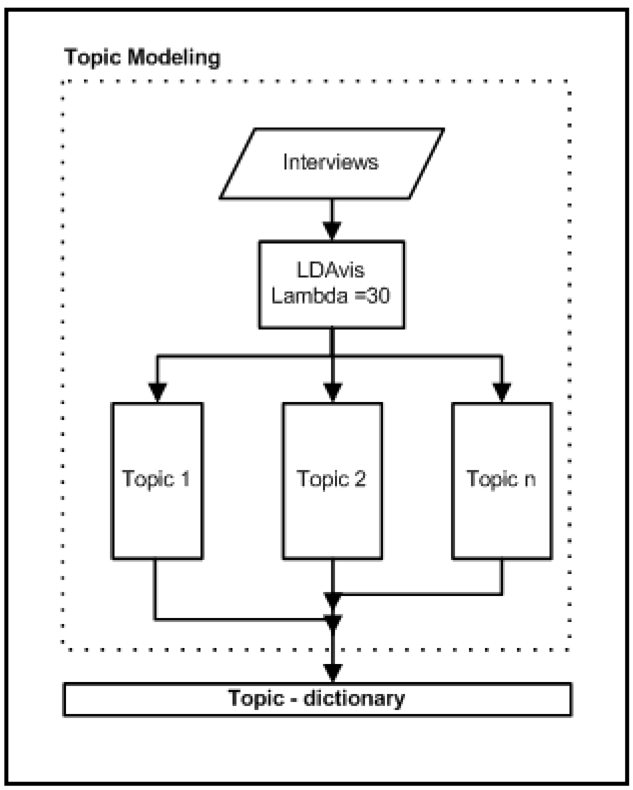

Figure 6.2 Topic modeling

6.4.2 Dictionary construction.

We used the 23 survey’s questions to construct the dictionary. Each question represented a characteristic student had to evaluate from the teacher. We hand-built each survey’s question description using educational articles and dictionary concepts. The description included the "what" a survey question represented and the "how" the teacher fulfilled the survey' question in class. For example, question 15 in table 6.2, expressed communication; we included the concept of communication in its description, we also included how people communicate using body language, friendly spoken language, and others. We did not include places (*where*) nor times (*when*) in this analysis because we wanted to focus only on the analysis of the action and possibilities of actions. Figure 6.3 shows the 23 students' survey questions as input and as output each survey question descriptions.

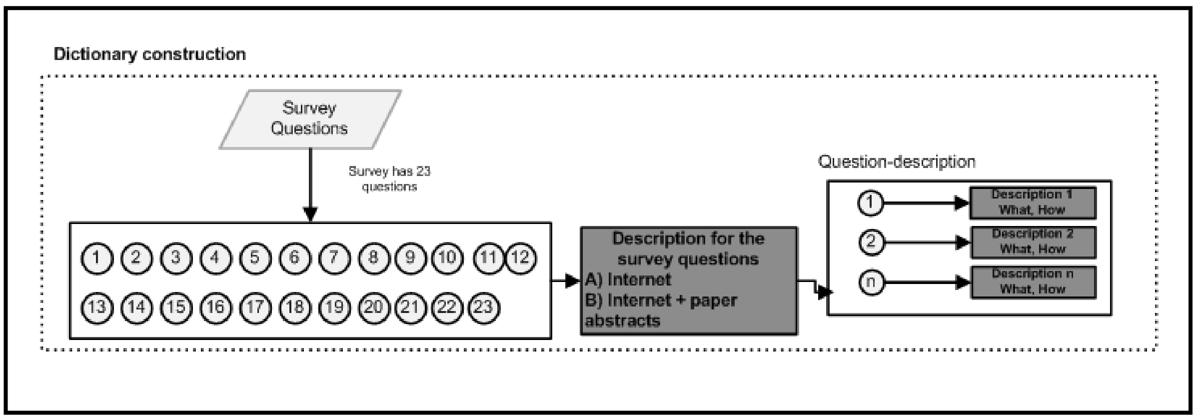

Figure 6.3 Dictionary construction

We distributed the survey' questions into the six topics obtained in section 6.4.1. We noticed two situations: In the first one, some survey questions didn't have a topic to relate to; in the second one, some topics could be empty. The first situation occurred when we couldn't assign any survey question to any topic. Survey questions evaluated topics that we didn't find in the interviews; to solve this, we created a new topic which included these survey questions. In the second situation, the teacher talked about topics that didn't have a representation through any survey question. In this case, we hand-built the description of this topic. Figure 6.4 shows topic modeling and dictionary construction sections.

Finally, we represented each topic with a group of survey' questions. We repeated this process for each of the found topics. Our dictionary contained the topics we obtained from topic modeling (we add the one we created) and their corresponding descriptions.

6.4.3 Classification and evaluation between expert opinion and the automatic classification

After constructing the topic description, we classified the interviews into these topics using the classification model. For testing, we asked experts to classify manually the interviews themes

Model Classification

We made the model classification using Vector Space Model (VSM) to represent text documents. Vector Space Model is the baseline model used to compare text documents as vectors without using complexes approaches as Natural Language Processing. In the VSM, we placed the topics descriptions, converted them to vectors and compared them to each of the interview's themes. If the cosine value between the topic description and the interview' theme was close to 1, then the theme and the topic were very similar.

To do this, we constructed a script in Python to transform the dictionary into the VSM and to classify the interviews' themes. We used the dictionary class from Gensim in Python. To make the classification, we used LSI. The model classified each interview's theme into one or more topic categories.

Expert classification

We separated all interviews' themes for the two experts. We asked them to make a manual classification for each theme. Figure 6.5 presents expert classifications. Experts analyzed each theme per interview and found one or more topic that fulfilled it.

Each theme could represent more than one topic. In Figure 6.5 theme B and G have 2 possible topics.

We compared the model classification with the experts' classification using the Kappa measurement. The complete process is presented in Figure 6.5.

