

Chemistry Education Research and Practice

Accepted Manuscript



This is an *Accepted Manuscript*, which has been through the Royal Society of Chemistry peer review process and has been accepted for publication.

Accepted Manuscripts are published online shortly after acceptance, before technical editing, formatting and proof reading. Using this free service, authors can make their results available to the community, in citable form, before we publish the edited article. We will replace this *Accepted Manuscript* with the edited and formatted *Advance Article* as soon as it is available.

You can find more information about *Accepted Manuscripts* in the [Information for Authors](#).

Please note that technical editing may introduce minor changes to the text and/or graphics, which may alter content. The journal's standard [Terms & Conditions](#) and the [Ethical guidelines](#) still apply. In no event shall the Royal Society of Chemistry be held responsible for any errors or omissions in this *Accepted Manuscript* or any consequences arising from the use of any information it contains.

ARTICLE

A Data Mining Approach to Study the Impact of the Methodology Followed in Chemistry Lab Classes on the Weight Attributed by the Students to the Lab Work on Learning and Motivation

Cite this: DOI: 10.1039/x0xx00000x

Received 00th January 2012,
Accepted 00th January 2012

DOI: 10.1039/x0xx00000x

www.rsc.org/

M. Figueiredo,^a L. Esteves,^b J. Neves^c and H. Vicente^d

This study reports the use of data mining tools in order to examine the influence of the methodology used in chemistry lab classes, on the weight attributed by the students to the lab work on learning and own motivation. The answer frequency analysis was unable to discriminate the opinions expressed by the respondents according to the type of the teaching methodology used in the lab classes. Conversely, the data mining approach using k-means clustering models, allowed a deeper analysis of the results, i.e., enabled one to identify the methodology to teach chemistry that, in students' opinion, is important for learning chemistry and increasing their motivation. The sample comprised 3447 students of Portuguese Secondary Schools (1736 in the 10th grade; 1711 in the 11th grade). The k-Means Clustering Method was used, with k values ranging between 2 and 4. The main strengths of this study are the methodological approach for data analysis and the fact that the sample was formed by students with different school careers that enables the use of the individual as the unit of analysis.

Introduction

Chemistry plays a major role in developing economic growth and improving the quality of life. Many of the breakthroughs in areas like health and medicine, food and agriculture, energy and the environment have been heavily dependent on advances in chemical knowledge. Chemistry is also essential in many other industrial applications like the aerospace or the electronics sector. Indeed, the new developments in nanotechnology and materials have chemistry at their core. From the economic point of view chemistry is regarded significant for Gross Domestic Product (GDP) generation. For instance, the Oxford Economics Report (2010) refers that the UK's chemical sector contributed for 21% of UK GDP and supported over 6 million jobs. This framework can justify the commitment in the training in chemistry made by the developed countries and the countries seeking to develop economically. Nevertheless, in the member countries of the Organization for Economic Cooperation and Development (OECD), close to 40% of high school students who come top in science subjects have no interest in pursuing a science related career, while almost 45% do not want to continue studying science. This scenario is worrying for countries, like Portugal, that need to develop high level skills in order to drive productivity and innovation (OECD, 2009). This report emphasizes that schools do a reasonable job in transmitting science knowledge and skills but they fail to

engage students in science and science related careers. From 2006 to 2012, in the OECD countries the science performance has remained broadly stable, although Portugal shows a slight improvement (OECD, 2014).

According to Bopegedera (2011) the majority of the chemistry students are interested in other courses like medicine, allied health fields, engineering or other sciences where chemistry is only a requirement to pursue their studies. However, to some students chemistry is still perceived as a challenging subject to study. The report for SCORE – Science Community Supporting Education (Coe *et al.*, 2008) presents results from a large number of studies conducted over a long period, using many different methods and datasets. These studies showed that chemistry is one of the most difficult subjects, hence, students need to be well motivated and to have good knowledge prior to commencing post compulsory study.

The practical work is widely and frequently used in the teaching of chemistry in secondary schools as a methodology of teaching and motivation (Millar, 2002). However, the term “practical work,” is commonly used in the literature as an overarching term that refers to any type of science teaching and learning activity in which students, working either individually or in small groups, are involved in manipulating and/or observing real objects and materials (Millar, 2010). From this point of view, practical work is a broad category that includes for example, “laboratory work” (or “lab work”).