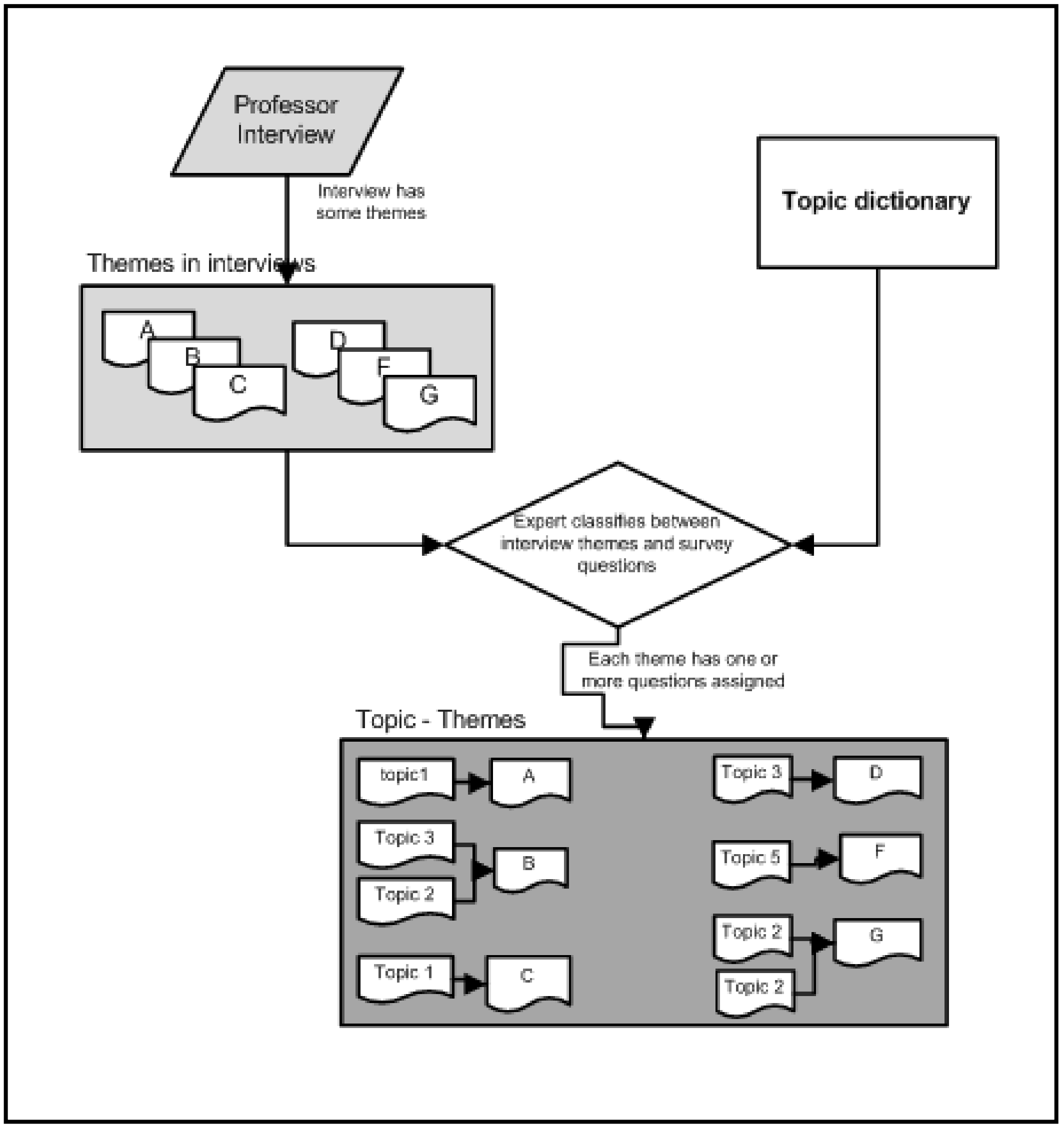

Figure 6.4 Expert classification process

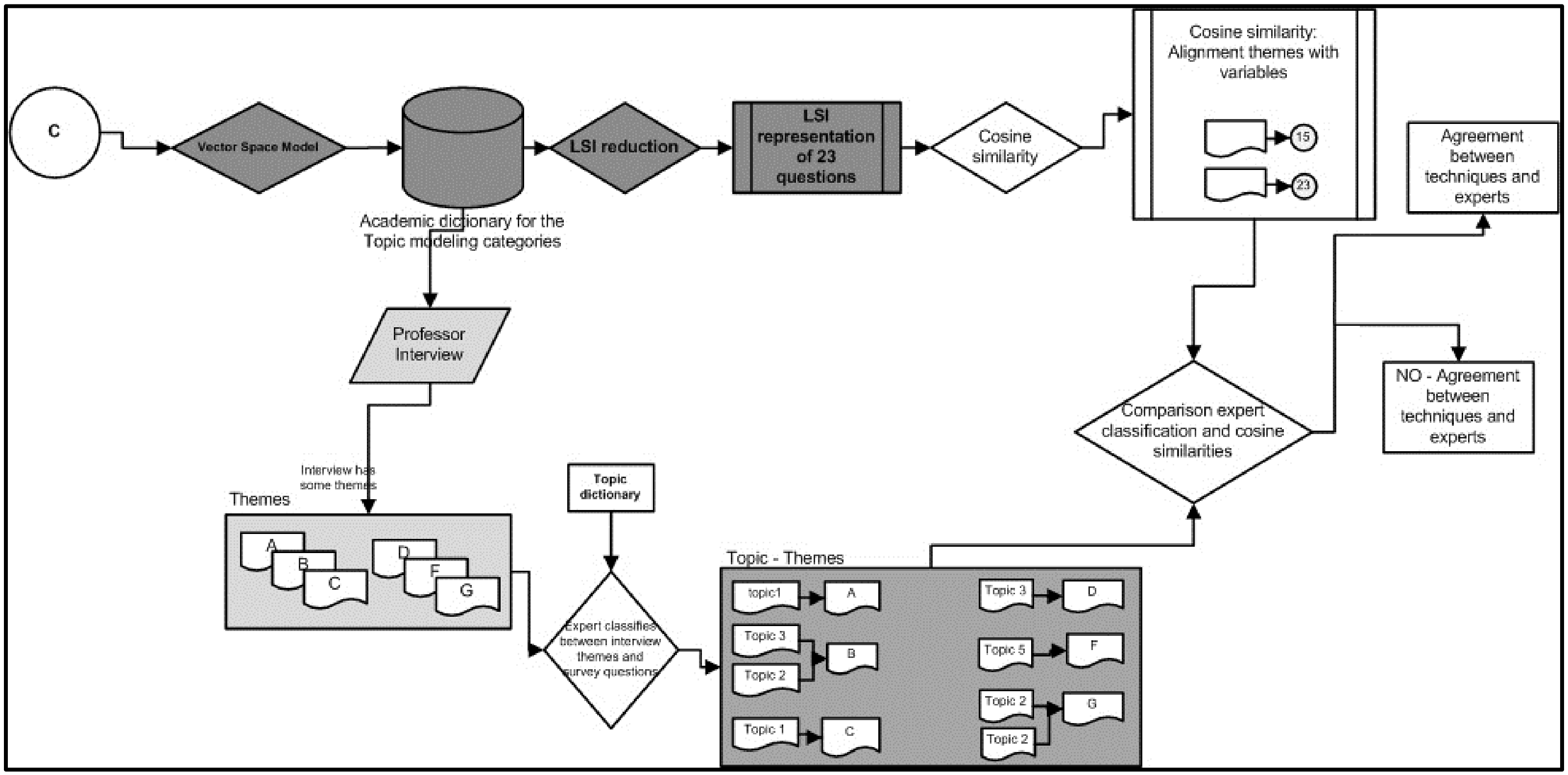

Figure 6.5 Classification and evaluation step

6.5 Results

6.5.1 Number of topics and representation of topics

We present the results of the four steps in the following section. We obtained six topics with their respective words as the most representative. The found topics were *evaluation and feedback*, *feeling or communication*, *time and effort*, *critical thinking (career, engineering professional interest)*, *content and delivery* and *technical aspects*. We present a brief description of the topics first and then the corresponding words with higher probability.

Topic 1 description: Evaluation and feedback: we described the evaluation and feedback as the critical thinking during the exams and evaluations. Evaluation described how teacher evaluated students, what techniques or type of exams they used (open questions or multiple choice questions). Feedback corresponded to what, how and when teacher informs students about their accomplishments or errors. Teacher evaluated if students knew the content of the course. Students evaluated the quality of the course and the quality of the course structure.
The following words represented the topic 1: *Understand, concepts, exam, feedback, evaluation, asks, evaluate, complicated, easy, applied, process, resolve, problems, techniques, component, mathematical, taught, receive, contact, fast, past, apply, analysis, comprehension, difficulty, grade, lack, solution, reflection.*

Topic 2 description: Feelings or communication: we described the feeling or communication topic as the combination of two elements: the communication channels, different ways to reach the students in class, and the reaction or repercussion of the communication channels over how a person felt.

The following words represented the topic 2: *grades, express, confidence, person, opinion, talk, questions, behaviour, likes, speak, afraid, extra, participate, personal, policies, prefer, feel, good, arguments, mood, question, situations, attitude, front, human, mad, technique, make, thinks, answer.*

Topic 3 description: time and effort: we described the time and effort topic as the time student dedicated to the class or their studies. It included learning and studying in group in a friendly environment.

The following words represented the topic 3: *time, attention, assistant, suggest, focus, part, styles, regular, topics, activity, dynamic, improving, lab, methods, negotiation, seller, solve, talking, activities, teaching, reflection, practice, labs, group, ideas, helps, life, working*.

Topic 4 description: critical thinking (career, engineering professional interest): We described the critical thinking topic as the reflection and summarization to the answers of questions, to the relation of concepts; it describes the interest of students in the professional life translated into the working or researching areas.

The following words represented the topic 4: *research, career, interested, motivated, friends, people, interest, motivate, moment, topic, curiosity, idea, low, money, stress, thesis, worried, lost, produce, work, knowledge, considers, end, bad, university, level, study, projects, present show*

Topics 5 description: Content and delivery: This topic comprised the structure and organization of the course and the content and the delivery of it.

The following words represented the topic 5: *pass, practical, respect, giving, motivation, organization, advice, steps, style, capable, easily, jokes, elements, day, material, remember, exercises, comfortable, presentations learn, levels, learning, reach, include, courses, specific, social, information, application, situation*.

Topic 6 description: technical aspects: This topic presented different words that related to the technical nature of the class, such as mathematics, programming, dynamics, research, structures, abilities, etc.

The following words represented the topic 6: *code, project, small, structures, programming, instructions, program, abilities, design, develop, critical, functions, prove, results, works, difficult, test, thinking, writing, objective, worried, applications, shows, style, practice, importance, hours, capable, improving.*

6.5.2 Dictionary construction

Table 6.3 presents the list of the topics obtained from step 6.4.1. In the first column, we placed all the topics from the topic modeling step. In the second column, we placed students' survey questions assigned to that topic. We found three topics without survey questions or that need some complementation. In topic 1, we needed to complement the topic. Thus, we included the feedback section. In topic 3 and 6, we didn't have survey questions to assign. In this case, we had to create the dictionary's description without question descriptions. Instead of using the survey question complete name, we used the short description mentioned in section 6.3.1.

Table 6-3 Topics vs. Survey questions short descriptions

Topic	Survey question
Topic 1: Evaluation and ***feedback***:	X3. Participation evaluation X4. Evaluation fit themes X5. Assessment instructions X22. Assessment results
Topic 2: Feelings or communication	X6. Motivation research X9. Encourage participation X12. Motivation learning material X13. Advise X15. Communication X16. Respect
Topic 3 Time and effort:	

Topic	Survey question
Topic 4: Critical thinking:	X14. Reflection X10. Summarization X11. Relation between concepts X17. Answer questions
Topics 5 Content:	X1. Audiovisual X2. Fulfilled program X7. Course schedule X8. Class policies X18. Class content organize X19. Class content understandable X20. Material preparation X21. Graphics and schemes X23. punctuality
Topic 6: Technical aspects	

In Figure 6.6**,** we present a graphic with the topics and the survey questions that represented them. Topic 1, evaluation and feedback, needed to include the word feedback to complement the topic. In the description of the topic, we used the descriptions of questions X3, X4, X5 and X22. We included this word because none of the questions in the survey expressed explicitly the feedback students should receive from their work. Topic 3, time and effort, did not have any question associated. So, we had to define this topic using the words detailed in the topic modeling step for this topic. Topic 6, Technical aspects, considered technical words or terminology related to different types of courses.

Figure 6.6 presents in grey the words that came from the topic modeling step; and, we present in black survey's questions considered in the analysis. The descriptions for each survey question were the "what" and the "how" of this survey question i.e. Topic 2 included feelings and emotions or motivations. We found question X16, X12, X13, X9, X15 and X6 in topic 2. We described survey question X16 as a complete paragraph related to the concept of "Respect". We made the same with survey questions X12, X13, X9, X15 and X6. We got descriptions for each of the survey question in complete sentences and paragraphs.

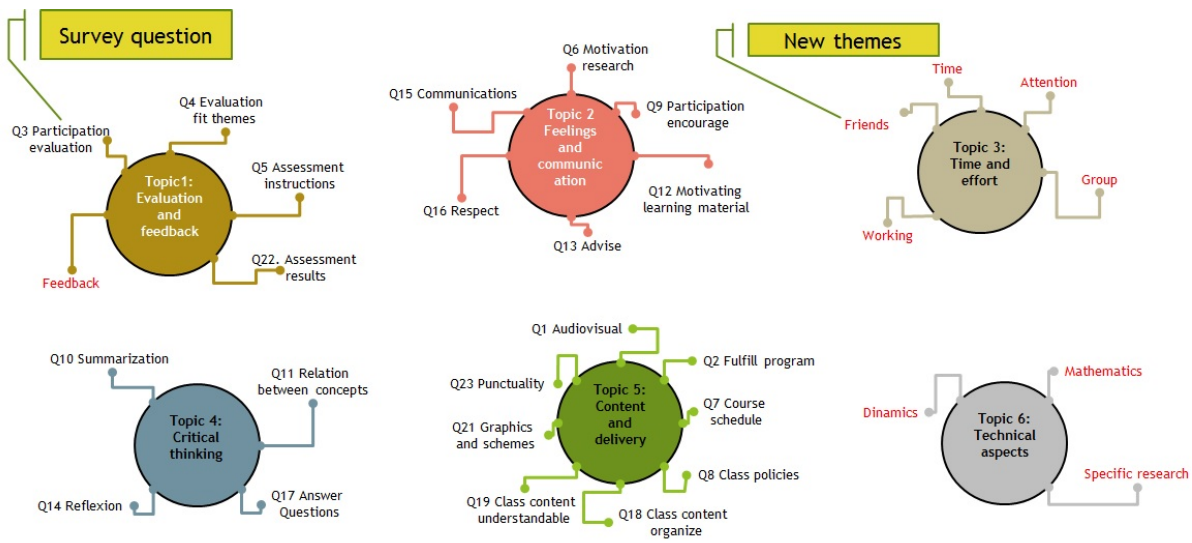

Figure 6.6 Topics survey questions

4.1 Classification and evaluation

In order to apply vector space model, we cleaned up the survey questions descriptions doing the following; we eliminated basic stop words. Also, we omitted repetitive words such as: “student”, “teacher”, “class”, “course” and “numbers”. Every sentence was lower-case and finally tokenized into words. As an example, we made a manual check of the words appearing in the topic’ descriptions of survey question X16 (Respectful), already cleaned. Some of the words related to this topic description were:

('cowardly', 1), ('save', 1), ('irritable', 1), ('clearly', 1), ('resistant', 1), ('futile', 1), ('regimented', 1), ('inspire', 1), ('consoled', 1), ('natural', 1), ('inconsolable', 1), ('looks', 1), ('lawless', 1), ('overwhelming', 1), ('allow', 1), ('fun', 1), ('satisfaction', 1), ('alert', 1), ('joy', 1), ('objectified', 1), ('rarely', 1), ('convinced', 1), ('excited', 1), ('lazy', 1), ('comfy', 1), ('satisfaction', 1), ('discontented', 1), ('stress', 1), ('motivation', 1), ('captivating', 1)

The model converted each interview’s theme in a vector and compared it against the six topic description. These descriptions were also vectors. Cosine measure helped with the comparison. The angle between the interview theme and the topic description should be very small to consider the interview-theme similar to the topic. We used a cosine measure range of 0.7 and above to select the possible classifications.

We classified all the themes from the teachers' interviews into the six topics. We asked two experts to classify manually the interview themes using the topics. As a result, we had three classifications. Incidentally, two experts worked on them. The model made one. We compared the results between each of the experts and the classification model.

Evaluation

We compared the results from experts' classification against model's classification using the kappa's levels. Kappa suggests a range of 0.41 or above, to consider moderate an agreement between the model classification and the experts' classification. The obtained results were the following:

Table 6-4 Evaluation Experts classification vs Model classification

	Expert 1 vs. Model	Expert 2 vs. Model
Interview 1	0.46	0.518
Interview 2	0.63	0.63
Interview 3	0.53	0.40
Interview 4	0.56	0.72
Interview 5	0.28	0.78
Interview 6	0.23	0.46

The results in table 6.4, showed values above .40. Expert 1 and Expert 2 had good kappa values. We considered the cosine value threshold of 0.7 to obtain closer vectors. Using this threshold, the model suggested one, two, three or four answers. This was possible because the 146 themes obtained from the interviews presented more than one category at a time(Kuralenok and Nekrest'yanov 2000) when the model classified the themes. Experts considered in their evaluations only one or two possible topics.

Table 6-5 Percentages of the number of topics per theme

Number of topics	**Percentage of tokens classified with model**
4	2.0%
3	15,86%
2	41,78%
1	21,23%

Table 6.5 shows the number of topics classified by the model in percentages. The model classified the themes between one or two topics. Table 6.6 shows the 146 themes classified in the following topics:

Table 6-6 Number of themes classified by topic

Content	**Critical Thinking**	**Evaluation and feedback**	**Feeling and motivation**	**Technical aspects**	**Time and effort**
25	15	23	49	24	10

Topic "Feeling and motivation" got the highest amount of classifications; the second higher topic was "Content".

4. Discussion

We presented the comparisons between the experts and the model using the kappa measure. We saw a *moderate and substantial* agreement between the expert 2 and the model, unlike the expert 1 who had only *fair and moderate* agreement. Expert 2 agreed strongly with the model than the expert 1. Subjective or objective considerations could influence in expert classification. Experts read the theme and classified it depending on his subjective considerations.

Experts mostly found two possible classifications for each theme. The model found until four possible classifications, but the higher concentration of the model classification was for two topics. This coincided with the experts too.

Finally, the themes were classified through the six topics. Some themes (49/146) were classified as communication and feeling, others as content (25/146). Experts considered that some themes had a stronger component of communication and feeling. This suggested teachers` had a big concern about the content and the communication and feelings of the students during the interview.

6.6 Conclusions

During the experiment, we reached two important conclusions: one related to the topics obtained among the interviews with topic modeling, and a second one, related to the categorization of the themes using the model.

When we examined the topics obtained from the interviews, we found that students' survey questions evaluated very objective topics from the course, such as: activities, content, organization, punctuality, etc. As we have demonstrated, the survey did not cover some topics presented in the interviews. However, we found them using topic modeling. The discovered topics were feeling and emotions, technical aspects and time and effort. The other three topics, content, critical thinking and evaluation and feedback were in the survey.

Teachers connected a component of emotions to the evaluation and feedback~~,~~ because this topic had a lot of words that focused on the evaluation and few words focused on emotions. The time and effort topic focused on some words (such as friends and groups) inside the group of words. Finally, critical thinking suggested career, money and job as principal interests.

In the categorization of the themes using the model, the themes presented in the interviews contained more than one topic. When we applied the categorization using the model, it retrieved one to four possible topics to classify a theme. The most common topic was “feeling and emotions or motivation” and the next frequent mentioned topic was “Content”.

This work suggested the inclusion of new elements which could be part of the students' survey. These elements are part of the perception of the teachers when they work with students during class. This improvement in the content of the survey could help to evaluate important elements considered by teacher and students.

6.7 Future work

Future research in this field should explore deeper the understanding on obtained topics, focusing into the (a) evaluation and feedback, (b) time and effort and (c) students' critical thinking. The application of the probabilistic topic modeling to new interviews could help define more precisely these three components. Hence, this will provide more insights about the type of feedback related to each type of course or to learning style in the students. Consequently, the researcher will have a better understanding about how much time help students understand topics or exercises. Finally, teacher could discovery new insights about ways to improve the students' critical thinking.

Acknowledgments

"CISE, Education Service and Research Center" from Escuela Superior Politécnica del Litoral in Ecuador, for contributing with the data.

This research is sponsored by a scholarship for postgraduate students given by SENESCYT, Secretaría Nacional de Educación Superior, Ciencia y Tecnología e Innovación, at "Academia 2010" Ecuador.

CHAPTER 7

GENERAL DISCUSSION

When teacher receives his evaluation results, the university provides the analysis using descriptive statistic and the evaluation of each question belongs completely to students. The results only reveal the highest and lowest evaluated questions, and the highest and lowest evaluated areas, but provide no feedback from the teacher's point of view. The teacher is left alone with no additional support that could help him reflect on the causes of the low scored questions, and potential solutions to improve those scores. Our methodology provided teacher with a mechanism to incorporate his own perceptions along with his colleagues' perceptions.

In our research, we proposed a solution to this problem. We took the results from the students' survey, applied interviews to the teacher and obtained knowledge from him. This knowledge helped the teacher to review his strategies, understand relations and identify new areas to be evaluated. The new areas were the result of the teachers' reflection on the students' interest.

The objective of this research was to create a methodology to integrate the objective analysis of students with the subjective considerations of teachers. With this objective in mind, we applied D3M that uses objective and subjective data sets. We defined three objectives that guided us to reach the general objective: to build a representation using students' model and teacher's model based on actionable knowledge, to construct a utility semantic measure to evaluate the usefulness of association rules within the model, and to discover new topics from the analysis of interviews to improve the teacher evaluation.

We are going to present the general discussion in three steps: the discussion of the results of the methodology in ESPOL domain, the possible results in the case that we wanted to apply the same methodology in another university, and finally the application of the methodology in a completely different domain.

7.1 Application of the methodology in ESPOL domain

We applied a case study methodology. We construct six model representations, one for each teacher. Each representation expressed the teacher and students' perceptions combined in the model representation. The utility measure we proposed integrate objective analysis and subjective analysis and evaluate association rules considering teacher best patterns, colleague's best considerations and students' perceptions. Each question from the survey has a weight expressed in numbers. When we integrate the best evaluated questions in the measure, we obtain the utility of the rule to be accomplished. The proposed topic analysis reveal that the students 'survey needs to be updated because there are hidden topics the teacher talked about in the interviews which are not evaluated in the survey. In addition, results indicated that it is important to add elements such as critical thinking, feelings, career worries as part of the elements that teachers address in class but students never evaluate.

The original procedure where the teacher received the results of the students' survey only presented the score for each question. The teachers felt disappointed with the low scores obtained in the students' survey results when they considered they have tried their best in the course.

In this research, we implemented a new approach to obtain tacit knowledge from the teachers which can help them improve their academic performance. Hence, this tacit knowledge was transformed into association rules that helped to understand the interaction among different survey questions. Additionally, the tacit knowledge and the rules created new knowledge to understand survey questions. This constitutes a great contribution for new teachers. Now, with the graphical model, they can see patterns within the students' survey question.
Moreover, the analysis of the rules that illustrate the relation between students' considerations and teachers' opinion provided a new way to enhance the students' perception

of the course. The utility measure evaluated the teacher’s own rules and suggest working on an attribute that could improve the teacher's performance. This utility measure is related to the Recuperation-Effort that is a function of the relation usefulness and actionability of the perception of teachers. The teacher perception about the effort to improve could change with time and experience. This fact placed him in a different position of the matrix. We illustrated this in Figure 3.2.

Finally, to correctly redirect the evaluation of teacher’s performance in class, it is important to ensure that students evaluate all the aspects that teachers discuss in class. From the topic modeling, we obtained different areas that are not considered in the evaluation. Critical thinking, technical expressions, and feelings are the new elements that teacher have to work on in class but are not described specifically in the survey.