The role of lab work in teaching and learning of chemistry

In the traditional instruction of the sciences predominate the lectures that aim to “deliver” ideas or information from the teacher to the students (Johnstone, 1993). However, this methodology is inappropriate for the study of most topics in chemistry because there are many skills involved in being a chemist, including observation, discussion, data-collection which cannot be developed in theoretical lectures. For this reason the laboratory work has a fundamental role in chemistry teaching, having been officially included in the curricula of Sciences since the nineteenth century. In fact, the relevance of lab work has been acknowledged by different authors over the last decades (Lock, 1988; Miguéns and Garrett, 1991; Gee and Clackson, 1992; Josephsen, 2003; Keiler and Woolnough, 2002; Hofstein and Lunetta, 2004; Millar, 2004; Hofstein and Mamlok-Naaman, 2007; Abrahams and Reiss, 2012). However, doubts were sometimes raised about its importance as a means for promoting significant learning of chemistry (Hodson, 1990, 1993). What role should lab work play in such a teaching process? Should lab work continue to be carried out in a traditional manner or why not give it new dimensions according to the fundamental role that, in our opinion, it could play? Although it is not a consensual opinion, many authors consider that lab work represents a fundamental resource in the teaching of Science (Johnstone and Al-Shuaili, 2001), and different types of lab work with different objectives, leading to dissimilar learning outcomes, should be carried out (Cheung, 2007; Logar and Savec, 2011).

Lab work methodologies

Conducting lab work may follow different methodologies. The methodology that most limits the role of the students are the demonstrations done by the teacher (Logar and Savec, 2011). However, the demonstrations are still used as lab work methodology in some schools, due to the lack of material resources. In this case the students have no opportunity to develop any of the skills usually presented as advantages to adopt the realization of lab work in schools. Despite being a passive methodology that may limit the potential for student learning it is very useful and it is often used in lectures to illustrate specific subjects.

A second type of methodology consists in carrying out the lab work by students according to recipes executed step-by-step. Students focus their thoughts on finishing one step after another and many times they do not develop a deeper understanding of the experiments. For many students lab work means just working, handling laboratory equipment, not including, in many cases, the development and the understanding of scientific thinking (Hofstein and Lunetta, 2004).

In a third type of methodology, the students that conducted experiments based on a receipt made by themselves, under teacher guidance, were frequently more motivated for the subject. Laboratory works developed based on constructivism had great role for increasing students' learning achievements

and developing students' positive attitudes towards chemistry laboratory, in contrast to traditional teacher centred approach (Tarhan and Sesen, 2010). From this perspective it makes sense to introduce a problem and guide the students in finding solutions. In fact, this methodology implies fundamental steps in the teaching/learning process of tentative skills, like collecting information or doing planning, promoting the acquisition of key abilities. Some studies show that learning experiences based on concrete situations are authentic, meaningful, challenging, and based on the choice and on the students' work, not only increasing the intrinsic motivation of the students to learn Science (Yair, 2000; Koballa and Glynn, 2007), but also improving their attitudes towards Science and Learning (Sherz and Oren, 2006). Furthermore, it is not relevant to continue to use lab work only as a mere illustration of theories or as a means to train manipulative abilities, like measuring volumes or masses, although attaining accurate and precise results has been always quite desirable (Bennett and O'Neale, 1998).

Recent studies confirm that laboratory based learning quality is increased as students have an active role in the process of gaining knowledge (Cheung, 2007; Bennett *et al.*, 2010; Kind *et al.*, 2011). There are several methods that allow to explore this type of learning like class research seminars, problem based learning, case studies, project-based learning, role playing, cooperative and cooperation learning, group debate, development of mind maps, experience based learning. However, Bopegedera (2011) points out the importance of the connections between theory (presented in the textbook and lectures) and practice (in the laboratory and problem-solving workshops) to provide a holistic learning experience. Indeed, students need a good balance between teacher guidance and independent thought.

Students' motivation to learning chemistry

Besides the advantages that conducting lab work represents, in terms of learning, lab work may also influence the student motivation. Recently, several studies have been published, aiming to investigate the role of motivation in knowledge transfer (Pugh and Bergin, 2006; Nokes and Belenky, 2011; Engle, 2012; Perkins and Salomon, 2012; Nokes-Malach and Mestre, 2013; Richey and Nokes-Malach, 2013). Other researchers investigated the effects of students' achievement motivation on learning outcomes (Ames and Archer, 1988; Dweck, 1986; Elliot *et al.*, 1999; Harackiewicz *et al.*, 2002; Grant and Dweck, 2003). These studies have examined the relationships between different types of achievement goals, learning and the motivational outcomes. The guidelines for considering these goals depend on two main vectors, i.e., how a person defines competence and the valences for achieving that competence.

The studies about motivation for learning usually distinguish between intrinsic and extrinsic motivation (Stipek 1996, 2014). The former is understood to be a personal interest in pursuing a goal without any palpable reward, i.e., the goal is considered to be an own wish and not required by external