The applied methodology combined two aspects, teachers` insights and students` evaluation. It provided a compound representation of the teacher and students' thoughts. Students provided their judgments, and the teacher contributed with his experience, and other colleagues complemented with perceptions from their teaching experience. These rules and other rules generated from different groups of students can constantly help improve the model. Using this model, the teacher can make comparisons of his actions among semesters.

7.2 Consideration to apply the methodology in another university

The methodology proposed can be applied to others universities that use surveys in their regular teacher evaluation process. For example, in the ETS or other university, we should expect differences such as the type of the surveys students fill out and the amount of time teachers are willing to spend on the surveys and the interview. We recommend using Likert scale or closed questions during the objective analysis. In case the university has a survey with open questions the generation of association rules is not possible. However, we can obtain clusters of words and topics from the teachers’ interviews. The teacher`s survey S1 should be constructed based on the university survey questions. The researcher must contemplate the comprehension of the questions inside the students` survey and the survey's areas the university aims to evaluate. In the case of the application of an interview to a teacher, it is important to prioritize the specific areas of interest and construct the interview with focusing on the areas of interest. For example, the interview in our case analysed three specific areas (design, promotion and administration of the course). Thus, the university that wants to apply the methodology needs to choose areas of analysis. The amount of time assigned to the tools depends on the amount of areas to be analysed in the interview and in the teacher`s survey. The researcher must define the areas and the questions of the survey. The suggested methodology contemplates: objective analysis of the data, retrieval of the knowledge and meta-knowledge from professor, construction of the representation, evaluation of new rules with the utility measure, application of the interview to the teacher, and the discovery of the new topics (missing in the students' survey)

7.3 Application of the methodology in a completely different domain

Although, the context is slightly different, companies that provide client service and want to better understand their clients can also use our methodology. Service companies already manages client service departments. This department regularly apply surveys to clients using different scales than the Likert scale. Our methodology is applicable using Likert scale; although it is possible to use another scale. Some redefinition of ranges will be needed in this case. As we mentioned before, open questions cannot be analysed with this methodology. The analysis of the objective results from the client's needs should provide the most important attributes for clients. In the company, we could apply the subjective analysis to the client service experts, who might be the Head of the Client Service department in the company and members of his team, specifically the team that works directly with the client. They are the ones who understand the problems clients have regularly. The researcher should apply the surveys and interviews to them. It is vital to create a survey that captures the perceptions about people who provide phone service, or personal service to their clients. We need to define the areas of the service the client survey wants to evaluate, and create an interview question guide that focuses on the client service areas. Once the researcher has defined the areas and the survey questions from client, the application of the methodology should be straightforward: objective analysis from the client`s survey; subjective analysis and retrieval of the domain knowledge from the service experts using the association rules generated in the objective analysis; generation of the model based on client and service experts; the evaluation of new association rules with the utility measure that can improve the client service and finally the topic analysis to the experts interviews to identify the additional topics in the client evaluation. For example, some new topics could be to help change the client's mood during telephone service, understand the technology with a problem or help explain the problem to the client service expert.

GENERAL CONCLUSION

Education settings where the students' objective evaluation plays a crucial role commonly need other perspectives to evaluate teachers` performance. Evaluators do not take into account the subjective considerations because subjectivity depends on teachers' opinions and perceptions. Our research provides a methodology to combine these two and adds an additional one, the semantic consideration, -what is important to each teacher.

Our three main contributions within this research are: First, the methodology to construct a representation based on actionable knowledge; the method generated relationships among questions representing teacher's perceptions, teachers' rules of thumbs and include students` perception; Second, the utility assess of survey's questions using association rules based on objective and subjective measures, where we evaluate association rules to select the one that improved the model. Finally obtaining new topics from teacher's interview, where we classify the interview themes and obtained topics related to the students` survey and discovered new ones. We will explain all these contributions as conclusion:

The methodology constructs a model that includes students' perspectives and teacher's perspectives and gives importance to the rules of thumb teacher uses in classes. The presence of interesting rules (or rules that contradict beliefs) gives the model the ability to discover tacit knowledge already known by the teacher but never explicitly formulated as rules to follow.

In addition, to model rules that represent something interesting-and-new or something old-and-already-known, we evaluate them identifying whether these representation rules stand for stronger or weaker elements in teachers. This means that even though objective analysis provides rules derived from students` objective considerations, the teacher needs to confirm that these attributes evaluated high are indeed part of his strengths. Also, he needs to identify what he considers are his additional strengths. We do the same with teacher's weaknesses, giving a measuring tool to evaluate his attributes.

The model helps the teacher to see what he tacitly knows and also knowledge he didn't have and compares it to what students saw and understood from the time spent in class. There are many objective and subjective measures that evaluate the objectivity or subjectivity of a pattern. In our case, our utility measure evaluates teacher attributes and students` perceptions. This measure is interesting for the teacher because it includes both teachers' opinion, students' opinion and the opinion of other teachers. The measure helps teacher focus on specific attributes e.g. those with the greatest impact on his evaluation, and not in all of the 23 ones. The teachers can focus on key attributes, identifying strengths and weaknesses, and the effort required to compensate the weakness, thereby improving considerably their performance as teachers. The effort variable from the utility measure presents two dimensions, *knowledge and actionability*. These dimensions create four quadrants that identify the poor and rich levels of knowledge and easy and difficult levels of actionability of the attribute and recommend teachers to work on those two quadrants where it is worth improving the attribute.

Once we accomplished research objective 1 and 2, we want to evaluate the similarity between the students' survey and the teachers' interviews. The teachers may be interested in topics that are not considered in the student's survey. The methodology retrieves new knowledge obtained from the teachers' regular practices that may not be part of what universities measure. Our contribution with this research is a methodology to improve evaluation of teachers with additional considerations and other subjective and semantic elements usually not taken into account in the process of teacher evaluation.

With the use of these tools, the teacher will have a clearer picture of his teaching practices. The results of the teacher's survey and interview will explain data that otherwise would have remained hidden. This will give teacher a fresh perspective on his teaching style and will allow him to better understand the impacts of changes introduced in teaching his class. This information could also facilitate the transfer of knowledge to other teachers.

Future work

To expand this research scope, we plan to keep working on subjective elements related to students and teachers. We would like to focus deeper on the following topics obtained from the research: Understanding the relation between teaching and how feelings impact learning. One of the results from this research is the presence of feelings during teachers teaching process, therefore we would like to measure the presence of feelings during the learning students process; additionally, the presence of feelings could be related to the type of course. Then we plan to apply the methodology and to interview teachers teaching same courses, and analysing the type of feelings students present during different course with different difficulty level.

APPENDIX I

SURVEY S1: MODEL QUESTIONNAIRE ON KNOWLEDGE RETRIEVEMENT

Section A: Identification of the most interesting questions

From the following statements used in the students' evaluation, rank each statement from "not applicable" to "extremely important", based on your own judgement.

Table-A I-1 Survey S1 Interest Evaluation of the Survey Questions

	Statements	**NA**	**Not important**	**Less important**	**+ o - important**	**Very important**	**Extremely important**
1.	Professor uses audiovisual help to support the content of the class	☐	☐	☐	☐	☐	☐
2.	Professor fulfills the program proposed at the beginning of course	☐	☐	☐	☐	☐	☐
3.	Professor evaluates periodically student participation in class	☐	☐	☐	☐	☐	☐
4.	Evaluations used in class fit the themes developed in class	☐	☐	☐	☐	☐	☐

	Statements	NA	Not important	Less important	+ o - important	Very important	Extremely important
5.	Professor provides clear instructions for learning assessment (tests, quizzes, presentation, simulations, dramatic representation, role playing, etc)	☐	☐	☐	☐	☐	☐
6.	Professor motivates students to do additional research.	☐	☐	☐	☐	☐	☐
7.	Professor explains the course schedule at the beginning of the course	☐	☐	☐	☐	☐	☐
8.	Professor delivers class content in an organized way	☐	☐	☐	☐	☐	☐
9.	Professor develops class content in an understandable way	☐	☐	☐	☐	☐	☐
10.	Professor prepares instructional, bibliographic or other resources to facilitate learning	☐	☐	☐	☐	☐	☐
11.	Professor responds to questions in class about subjects related to the field	☐	☐	☐	☐	☐	☐

	Statements	NA	Not important	Less important	+ o - important	Very important	Extremely important
12.	Professor explains class policies at the beginning of the course	☐	☐	☐	☐	☐	☐
13.	Professor encourages active student participation in class	☐	☐	☐	☐	☐	☐
14.	Professor summarizes key ideas discussed before moving to a new unit or topic	☐	☐	☐	☐	☐	☐
15.	Professor establishes relationships between new concepts and already known concepts, whenever possible	☐	☐	☐	☐	☐	☐
16.	Professor motivates learning of the course material	☐	☐	☐	☐	☐	☐
17.	Professor is willing to answer questions and offer advice within and outside the classroom	☐	☐	☐	☐	☐	☐
18.	Professor promotes reflection on topics covered	☐	☐	☐	☐	☐	☐
19.	Professor maintains fluid communication with students	☐	☐	☐	☐	☐	☐
20.	Professor is respectful to students	☐	☐	☐	☐	☐	☐

	Statements	NA	Not important	Less important	+ o - important	Very important	Extremely important
21.	Professor frequently uses schemes and graphics to support explanations	☐	☐	☐	☐	☐	☐
22.	Professor provides the results of assignments on time	☐	☐	☐	☐	☐	☐
23.	Professor attends classes on time	☐	☐	☐	☐	☐	☐

Section B : Selection of the most interesting area and the most interesting questions per area

Question 1:

Attributes from 1 to 23 (Section A), focus on the evaluation of specific areas of the courses ("Design", "Learning Promotion", "Production and Material", and "Education Management" area). Which of these areas are more important for you? Choose all the areas that are important for you (You can choose one or more)

☐Design of the course
☐Learning promotion of the course
☐Production and teaching materials & Education management of the course
☐None of the above

Question 2:

If you checked **Design of the course** in **question 1 - Section B**, which of the following attributes are more important when you analyze your own development as professor. **Select up to 3** attribute.