agents. Extrinsic motivation, in pursuit of a task, is required or directed by external factors not on the basis of the own wishes. Learning is more likely to be significant when it increases the degree of intrinsic motivation that leads to personal fulfilment. Indeed, some studies point out the many advantages for students who enjoy learning compared with those who do that because they feel they must achieve extrinsic rewards or avoid punishment (Stipek, 2014). According to this author, students who enjoy learning for their own sake seem to learn at more conceptual levels, seek intellectual challenges more frequently, and persist longer during difficult tasks than students who focus on external rewards and punishments. Considering the fact that in schools learning is something that is imposed to students, is promoted primarily by factors of extrinsic motivation, like educational attainment, progression to the next level, among others. These factors are often overvalued both in school and in the family context (Heyman and Dweck, 1992; Stipek, 2014). Nevertheless, the role of extrinsic motivation cannot be undervalued in the educational context. Indeed, through appropriate stimuli (i.e., teaching strategies) the teacher can help the student to redefine goals, attributions, interests and/or self-concepts. The relations between extrinsic stimuli and motivation are neither linear nor uncomplicated, since these stimuli can trigger/influence in different ways the students' motivation. Some studies about students' motivation refer five criteria to distinguish between intrinsic and extrinsic motivation (Harter 1981; Shachar and Fischer, 2004), namely preference for challenge; curiosity and interest; independent mastery; independent judgment and internal criteria. These criteria are linked, respectively, to the questions "*Does the student like hard challenging work as opposed to easier assignments?*"; "*Does the student work to satisfy his/her own interest and curiosity rather than to satisfy the teacher?*"; "*Does the student prefer to acquire his own skills of logical thinking instead of relying on the teacher for help and guidance?*"; "*Does the student prefer self-directed learning instead of learning directed by the teacher?*"; and "*Does the student know when he/she has succeeded or failed on school assignments instead of being dependent on external evaluation only?*".

Recent studies have shown that the use of diversified teaching strategies can significantly increase the intrinsic motivation of students. Baeten *et al.* (2012) shows the importance of gradually introducing students to case-based learning, in terms of their autonomous motivation and achievement. A study conducted by Changeiywo *et al.* (2011) highlights that students exposed to mastery learning approach have significantly higher motivation than those taught through regular methods.

In this context the completion of lab work emerges as a factor in the students you want to unleash the mechanisms of intrinsic motivation for learning science, especially chemistry.

Knowledge discovering from databases

In recent years, the advances in information technologies have made it possible to collect and store a large volume of data. In fact, the high amount of data stored has far exceeded the human

ability for analysis, interpretation and comprehension. Thus, in the last years, the use of data mining tools has become mandatory. The process of knowledge extraction (that includes a data mining stage) is designated Knowledge Discovery from Databases (KDD). The designation KDD was formally adopted in 1989 and refers to a process that involves the identification and recognition of patterns in a database, in an automatic way, i.e., obtaining relevant, unknown information, that may be useful in a decision making process, without a previous formulation of hypothesis. This process of knowledge extraction may be oriented to attain different objectives (e.g. classification, clustering, forecasting, optimization or summarization), and may entangle different undertakings (e.g. selection, pre-processing, transformation, data mining and interpretation) as depicted in Fig. 1. Indeed, the data mining stage is the core of the process of KDD and is centred in the application of algorithms that cater for the identification and the recognition of patterns from large volumes of data (Klosgen and Zytkow, 2002; Han *et al.*, 2011).

The classical KDD application areas include, among others, marketing, finance, fraud detection, manufacturing, telecommunications, internet or medicine (Han *et al.*, 2011; Witten *et al.*, 2011). Recent studies show the applicability of KDD to other areas like production of water to human consumption (Pinto *et al.*, 2009; Couto *et al.*, 2012) or prediction of the availability of nitrogen in soils (Nunes *et al.*, 2012). Regarding education research, data mining is still considered a new paradigm and a promising challenge. Indeed, educational data mining can be considered an emerging theme, concerned with developing methods for exploring the various types of data that come from the educational context. A few studies that illustrate the applicability of these tools to different problems in educational field can be found in literature.

A specific application of data mining tools in learning management systems was presented by Romero *et al.* (2008). The main objective of this study was to classify students into different groups with equal final marks depending on the activities carried out in Moodle. The C4.5 algorithm was used to induce decision trees and a set of interesting rules were obtained by the authors. For instance, students with a low number of quizzes passed were classified as *fail*. Students with a high number of passed quizzes are directly classified as *excellent*. Finally, students with a medium number of passed quizzes are classified as *fail*, *pass* or *good* depending on other variables like *total time of assignments*, *number of quizzes*, *number of quizzes failed*, or *number of assignments*. The knowledge discovered can be used by the instructor in different ways. On the one hand, to classifying new students in order to detect early students with learning problems and, on the other hand, to decide about the use of some types of activities that conduce to higher marks, or on the contrary, to decide to eliminate some activities related to low marks. Also in the scope of distance education platforms, Sevindik and Cömert (2010) compare different data mining algorithms like *k*-Means, Apriori, C4.5, Support Vector Machines, *k*-Nearest Neighbours and Naive Bayes. According to the authors the algorithm C4.5,

used to induce decision trees, shows to be the more effective one in classifying students' characteristics and academic success and can be used by the teacher to anticipate possible scenarios and avoid academic failure.

Şen and Uçar (2012) developed a process of knowledge discovery from databases, using artificial neural networks and decision trees, in order to study the students' achievements. The input variables were *gender, age, type of high school graduation, education type* (i.e., distance/regular) and *lesson type*, while the output variable was the *students' scores*. Both classification methods exhibited values of overall accuracy higher than 94%. The results show that the students' success rate has inverse ratio with students' age and the success score decreases with increasing age. Another interesting feature is related with type of education. The students with best scores (ranging between 80 and 100) are studying in the formal education while the students with scores varying between 65 and 80 are studying in the distance education. The study also shows that the scores less than 60 were obtained mainly by students in the distance education.

Şen *et al.* (2012) developed models to predict secondary education placement test results using C5 decision tree algorithm, vector machines, artificial neural networks and logistic regression. The overall accuracy of models ranges between 82% (logistic regression model) and 95% (decision tree model). The authors used 24 input variables that include, among other, *gender, marital status* or the *scores* obtained by the students in various subjects like *mathematics, science and technology* or *foreign language*. An import aspect of this study, that should be noted, was related with the sensitivity analysis performed on the models in order to determine the importance of the input variables. The sensitivity analysis showed that *previous test experience*, whether a student has a *scholarship, number of siblings, previous years' grade point average* are among the most important predictors of the placement test scores. Undeniably, knowing the factors that more directly or indirectly affect achievement is valuable to all actors involved in the educational process (i.e., students, parents, teachers, administrators) in order to maximize success.

Neves *et al.* (2015) and Figueiredo *et al.* (2014) present the development of decision support systems to evaluate the quality of learning and to evaluate potential situations of school dropout, respectively. These systems were built under a formal framework based on Logic Programming, in terms of its knowledge representation and reasoning procedures, complemented with an approach to computing grounded on Artificial Neural Networks. This approach not only allows to obtain the evaluation of quality of learning (or school dropout risk) but it also permits the estimation of the confidence that one has on the model prediction.

All these studies exemplify the use of different data mining algorithms (e.g. cluster analysis, decision trees, association rules, support vector machines, artificial neural networks) and illustrate the potential and the central role that such tools could play in the educational context.

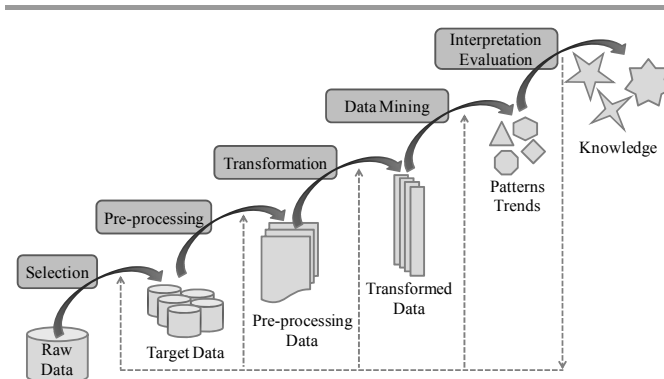


Fig. 1 Data mining as a stage in the process of knowledge discovery in databases

Integration of the study into the secondary school context

In Portugal, until the academic year of 2003/2004 the role of laboratorial work in chemistry teaching was greater, since the subjects Laboratorial Techniques I, II and III were included in the secondary school level curriculum, allowing students to acquire practical skills. Currently, the secondary school curriculum includes only six sessions of 135 minutes per academic year to perform lab work. In this scenario one of the main teachers' challenges is how to conduct the laboratory work in order to promote a more effective learning and reinforce the students' motivation. In this context, the methodology followed by the teacher is of utmost importance to achieve these goals.

Study aims

The present work reports the use of data mining tools in order to examine the influence of the methodology used in chemistry lab classes on the weight attributed by the students to the lab work on learning and own motivation.

The major contribution of this work is related with the methodological approach and the use of data mining tools for data analysis. In other words, one of the strengths of this study lays in the fact that the sample was formed by students with different school careers, and consequently exposed to different teaching methodologies in lab classes. Through the use of data mining tools it was possible to correlate the importance attributed by the students to the lab work on their learning and motivation with the teaching methodologies followed in their lab classes, considering their integral school careers. Indeed, in the studies present in literature that address and discuss the problem of the influence of the methodologies followed in lab classes on learning and/or motivation of students, the sample is usually formed by a specific group of students submitted to the same methodology, i.e., homogeneous samples, in the sense that all students have the same educational experience where the unit of analysis is the group (not the individual).

Methods

Study design

PARTICIPANTS. The Portuguese educational system at the secondary school level comprises three years of high school (10th to 12th grade). Chemistry is only taught as a separate subject at the 12th grade. Physics and chemistry are combined into a single subject, called physics and chemistry, at the 10th and 11th grades. Chemistry is taught during half of the school year and physics is studied during the other half. The subject is taught in three sessions per week, two of them with 90 minutes and the remaining one with 135 minutes, where the class is divided into two groups, planned for the realization of practical work. The curriculum is the same for all schools in the country and imposes some specific laboratory activities. A total of 3447 physics and chemistry students comprising 1736 of the 10th grade (15–16 years old) and 1711 of the 11th grade (16–17 years old) were enrolled in this study. The students came from Portuguese secondary schools located in the north (districts of Bragança and Oporto), centre (districts of Castelo Branco and Lisbon), and south (districts of Beja, Évora and Faro). The districts of Beja, Bragança, Castelo Branco and Évora are situated in the interior region of the country, while the remaining ones are located in the coastal line.