☐1 Professor uses audiovisual help to support the content of the class
☐2 Professor fulfills the program proposed at the beginning of course
☐3 Professor evaluates periodically student participation in class
☐4 Evaluations fit the themes developed in class
☐5Professor provides clear instructions for learning assessment (tests, quizzes, presentation, simulations, dramatic representation, role playing, etc)
☐6 Professor motivates students to do additional research.
☐None of the above

Question 3:

If you checked the **<u>Learning promotion of the course</u>** area in **<u>question 1 – Section B</u>**, which of the following attributes are more important when you analyze your own development as professors. Select up to 7 attributes.

☐7 Professor explains the course schedule at the beginning of the course
☐8 Professor explains class policies at the beginning of the course
☐9 Professor encourages active student participation in class
☐10 Professor summarizes key ideas discussed before moving to a new unit or topic
☐11 Professor establishes relationships between new concepts and already known concepts, whenever possible
☐12 Professor motivates learning of the course material
☐13 Professor answers questions and offers advice within and outside the classroom
☐14 Professor promotes reflection on topics covered
☐15 Professor maintains fluid communication with students
☐16 Professor is respectful to students
☐17 Responded to questions in class about subjects related to the field
☐18 Professor delivers class content in an organized way
☐19 Professor develops class content in an understandable way
☐None of the above

Question 4:

If you checked the **<u>Production of teaching materials, & education Management</u> area** in **<u>question 1-Section B</u>**, which of the following questions are more important when you analyze your own development as professors. Select up to 2 attributes.

☐X20 Prepared instructional, bibliographic or other resources to facilitate learning
☐X21Frequently used schemes and graphics to support their explanations
☐X22Provided the results of the assessments on time
☐X23Attended classes on time
☐None of the above

Section C: Identification of actionability in the low evaluated questions

Let's suppose that students evaluate you as **REGULAR** in all the 23 statements. How could you use these results to improve your teaching? Would you say that this information is "**useful**", "**related to another answer too**", "**useful but not easy to change**" or you "**don't know how to use that information**"? Remember that:

- **Useful and straightforward:** If the statement can be used immediately or applied immediately, or find a solution immediately as it is, if you know it.

- **Maybe the result is related to something else:** when you think that the result of that statement maybe is also related to another statement or situation.
- **Useful but not easy to change:** When you will need to work a lot to change it
- **I don’t know how to use the results:** When you don’t know what to do with a regular evaluation in that statement.

Select only **ONE** of the possible options per statement.

Table –A I-2 Low Evaluation Reaction

	Statements	**I don’t know how to use the results**	**Useful but not easy to change**	**Maybe the result is related to something else**	**Useful and straightforward**
1.	Professor seldom uses audiovisual help to support the content of the class	☐	☐	☐	☐
2.	Professor seldom fulfills the program proposed at the beginning of course	☐	☐	☐	☐
3.	Professor doesn’t evaluate periodically student participation in class	☐	☐	☐	☐
4.	Evaluations used in class don’t fit the themes developed in class	☐	☐	☐	☐
5.	Professor seldom provides clear instructions for learning assessment (tests, quizzes, presentation, simulations, dramatic representation, role playing, etc.)	☐	☐	☐	☐
6.	Professor doesn’t motivate students to do additional research.	☐	☐	☐	☐
7.	Professor doesn’t explain the course schedule at the beginning of the course.	☐	☐	☐	☐
8.	Professor doesn’t explain class policies at the beginning of the course.	☐	☐	☐	☐

	Statements	I don't know how to use the results	Useful but not easy to change	Maybe the result is related to something else	Useful and straightfo rward
9.	Professor seldom encourages active student`s participation in class.	☐	☐	☐	☐
10.	Professor doesn't summarize key ideas discussed before moving to a new unit or topic.	☐	☐	☐	☐
11.	Professor doesn't establish relationships between new concepts and already known concepts, whenever possible.	☐	☐	☐	☐
12.	Professor doesn't motivate learning of the course material.	☐	☐	☐	☐
13.	Professor wasn't willing to answer questions and offer advice within and outside the classroom.	☐	☐	☐	☐
14.	Professor doesn't promote reflection on topics covered.	☐	☐	☐	☐
15.	Professor doesn't maintain fluid communication with students.	☐	☐	☐	☐
16.	Professor isn't respectful to students.	☐	☐	☐	☐
17.	Professor doesn't respond to questions in class about subjects related to the field.	☐	☐	☐	☐
18.	Professor doesn't deliver class content in an organized way.	☐	☐	☐	☐
19.	Professor doesn't develop class content in an understandable way.	☐	☐	☐	☐
20.	Professor doesn't prepare instructional, bibliographic or other resources to facilitate learning.	☐	☐	☐	☐
21.	Professor rarely uses schemes and graphics to support their explanations.	☐	☐	☐	☐

	Statements	I don't know how to use the results	Useful but not easy to change	Maybe the result is related to something else	Useful and straightforward
22.	Professor doesn't provide the results of the assessments on time.	☐	☐	☐	☐
23.	Professor doesn't attend classes on time.	☐	☐	☐	☐

Section Da: Creation of open association rules

Question 1:

Choose **<u>up to 5 attributes</u>** from **section A,** evaluated as "extremely important". Place the number of the attribute in **<u>Table-A I-4, "Attribute Y" column</u>**. Choosing from the remaining attributes (it doesn't matter the rank of the attribute), select only 2 attributes that you think can produce the already selected "Y" attribute. **<u>You can use the same attribute in different consequents but each rule has to have different attributes in the antecedent. Don't use the same attribute in the antecedent and consequent at the same time.</u>**

The way to read the sentence in Table-A I-3 is the following: If professor is excellent motivating students to do additional research **AND** if professor is excellent developing class content in an understandable way (19) **THEN** automatically professor is evaluated as excellent in attending classes on time(23)

Table-A I-3 Example of open association rules

	Attribute X1	Attribute X2		Attribute Y
RULE A	6	19	THEN	23

Table-A I-4 Teacher's open association rules

	Attribute X1	Attribute X2		Attribute Y
RULE A			THEN	
RULE B				
RULE C				
RULE D				
RULE E				

Question 2:

From Table-A I-4 which of the two attributes in the antecedent per rule, is more important? Choose one attribute from the antecedent and copy the respective "attribute Y" to the Table-A I-5.

Table - A I -5 Reduced open association rules

	Attribute X		Attribute Y
RULE A		THEN	
RULE B			
RULE C			
RULE D			
RULE E			

Section Db: Creation of contextual association rules

Question 1:

Using the **<u>three attributes</u>** selected in **section B - <u>question 2 (Design area)</u>**, place their numbers in **<u>Table-A I-7, "Attribute Y" column</u>**. Choosing from the 6 attributes of the design area, section B - question 2, select 2 attributes you think can produce "attribute Y" per area. For example:

Table-A I-6 Example design area association rules

	Attribute X1	Attribute X2	THEN	Attribute Y
RULE A	1	5		6

<u>The way to read the sentence is the following:</u> If professor is excellent in using audiovisual help to support the content of the class **AND** if Professor is excellent providing clear instructions for learning assessment (5) (tests, quizzes, presentation, simulations, dramatic representation, role playing, etc.), **THEN** automatically the professor will be evaluated as excellent in motivating students to do additional research. . **<u>You can use the same attribute in different consequents but each rule has to have different attributes in the antecedent. Don't use the same attribute in the antecedent and consequent at the same time.</u>**

Table-A I-7 Teacher's design areas association rules

	Attribute X1	Attribute X2		Attribute Y
RULE A			THEN	
RULE B				
RULE C				

Question 2:

From Table-A I-7, which of the two attributes per rule, do you think is more important? Copy the "attributes Y" in Table-A I–8 and choose only one of the antecedent attributes that can produce "Attribute Y".

Table-A I-8 Reduced context design association rules

	Attribute X		Attribute Y
RULE A		THEN	
RULE B			
RULE C			

Question 3:

Using the **<u>seven attributes</u>** selected in **<u>Section B - question 3 (Learning promotion of the course area)</u>**, place their numbers in the **T<u>able-A I-10, "Attribute Y"</u>**. Choosing from the 13 attributes of the Learning promotion of the course area, Section B - question 3, select 2 attributes you think can produce the "attribute Y". For example: If you select attribute 9 as result in column Y you can say that attribute 17 and 11 can produce attribute 9.

Table-A I-9 Example of Learning promotion course association rule

	Attribute X1	Attribute X2		Attribute Y
RULE A	17	11	THEN	9

.

<u>The way to read the sentence is the following:</u> If professor is excellent responding questions in class about subjects related to the field (17) **AND** if professor is excellent establishing relationships between new concepts and already known concepts whenever possible, **THEN** automatically professor will be evaluated as excellent encouraging active student participation in class. **<u>You can use the same attribute in</u> <u>different rule consequents but each rule has to</u>**

<u>have different attributes in the antecedent. Don't use the same attribute in the antecedent and consequent at the same time.</u>

Table-A I-10 teacher's learning promotion association rules

	Attribute X1	Attribute X2		Attribute Y
RULE A				
RULE B				
RULE C				
RULE D			THEN	
RULE E				
RULE F				
RULE G				

Question 4:

From Table A I - 10, Which of the two antecedent attributes per rule, do you think is more important?. Please choose one and then copy it and the corresponding consequent attribute in Table A I - 11. Repeat this procedure for each rule.