SAMPLE CHARACTERIZATION. Table 1 shows the sample characterization in terms of age, gender, grade and district. With regard to the 10th grade sample, a perusal of Table 1 reveals that 45.5% of students are male and 54.5% are female. Concerning the students' age, 85.7% of them did not exceed 16 years old, which suggests that the rate of school failure is low. The geographical location seems not to influence the results, since the percentage of this class of students varies between 79.9% (district of Bragança) and 87.7% (district of Oporto).

Concerning the 11th grade sample, a glance to Table 1 shows that 46.8% of students are male and 53.2% are female. Regarding the students age, only 35.4% of them do not exceed 17 years old, which insinuates high rate of school failure. The geographical location seems not to influence the results since

there are no significant differences between the percentages of students under 17 (seventeen) years old (lies among 33.1% in the district of Évora and 37.4% in the district of Faro).

ETHICAL ASPECTS OF THE STUDY. Students participated in the study voluntarily without any pressure or coercion and were informed that their grades would not be affected. Each of the participants gave an informed consent form to participate in the study. The study was conducted in compliance with the relevant laws and institutional guidelines, and was approved by the relevant authorities.

ENDOGENOUS ASPECTS OF THE STUDY. The curriculum of the discipline of physics and chemistry imposes a set of obligatory laboratory activities. The mandatory nature of these activities eliminates the problems of endogeneity that could influence the study by introducing other variables beyond the teaching methodology.

DATA COLLECTION. In order to fulfil the goals defined so far, a versatile tool to data collection was essential, with the potential to be used in a wide geographical area and on time (DeKetele and Roegiers, 2009; Cohen *et al.*, 2011). After considering and analyzing the advantages and limitations intrinsic to the various techniques available (McMillan and Schumacher, 2009), a practice based on the inquiry by questionnaire was chosen. Indeed, this kind of instrument has a well-defined structure and allows for the conversion of the information reported by the respondents, into a quantitative one. The questions included in the questionnaire were planned, on the one hand, to allow for the gathering of information on the learning methodologies followed in the lab classes and, on the other hand, to scrutinize the influence of such methodologies on chemistry learning and on the students motivation.

DATA ANALYSIS TOOLS. Beyond the answers frequency analysis, knowledge discovery from databases was the strategy followed to treat the experimental results. In data mining stage a cluster analysis was carried out. The technique used to induce clusters was the k-means clustering method (MacQueen, 1967), implemented using the software WEKA (Hall, *et al.*, 2009).

Table 1 Sample characterization in terms of age, gender, grade and district

District	10 th Grade					11 th Grade								
	Age					Gender		Age					Gender	
	<15	15	16	17	>17	F	M	<16	16	17	18	>18	F	M
Beja ^{c,*}	0	71	52	17	10	89	61	2	7	37	13	64	72	51
Bragança ^{b,*}	0	43	68	21	7	75	64	0	0	43	25	37	43	62
Castelo Branco ^{b,*}	0	62	46	11	3	64	58	1	3	48	17	59	67	61
Évora ^{c,*}	1	52	43	15	9	66	54	0	0	47	16	79	85	57
Faro ^{c,+}	0	89	58	8	16	92	79	2	5	57	27	80	98	73
Lisbon ^{b,+}	1	292	241	48	31	341	272	1	4	226	77	316	326	298
Oporto ^{a,+}	0	204	165	32	20	219	202	1	3	149	28	237	219	199

Portugal regions – ^a north; ^b centre; ^c south; * interior; + coastal.

1
2
3
4
5
6
7
8
9
10
11
12
Some implementations of k -means only allow numerical values for attributes, (i.e., it may be necessary to convert categorical attributes). However, this is not necessary for clustering in WEKA since the WEKA Simple k -Means algorithm automatically handles a mixture of categorical and numerical attributes. In addition, the algorithm also normalizes the numerical attributes automatically, when computes the Euclidean distance. A more detailed description of the WEKA Simple k -Means algorithm can be found in Witten *et al.* (2011); Sharma(Sachdeva) *et al.* (2012).

13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60

Questionnaire

In order to collect data a questionnaire was designed specifically for this study (presented in Supplemental Materials section). The questions included in the questionnaire were organized into three sections, aiming the characterization of the sample, the characterization of the methodology followed in chemistry lab classes, and finally to collect the opinion of students about the importance of lab work on chemistry learning and on own motivation. The former section includes the questions related with age, gender, grade and provenience (i.e., district). The second one comprises the questions *Q1 – Who does the lab work?*; *Q2 – How are the students organized in the lab classes?*; *Q3 – Which is the basis of the lab work?* and *Q4 – What type of post-lab work is done?*. The latest includes the issues *Q5 – How do you classify the importance of lab work on the learning of Chemistry?*; and *Q6 – How do you classify the importance of lab work to increase your motivation to study Chemistry?*.