Table-A I-11 Final Learning promotion association rules

	Attribute X		Attribute Y
RULE A			
RULE B			
RULE C			
RULE D		THEN	
RULE E			
RULE F			
RULE G			

Question 5:

Using the **<u>two attributes</u>** selected in **<u>section B- question 4 (</u>**Production of teaching materials, & Management of education), place their numbers in the **Table A I -13 "Attribute Y" column**. Choosing from the 4 attributes of the Production and teaching materials for the course area, section B - question 4, select 2 attributes you think can produce the Y attribute. For example

Table-A I-12 Example Production of teaching materials & Management of education

	Attribute X1	Attribute X2	THEN	Attribute Y
RULE A	20	23		21

The way to read the sentence is the following: If professor is excellent preparing instructional, bibliographic or other resources to facilitate learning **AND** if professor is excellent on attending classes on time (23) **THEN** automatically professor will be evaluated as excellent in the frequently use of schemes and graphics to support explanations (21). **You can use the same attribute in different rule consequents but each rule has to have different attributes in the antecedent. Don't use the same attribute in the antecedent and consequent at the same time.**

Table-A I-13 Example Production of teaching materials & Management of education association rules

	Attribute X1	Attribute X2		Attribute Y
RULE A			THEN	
RULE B				

Question 6:

From Table A I-13 which of the two attributes per rule, do you think is more important? Choose one antecedent attribute and copy the same "attribute Y" per rule in table A I-14.

Table -A I-14 Final teacher Example Production of teaching materials & Management of education association rules

	Attribute X		Attribute Y
RULE A		THEN	
RULE B			

Attribute description sheet

Design area	1	Used audiovisual help to support the content of the class
	2	Fulfilled the program proposed at the beginning of course
	3	Evaluated periodically student participation in class
	4	Evaluations fit the themes developed in class
	5	Provided clear instructions for learning assessment (tests, quizzes, presentation, simulations, dramatic representation, role playing, etc)
	6	Motivated students to do additional research.
Learning promotion area	7	Explained the course schedule at the beginning of the course
	8	Explained class policies at the beginning of the course
	9	Encouraged active student participation in class
	10	Summarized key ideas discussed before moving to a new unit or topic
	11	Established relationships between new concepts and already known concepts, whenever possible
	12	Motivated learning of the course material
	13	The professor was willing to answer questions and offer advice within and outside the classroom
	14	Promoted reflection on topics covered
	15	Maintained fluid communication with students
	16	He/she was respectful to students
	17	Responded to questions in class about subjects related to the field
	18	Delivered class content in an organized way
	19	Developed class content in an understandable way
Production and teaching materials and **education management area**	20	Prepared instructional, bibliographic or other resources to facilitate learning
	21	Frequently used schemes and graphics to support their explanations
	22	Provided the results of the assessments on time
	23	Attended classes on time

APPENDIX II

SURVEY S2: QUESTIONNAIRE FOR TEACHER'S EVALUATION RULES

Instructions

ASSOCIATION RULES EXAMPLE: TEACHER 011

In the following tables:

Read all the rules from the page or group of pages that evaluate the same consequent (element after the THEN operator), select only one **interesting rule** and mark it in the **interesting column** (each group should have only one rule marked on the **"interesting"** column).

Notice that **interesting** means: "if the rules surprise the user (because it is new and different from previous concepts or believing) and if user can use it taking any advantage of this information".

Select a knowledge category for each rule. The knowledge categories that you will have are: "Unexpected", "Useful", "Obvious", "Previous" or "Irrelevant":

Unexpected knowledge: "is a rule that represents a novelty, something that a user never thought of or something that contradicts his /her previous knowledge". In case that you categorize any rule with the unexpected knowledge, please select which side of the rule was unexpected for you:

(LS) Left side (Antecedent Unexpected),
(RS) Right side (Consequent Unexpected),
(Labský and Svátek) Both sides (Antecedent and Consequent Unexpected)

Useful knowledge: "is a rule representing a knowledge which can be used to assist the user in some decision making"
Obvious knowledge: "is a rule that indicates solid domain knowledge"
Previous Knowledge: "a rule with this evaluation represent user's knowledge formed with past experiences, something that you realize because of your experience"
Irrelevant Knowledge: "is a rule that represents knowledge that is not important or necessary according to the user".

Observation: If conjunction (an attribute inside the antecedent of a rule) is evaluated as **ALWAYS,** it means that professor received **9 or 10** in the same scale.

	RULES EVALUATED AS <u>EXCELLENT(ALWAYS),</u> (ANTECEDENT AND CONSEQUENT SIDE)	IINTERESTING	UNEXPECT. knowledge LS BS	RS		USEFUL knowledge	OBVIOUS knowledge	PREVIOUS knowledge	IRRELEVANT knowledg
1	**IF** **X15** Professor **ALWAYS** maintains fluid communication with students **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐
2	**IF** **X19** Professor **ALWAYS** develops class content in an understandable way **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐
3	**IF** **X15** Professor **ALWAYS** maintains fluid communication with students **AND** **X19** Professor **ALWAYS** develops class content in an understandable way **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐
4	**IF** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐

	RULES EVALUATED AS <u>EXCELLENT(ALWAYS),</u> (ANTECEDENT AND CONSEQUENT SIDE)	IINTERESTING	UNEXPECT. knowledge LS	RS	BS	USEFUL knowledge	OBVIOUS knowledge	PREVIOUS knowledge	IRRELEVANT knowledg
5	**IF** **X19** Professor **ALWAYS** develops class content in an understandable way **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐
6	**IF** **X15** Professor **ALWAYS** maintains fluid communication with students **ANDx** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐
7	**IF** **X15** Professor **ALWAYS** maintains fluid communication with students **AND** **X19** Professor **ALWAYS** develops class content in an understandable way **AND** **X22** Professor ALWAYS provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X12 ALWAYS** motivates learning of the course material	☐	☐	☐	☐	☐	☐	☐	☐

	RULES EVALUATED AS <u>EXCELLENT(ALWAYS),</u> (ANTECEDENT AND CONSEQUENT SIDE)	IINTERESTING	UNEXPECT. knowledge			USEFUL knowledge	OBVIOUS knowledge	PREVIOUS knowledge	IRRELEVANT knowledg
			LS	RS	BS				
1	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **THEN** Professor is evaluated as she/he... **X15** **ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐
2	**IF** **X19** Professor **ALWAYS** develops class content in an understandable way **THEN** Professor is evaluated as she/he... **X15** **ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐
3	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **AND** **X19** Professor **ALWAYS** develops class content in an understandable way **THEN** Professor is evaluated as she/he... **X15** **ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐
4	**IF** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X15** **ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐

	RULES EVALUATED AS **EXCELLENT(ALWAYS),** (ANTECEDENT AND CONSEQUENT SIDE)	IINTERESTING	UNEXPECT. knowledge LS	RS	BS	USEFUL knowledge	OBVIOUS knowledge	PREVIOUS knowledge	IRRELEVANT knowledg
5	**IF** **X19** Professor **ALWAYS** develops class content in an understandable way **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X15 ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐
6	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X15 ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐
7	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **AND** **X19** Professor **ALWAYS** develops class content in an understandable way **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X15 ALWAYS** maintains fluid communication with students	☐	☐	☐	☐	☐	☐	☐	☐

	RULES EVALUATED AS **<u>EXCELLENT(ALWAYS),</u>** **(ANTECEDENT AND CONSEQUENT SIDE)**	**IINTERESTING**	**UNEXPECT. knowledge**			**USEFUL knowledge**	**OBVIOUS knowledge**	**PREVIOUS knowledge**	**IRRELEVANT knowledg**
			LS	**RS**	**BS**				
1	**IF** **X15** Professor **ALWAYS** maintains fluid communication with students **THEN** Professor is evaluated as she/he... **X19 ALWAYS** develops class content in an understandable way	☐	☐	☐	☐	☐	☐	☐	☐
2	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **THEN** Professor is evaluated as she/he... **X19 ALWAYS** develops class content in an understandable way	☐	☐	☐	☐	☐	☐	☐	☐
3	**IF** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X19 ALWAYS** develops class content in an understandable way	☐	☐	☐	☐	☐	☐	☐	☐
4	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X19 ALWAYS** develops class content in an understandable way	☐	☐	☐	☐	☐	☐	☐	☐

	RULES EVALUATED AS <u>EXCELLENT(ALWAYS),</u> (ANTECEDENT AND CONSEQUENT SIDE)	IINTERESTING	UNEXPECT. knowledge LS RS BS			USEFUL knowledge	OBVIOUS knowledge	PREVIOUS knowledge	IRRELEVANT knowledg
5	**IF** **X15** Professor **ALWAYS** maintains fluid communication with students **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X19 ALWAYS** develops class content in an understandable way	☐	☐	☐	☐	☐	☐	☐	☐
6	**IF** **X12** Professor **ALWAYS** motivates learning of the course material **AND** **X15** Professor **ALWAYS** maintains fluid communication with students **AND** **X22** Professor **ALWAYS** provides the results of the assessments on time **THEN** Professor is evaluated as she/he... **X19 ALWAYS** develops class content in an understandable way	☐	☐	☐	☐	☐	☐	☐	☐

APPENDIX III

INTERVIEW

Question 1: How much time do you assign for the preparation of resources like quizzes, homework, new material etc?

Question 2: What techniques do you use to improve learning or make it faster? (Use of evaluations, types of evaluations like surprise evaluations).

Question 3: How do you motivate your students to do research? (Do you give resources to do research, research techniques?

Question 4: Could you list examples of "demonstration of respect" you have exercised with your students? (Negotiation with students? make jokes in class? Attend classes on time?).

Questions 5: What is your communication strategy during your classes and outside your classes? (Talkative teachers, non-talkative teachers, students' preference about teacher communication characteristics).

Question 6: How do you teach reflection of topics to your students (best moment to do this during class?, when do you think is better, when could be wrong?).

Question 7: What techniques do you use to summarize key ideas?

Questions 8: According to your experience is there any specific strategy of availability that works for you? For example, giving your availability to the students using facebook, cellphone and email? (Other options of mix strategy of availability: sms, chat systems, other social Medias (Flickr, Twitter), content systems with chat systems included, office hours) (Are you available in all those applications, all the courses need that kind of availability, past experience of good or bad availability using technology?)

Question 9: Students are motivated to participate in class? (do you motivate the participation? How? What do you use for motivate the participation? In your own words, what does the expression "participation of the students in class" means?

Question 10: What activities reflect that you are an organized professor inside and outside your class? (Examples of activities, can you tell me if these are part of class organization: Prepare good material for class, showing material following the same order of the course schedule, showing material following the same order and dates of the course schedule, showing material organize by you but doesn't follow the course schedule

Question 11: How you identify students didn't understand you? (Which one of the following are more applicable in your case: Students don't ask after the lecture, the students faces' expressions give you an idea, you make some questions to get what students understood, Ask students to make an exercise)

Question 12: Mention some techniques you use to facilitate students learning. (what do you use as learning resources (Research papers or journals, Books or magazines, your own presentations, Material prepared by you (specific documents created and typed by you), Pieces of movies, Complete movies, TV shows, Exercises in class, Cases in class, Students helping other students)

Question 13: How important is the use of graphical medias (what do you use, why)

Question 14: What do you do when you sense that students are frustrated because they don't understand (students' expressions, offering more coaching, reading lectures, intellectual help?)

APPENDIX IV

ETHICAL CERTIFICATIONS

Certificate of Completion

The National Institutes of Health (NIH) Office of Extramural Research certifies that **Otilia Alejandro** successfully completed the NIH Web-based training course "Protecting Human Research Participants".

Date of completion: 08/03/2014

Certification Number: 1512650

Le génie pour l'industrie

Comité d'éthique de la recherche
École de technologie supérieure

Date : 8 juillet 2015 **H20140102** **Nouvelle**

OBJET : **Titre du projet :** Méthodologie de forage de données dirigée par le domaine pour les systèmes d'aide à la décision, fondée sur des mesures objectives, subjectives et sémantiques

Responsable du projet : Sylvie Ratté

Décision : APPROBATION FINALE

Monsieur,

Les modifications et précisions demandées par le CÉR dans sa lettre du 14 mai 2014 ayant été apportées adéquatement, votre projet peut aller de l'avant.

Veuillez toutefois noter que cette approbation n'est valable que pour une année, soit jusqu'au 8 juillet 2016. Vous devrez donc annuellement demander le renouvellement de l'approbation au Comité, sans quoi le projet sera considéré comme terminé. Dans la perspective où il devait être terminé, vous devrez fournir un rapport final suivant le modèle disponible sur la page Internet de l'ÉTS. Ce rapport est attendu pour le 31 décembre 2016.

Veuillez agréer, Monsieur, l'expression de mes sentiments les meilleurs.

Caroline Chartrand, M.Sc.
Coordonnatrice
Comité d'éthique de la recherche

DATE DE FIN DE L'APPROBATION
8 juillet 2016

c.c. : Sylvain G. Cloutier, doyen de la recherche
Paul Gervais, président du CÉR de l'ÉTS

BIBLIOGRAPHY

Arthur, L. 2009. "From performativity to professionalism: lecturers' responses to student feedback." *Teaching in Higher Education*. vol. 14 nº 4, 441-454 p.

Badur, B. and S. MARDIKYAN. 2011. "Analyzing teaching performance of instructors using data mining techniques." *Informatics in Education-An International Journal*. vol. 10 nº 2, 245-257 p.

Bailey, C., P. R. Baines, H. Wilson and M. Clark. 2009. "Segmentation and customer insight in contemporary services marketing practice: why grouping customers is no longer enough." *Journal of Marketing Management*. vol. 25 nº 3-4, 227-252 p.

Bing, L., H. Wynne, C. Shu and M. Yiming. 2000. "Analyzing the subjective interestingness of association rules." *Intelligent Systems and their Applications, IEEE*. vol. 15 nº 5, 47-55 p.

Blei, D. M. 2012. "Probabilistic topic models." *Communications of the ACM*. vol. 55 nº 4, 77-84 p.

Burton, S. H., R. G. Morris, C. G. Giraud-Carrier, J. H. West and R. Thackeray. 2014. "Mining useful association rules from questionnaire data." *Intelligent Data Analysis*. vol. 18 nº 3, 479-494 p.

Cao, L., C. Zhang, P. Yu and Y. Zhao. 2010. *D3M AKD Frameworks*. In Domain Driven Data Mining. Longbing Cao,Chengqi Zhang,PhilipS Yu and Yanchang Zhao (Eds.). 93-112 p. vol. 2. Springer US.

Cao, L., C. Zhang, P. Yu and Y. Zhao. 2010. *D3M Methodology*. In Domain Driven Data Mining. Longbing Cao,Chengqi Zhang,PhilipS Yu and Yanchang Zhao (Eds.). 27-47 p. vol. 2. Springer US.

Cao, L., Y. Zhao, H. Zhang, D. Luo, C. Zhang and E. Park. 2010. "Flexible Frameworks for Actionable Knowledge Discovery." *Knowledge and Data Engineering, IEEE Transactions*. vol. 22 nº 99, 1-1 p.

Clark, M. 2000. "Getting participation through discussion." *SIGCSE Bull.* vol. 32 nº 1, 129-133 p.

Creswell, J. W. 2012. *Qualitative Inquiry and Research Design: Choosing Among Five Approaches.* SAGE Publications, 448 p.

Davenport T. H , P. L. 1999. *Working knowledge: How organizations manage what they know* 1e. Boston, Harvard Business Press 2013, 224 p.

Deakin, H. and K. Wakefield. 2013. "Skype interviewing: Reflections of two PhD researchers." *Qualitative Research*. vol. 14 nº 5, 603-616 p.

Deerwester, S. C., S. T. Dumais, T. K. Landauer, G. W. Furnas and R. A. Harshman. 1990. "Indexing by latent semantic analysis." *Journal of the American Society for Information Science and Technology*. vol. 41 nº 6, 391-407 p.

Eshach, H., Y. Dor-Ziderman and Y. Yefroimsky. 2014. "Question Asking in the Science Classroom: Teacher Attitudes and Practices." *Journal of Science Education and Technology*. vol. 23 nº 1, 67-81 p.

ESPOL. 2011. «Proyecto de Acreditacion Internacional de Programas de Ingenieria». Online <http://www.abet.espol.edu.ec/inicio.html>. Accessed Noviembre 23, 2015.

Flavell, J. H. 1979. "Metacognition and cognitive monitoring: A new area of cognitive–developmental inquiry." *American psychologist*. vol. 34 nº 10, 906 p.

Foundation, T. A. S. 2000. *SPSS Statistics for Windows.* 13.0.Software. Chicago SPSS Inc.

Fuller, C. M., D. P. Biros and D. Delen. 2011. "An investigation of data and text mining methods for real world deception detection." *Expert Systems with Applications*. vol. 38 nº 7, 8392-8398 p.

Geng, L. and H. Hamilton. 2007. *Choosing the Right Lens: Finding What is Interesting in Data Mining*. In Quality Measures in Data Mining. Howard Hamilton Fabrice Guillet (Eds.). 3-24 p. vol. 43. Springer Berlin Heidelberg.

Geng, L. and H. J. Hamilton. 2006. "Interestingness measures for data mining: A survey." *ACM Comput. Surv.* vol. 38 nº 3, 9 p.

Grimmer, J. and B. M. Stewart. 2013. "Text as data: The promise and pitfalls of automatic content analysis methods for political texts." *Political Analysis*. vol. 3 nº 21, 30 p.

Gwet, K. L. 2015. "Testing the Difference of Correlated Agreement Coefficients for Statistical Significance." *Educational and Psychological Measurement*. vol. 69 nº 4, 675-695 p.

Herrmann, T., A. Kienle and N. Reiband. 2003. "Meta-Knowledge-A Success Factor for Computer-Supported Organizational Learning in Companies." *Educational Technology & Society*. vol. 6 nº 1, 9-13 p.

Hirschberg, J., J. N. Lye, M. Davies and C. G. Johnston. 2015 "Measuring Student Experience: Relationships between Teaching Quality Instruments (TQI and Course Experience Questionnaire (CEQ)." Social Science Research Network Retrieved from http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2563326. Accessed September 2015.

Hirschowitz, A. 2001. "Closing the CRM loop: The 21st century marketer's challenge: Transforming customer insight into customer value." *Journal of Targeting, Measurement and Analysis for Marketing*. vol. 10 nº 2, 168-178 p.

IDEA. 2015. «Improving Learning in Higher Education ». Online <http://ideaedu.org/>. Accessed August 30, 2015.

Jackson, D. L., C. R. Teal, S. J. Raines, T. R. Nansel, R. C. Force and C. A. Burdsal. 1999. "The Dimensions of Students' Perceptions of Teaching Effectiveness." *Educational and Psychological Measurement*. vol. 59 nº 4, 580-596 p.

Jamoussi, S. and H. Ameur. 2013. "Dynamic construction of dictionaries for sentiment classification". *Cloud and Green Computing (CGC), 2013 Third International Conference*, (Karlsruhe, Germany),418-425 p. IEEE.

Jiabin, D., H. Juanli, C. Hehua and W. Juebo. 2010. "A Study of Teaching Evaluation in Adult Higher Education Based on Decision Tree". *Information Technology and Computer Science (ITCS), 2010 Second International Conference* (Kiev),381-385 p. IEEE.

Jiang, Y., K. Wang, A. Tuzhilin and A. W.-C. Fu. 2005. "Mining patterns that respond to actions". *Data Mining, Fifth IEEE International Conference*, (Houston, TX), 669-672 p. IEEE Computer Society.

Khan, M. S., M. Muyeba and F. Coenen. 2008. "A Weighted Utility Framework for Mining Association Rules". *Computer Modeling and Simulation, 2008. EMS '08. Second UKSIM European Symposium*, (Liverpool, England), 87-92 p. IEEE.

Kipper, H. and T. Ruutmann. 2012. "Teaching for understanding in engineering education". *Interactive Collaborative Learning (ICL), 2012 15th International Conference*, (Villach, Austria), 1-9 p. IEEE.

Kleinberg, J., C. Papadimitriou and P. Raghavan. 1998. "A Microeconomic View of Data Mining." *Data Mining and Knowledge Discovery*. vol. 2 nº 4, 311-324 p.

Kuralenok, I. and I. Nekrest'yanov. 2000. "Automatic document classification based on latent semantic analysis." *Programming and Computer Software*. vol. 26 nº 4, 199-206 p.

Kvale, S. and S. Brinkmann. 2009. *Interviews: Learning the craft of qualitative research interviewing.* Sage, 405 p.

Labský, M. and V. Svátek. 2008. *Combining Multiple Sources of Evidence in Web Information Extraction*. In Foundations of Intelligent Systems. Aijun An,Stan Matwin,Zbigniew W. Ras and Dominik Slezak (Eds.). 471-476 p. vol. 4994. Springer Berlin Heidelberg.