The validation of the questionnaire respects the practices recommended by Bell (2010). Subsequently, the questionnaire was evaluated by a group of experts that suggested some amendments. As soon as these revisions were done, the questionnaire was applied to a small group of students of both grades, not included in the sample, to evaluate the validity of the questionnaire and identify possible difficulties in the interpretation of the questions. The questionnaire was sent by mail to the schools that indicated their willingness to participate in this move. To ensure that the answers reflect the whole school career and were not influenced by the work developed in the present academic year, the questionnaires were applied in the beginning of academic year. Thus, in this study only the responses received until 31 October were considered, i.e., five weeks after the beginning of the school year. The return rate for the samples related to the 10th and 11th grades were, respectively, 39.2% (1776 questionnaires received in 4530 sent), and 34.5% (1754 questionnaires received in 5085 sent).

Knowledge discovering process

SELECTION. This is the first phase of the knowledge discovering process (Fig. 1). In this phase the problem was analyzed and the main goals were defined. As mentioned above, the present work seeks to study the influence of the methodology used in chemistry lab classes on the weight attributed by the students to the lab work on learning and own

motivation. Thus, based on the answers about the methodology followed in lab classes, the data mining goal was defined as the search of different clusters, so that the similarity is maximized within cluster and is minimized between clusters. To ensure that the clusters were formed based on the methodology followed in chemistry lab classes the input variables used were exclusively the answers to the questions related with the methodology (issues Q1, Q2, Q3 and Q4).

PRE-PROCESSING. The first step of this stage is the transposition of the data collected by questionnaires into quantitative data, in order to create the database. The data transposition process allowed the identification of anomalies in the answers to the questionnaires (e.g., missing and/or multiple answers to a given question). These records were discarded from this study. The database comprises a total of 3447 records (1736 concerning the 10th grade and 1711 regarding the 11th grade) and ten fields.

TRANSFORMATION. The data must be transformed to the required format of the algorithm. Thus the database was saved in a text file with ARFF format (Witten *et al.*, 2011).

DATA MINING. Clustering is one of the most appropriate assignment on the data mining phase for uncovering groups and identifying thought-provoking distributions and patterns in data. Clustering models focus on identifying groups of similar objects (respondents in the present study), and label the objects according to the group (i.e., the cluster) to which they belong. This is done without the use of prior knowledge about the groups and their characteristics. These models are often referred to as unsupervised learning models, since there is no external standard by which to judge the model's performance. Their value is determined by their ability to capture thought-provoking grouping in the data and provide useful descriptions of those assemblages, taking into account the goals set (Han *et al.*, 2011).

The basic idea in the k -means clustering method is to try to discover k clusters, according to the requirements:

- Each cluster must contain at least one object; and
- Each object must belong to exactly one cluster.

The k -means algorithm input parameters stand for the number of clusters, k , and a data set, D , with n objects. As soon as the algorithm is enforced, it selects, randomly, k points that denote the initial centres of the clusters, being the objects assigned to the cluster to which they are akin done according to the Euclidean distance between the objects and the cluster midpoint (Bradley and Fayyad, 1998). Next, the algorithm computes the new centre for each cluster. These processes iterate until further refinement may no longer improve the model or the number of iterations exceed a specified limit (Fig. 2). In this study k varied from 2 to 7 and the iterative process was stopped whenever the additional refinement does not improve the model.

INTERPRETATION/EVALUATION. Since in data mining stage the clusters were formed without the use of prior knowledge about the groups and their characteristics it is mandatory to understand how the clusters were formed. To achieve such goal

were used Decision Trees (DTs). DTs have many attractive features, such as allowing human interpretation, and hence making it possible for a decision maker to gain insights into what factors are critical for a particular classification. DTs adopt a branching structure of nodes and leaves, where the knowledge is hierarchically organized. Each node tests the value of a feature (i.e., an input variable), while each leaf is assigned to a class label (a cluster in the present study).

The basic strategy employed to generate DTs is the so-called recursive partitioning approach to problem solving. It works by partitioning the examples by choosing a set of conditions on an independent variable (i.e., an input variable), such that an error on the dependent variable (i.e., the output variable) is minimised within each group. The process continues recursively inside each subgroup until certain conditions are met, like the ones where the error cannot be further reduced (Han *et al.*, 2011).

Sometimes, it is useful to build a rule-based classifier by extracting IF-THEN rules from the DTs. The rule is created at each path, from the root (the first node of DT) to a leaf. Each splitting criterion along a given path is logically ANDed to form the rule antecedent (the IF part). The leaf node holds the class prediction, forming the rule consequent (the THEN part).

Early systems for generating DTs include CART (Breiman *et al.*, 1984) and ID3 (Quinlan, 1986), the latter being followed by the version C4.5 and C5.0. The C4.5 version was an improvement of the ID3 algorithm that allows the use of continuous values, support omitted values and tree pruning (Quinlan, 1993). The DT algorithm used in this study was the J48 as implemented in WEKA (Hall, *et al.*, 2009). This J48 implements the 8th revision of the C4.5 algorithm. A description of the J48 algorithm can be found in Witten *et al.* (2011).