Lanfang, L., P. Qingxian and Q. Xiuqin. 2010. "New application of association rules in teaching evaluation system". *Computer and Information Application (ICCIA), 2010 International Conference*, (Tianjin, China),13-16 p. IEEE.

Lee, D., S.-H. Park and S. Moon. 2013. "Utility-based association rule mining: A marketing solution for cross-selling." *Expert Systems with Applications*. vol. 40 nº 7, 2715-2725 p.

Lee, J. and Y. Jeong. 2012. "User-centric knowledge representations based on ontology for AEC design collaboration." *Computer-Aided Design*. vol. 44 nº 8, 735-748 p.

Lenca, P., B. Vaillant, P. Meyer and S. Lallich. 2007. *Association Rule Interestingness Measures: Experimental and Theoretical Studies*. In Quality Measures in Data mining. Guillet Fabrice and Hamilton Howard (Eds.). 51-76 p. vol. 43. Springer Berlin Heidelberg.

Longbing, C. 2008. "Domain Driven Data Mining (D3M)". *Data Mining Workshops, 2008. ICDMW '08. IEEE International Conference*, (Pisa, Italy),74-76 p. IEEE.

Longbing, C. 2010. "Domain-Driven Data Mining: Challenges and Prospects." *Knowledge and Data Engineering, IEEE Transactions*. vol. 22. nº 6, 755-769 p.

Longbing, C. C. Z. 2007. "The evolution of KDD: Towards Domain-Driven Data Mining." *International Journal of Pattern Recognition*. vol. 21. 1-16 p.

Majid, S., P. Yang, H. Lei and G. Haoran. 2014. *Knowledge Sharing by Students: Preference for Online Discussion Board vs Face-to-Face Class Participation*. In The Emergence of Digital Libraries – Research and Practices. Kulthida Tuamsuk,Adam Jatowt and Edie Rasmussen (Eds.). 149-159 p. vol. 8839. Cham. Springer International Publishing.

Marsh, H. W. 1984. "Students' evaluations of university teaching: Dimensionality, reliability, validity, potential baises, and utility." *Journal of Educational Psychology*. vol. 76 nº 5, 707 p.

Marshall, G. 2005. "The purpose, design and administration of a questionnaire for data collection." *Radiography*. vol. 11. nº 2, 131-136 p.

Merceron, A. and K. Yacef. 2008. "Interestingness Measures for Associations Rules in Educational Data." *EDM*. vol. 8. 57-66 p.

Minami, T. and Y. Ohura. 2013. "Investigation of Students' Attitudes to Lectures with Text-Analysis of Questionnaires". *Advanced Applied Informatics (IIAIAAI) International Conference*, (Los Alamitos, CA), 56-61 p. IEEE.

Oliviera Rezende Solange , A. M. E., Lika Fujimoto Magaly , Akemi Carvalho Reberta 2009. *Combining Data Driven and User Driven Evaluation measures toidentify interesting rules*. In Postmining of association rules. Ramamohanarao Kotagiri Jean-Francois Boulicaut, Jian Pei, Jaideep Srivastava, Philip S. Yu (Eds.). 38-55 p. vol. 1. IGI Global.

Padmanabhan, B. and A. Tuzhilin. 1999. "Unexpectedness as a measure of interestingness in knowledge discovery." *Decision Support Systems*. vol. 27. nº 3, 303-318 p.

Padmanabhan, B. T. A. 1998. "A Belief-Driven Method for Discovering Unexpected Patterns". *Third International Conference on Knowledge Discovery and Data Mining* (New York, USA),8 p.

Pan, q., L. Qu and L. Lou. 2009. "Data mining and application of teaching evaluation based on association rules". *Computer Science & Education, ICCSE. 4th International Conference*, (Nanning, China), 1404-1407 p. IEEE.

Paquette Gilbert, L. M., Lundgren-Cayrol Karin. 2011. *The Mot + Visual Language for Knowledge Based Instructional Design* In Instructional Design: Concepts, Methodologies, Tools and Applications. Murray E. Jennex Steve Clarke, Annie Becker, Ari-Veikko Anttiroiko (Eds.). 697-717 p. vol. Hershey, PA. IGI Global.

Pillai, J. and O. Vyas. 2010. "Overview of itemset utility mining and its applications." *International Journal of Computer Applications*. vol. 5 nº 11, 9-13 p.

Ramesh Sharda, M. H. 2009. "Information Extraction from Interviews to Obtain Tacit Knowledge: A Text Mining Application". *Americas Conference on Information Systems (AMCIS) Proceedings*, (San Francisco, USA),11 p. AIS Electronic Library (AISeL).

Raymond, C., Y. Qiang and S. Yi-Dong. 2003. "Mining high utility itemsets". *Data Mining, 2003. ICDM 2003. Third IEEE International Conference*, (Melbourne, FL, USA), 19-26 p. IEEE.

Řehůřek, R. 2015. «Topic Modeling for humans». Online <https://radimrehurek.com/gensim/>. Accessed January 2014.

Rubin, H. J. and I. S. Rubin. 2012. *Qualitative Interviewing: The Art of Hearing Data.* SAGE Publications, 265 p.

Sandhu, P. S., D. S. Dhaliwal, S. N. Panda and A. Bisht. 2010. "An Improvement in Apriori Algorithm Using Profit and Quantity". *Computer and Network Technology (ICCNT), 2010 Second International Conference* (Bangkok, Thailand), 3-7 p. IEEE.

Sievert, C. and K. E. Shirley. 2014. "LDAvis: A method for visualizing and interpreting topics". *Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces*, (Baltimore, USA), 63-70 p. Association for Computational Linguistics.

Silberschatz, A. and A. Tuzhilin. 1995. "On subjective measures of interestingness in knowledge discovery". *Proceedings of First International Conference on Knowledge Discovery and Data Mining, KDD-95*, (Menlo Park, CA, USA), 275-281 p. AAAI.

Silberschatz, A. and A. Tuzhilin. 1996. "What makes patterns interesting in knowledge discovery systems." *Knowledge and Data Engineering, IEEE Transactions*. vol. 8 nº 6, 970-974 p.

Tan, P.-N., V. Kumar and J. Srivastava. 2002. "Selecting the right interestingness measure for association patterns". In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, (Edmonton, Alberta, Canada), 32-41 p.ACM.

Tan, P.-N., V. Kumar and J. Srivastava. 2004. "Selecting the right objective measure for association analysis." *Information Systems*. vol. 29 nº 4, 293-313 p.

Tao, F., F. Murtagh and M. Farid. 2003. "Weighted Association Rule Mining using weighted support and significance framework". In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, (Washington, D.C.), 661-666 p.ACM.

University of Manitoba. 2015. «Center for the Advancement of teaching and learning». Online <http://intranet.umanitoba.ca/academic_support/catl/resources/learning/>. Accessed August 30, 2015.

Valot, C. and R. Amalberti. 1992. "Metaknowledge for time and reliability." *Reliability Engineering & System Safety*. vol. 36 nº 3, 199-206 p.

Viera, A. J. and J. M. Garrett. 2005. "Understanding interobserver agreement: the kappa statistic." *Fam Med*. vol. 37 nº 5, 360-363 p.

Wang, J., Y. Liu, L. Zhou, Y. Shi and X. Zhu. 2007. *Pushing frequency constraint to utility mining model*. In Computational Science–ICCS 2007. Geert Dick van Albada Yong Shi, Jack Dongarra, Peter M. A. Sloot (Eds.). 685-692 p. vol. 4489. Springer Berlin Heidelberg.

Wang, W., J. Yang and P. S. Yu. 2000. "Efficient mining of weighted association rules (WAR)". *Proceedings of the sixth ACM SIGKDD international conference on Knowledge discovery and data mining*, (New York, USA),270-274 p. ACM.

Wen, S., T. Rong, W. Jian, L. Hongmei and L. Jianye. 2014. *Applying data mining to improve faculty evaluation for educational organizations*. In Engineering Technology and Applications. Wise Shu Fun Shao, and Tracy Tian (Eds.). 183-189 p. vol. 28. CRC Press.

Wichita State University. 2015. «Student Perception of Teaching Effectiveness, SPTE». Online <http://webs.wichita.edu/?u=9112ssrl&p=/spte/spte_home/>. Accessed 30 August 2015.

Xin Chen, Y.-F. W. 2006. "Personalized Knowledge Discovery: Mining Novel Association Rules from Text". *Proceedings of the Sixth SIAM International Conference on Data Mining*, (Maryland, USA),589-593 p. Society for Industrial and Applied Mathematics.

Xuan-Hiep, H., F. Guillet and H. Briand. 2006. "Extracting representative measures for the post-processing of association rules". *Research, Innovation and Vision for the Future, 2006 International Conference*, (Ho Chi Minh City, Vietnam),100-106 p. IEEE.

Yang, H.-H., J.-C. Yu and H.-J. Yang. 2009. *Toward an Understanding of Knowledge Reuse in an On-Line Environment*. In Proceedings of the European Computing Conference. Nikos Mastorakis,Valeri Mladenov and Vassiliki T. Kontargyri (Eds.). 245-259 p. vol. 28. Springer US.

Yang, Q. 2009. *Post-processing Data Mining Models for Actionability*. In Data Mining for Business Applications. Longbing Cao,Philip S. Yu,Chengqi Zhang and Huaifeng Zhang (Eds.). 11-30 p. vol. 1. Springer US.

Yao, H., H. J. Hamilton and C. J. Butz. 2003. "A Foundational Approach to Mining Itemset Utilities from Databases". *SDM*, (San Francisco, CA),215-221 p. SIAM.

Yi-Dong, S., Z. Zhong and Y. Qiang. 2002. "Objective-oriented utility-based association mining". *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference*, (Maebashi City, Japan),426-433 p. IEEE.

Zhang, Z. and Z. Wang. 2012. "Non-teacher factors may affect teaching evaluation from students". *Robotics and Applications (ISRA), 2012 IEEE Symposium*, (Kuala Lumpur), 89-92 p. IEEE.