Results and discussion

Answer frequency analysis

This section presents the frequencies of answers to each question included in the questionnaire. In order to examine if the answers to the questionnaire are influenced by the geographical location of the schools, an attempt was made to relate the student's replies to the questions presented above, within the various districts. The results are depicted in Fig. 3 and show that the answers are not influenced by the geographical location of the schools and, therefore, may be analyzed together. Since the answers were not influenced by the geographical localization, an answer frequency analysis was performed (Fig. 4). With this methodology of data treatment it is only possible to state that:

- A small percentage of respondents claim that the lab work is done exclusively by the teacher;
- The answers related with the organization of the students in lab classes are not conclusive;
- About 80% of the students declare that the lab work is based on experimental guidelines; and
- About 75% of the respondents state that the post-lab work consists on the elaboration of written reports.

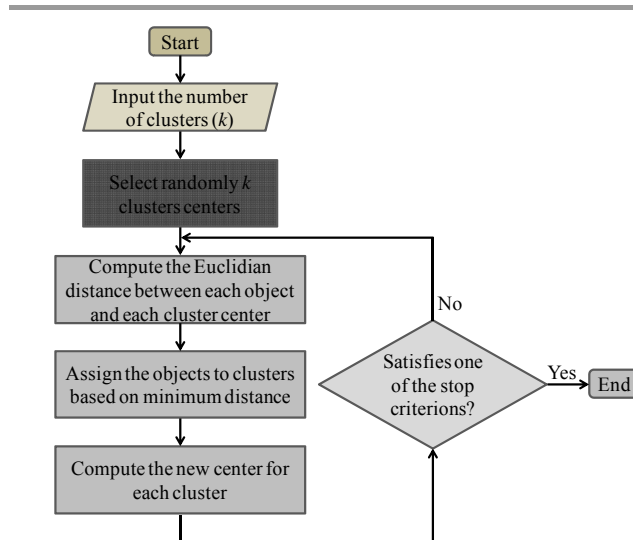


Fig. 2 *k*-means clustering process

The analysis of Fig. 4 also denotes, for both grades, that in the opinion of students the importance of lab work for learning chemistry and to increase their own motivation is very high or high. Only a few respondents answer moderate, low or very low. However, the percentages of the most positive responses were higher in the 10th grade, which may be attributed to two factors. The former one is linked to the fact that in the 11th grade the quantitative treatment and the discussion of the results obtained in lab work are deeper, and require the knowledge and the skills already acquired in subjects like chemistry, physics and mathematics. The latest is linked to the fact that, as previously noted, some of these students may be repeating the attendance on the subject, and that lab work does not constitute novelty and does not influence, to the same extent, either the importance attributed by the students to the lab work in learning or the motivation to study chemistry.

Since the overwhelming majority of the respondents declares that the importance of lab work for learning chemistry and to increase their own motivation is very high or high, it is impossible to conclude about the influence of the methodology used in chemistry lab classes, on the weight attributed by the students to the lab work on learning and own motivation. In order to overcome these limitations a methodology of data analysis based on cluster analysis and decision trees was carried out. The *k*-means clustering method is one of the most efficient data mining algorithms that seek to identify groups of similar objects (i.e., respondents) in complex samples. Decision trees were used to understand how the clusters were formed. To ensure that the clusters are formed based on the methodology followed in chemistry lab classes the input variables used are exclusively the answers to the questions related with the methodology (issues Q1, Q2, Q3 and Q4).

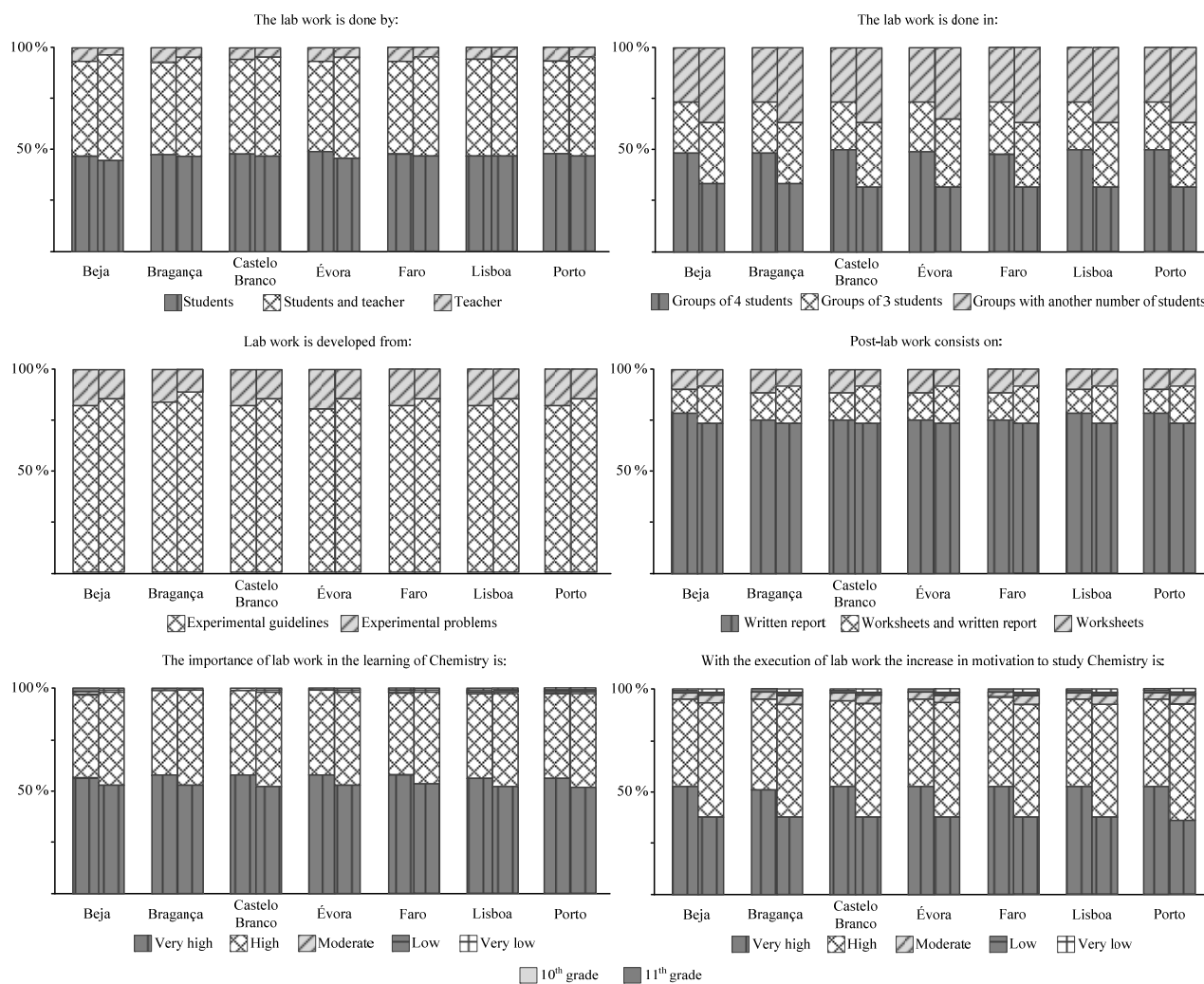


Fig. 3 Frequencies of answers to the questions “Who does the lab work?”, “Which is the basis of the lab work?”, “How are the students organized in the lab classes?”, “What type of post-lab work is done?”, “How do you classify the importance of lab work on the learning of Chemistry?” and “How do you classify the importance of lab work to increase your motivation to study Chemistry?” split by districts

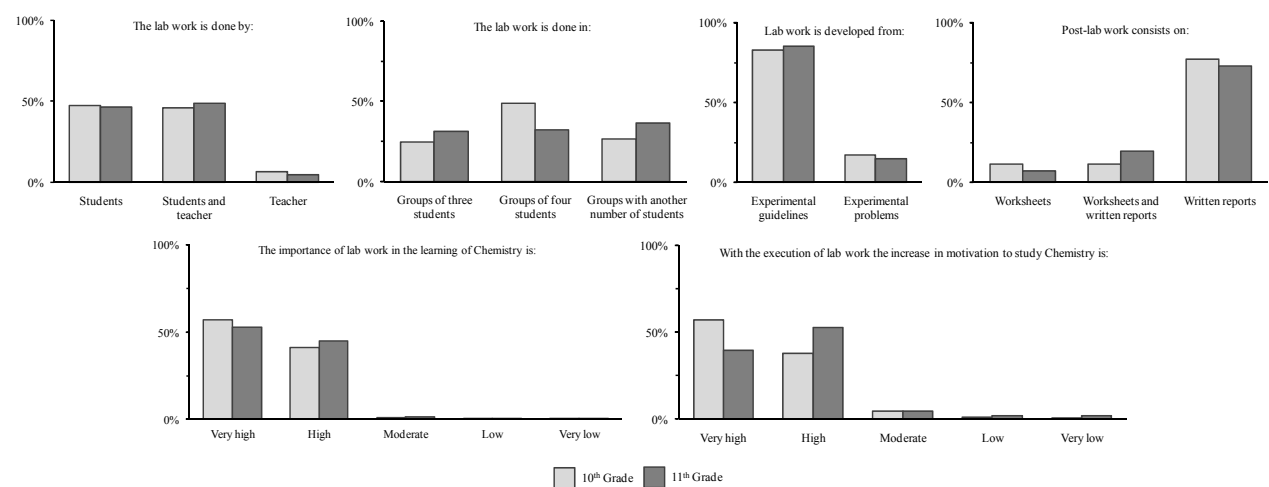


Fig. 4 Frequencies of the answers given to each question by the respondents of the 10th and the 11th grades

Clustering models – Assessment and interpretation

The starting point of k -means clustering algorithm is the choice of number of clusters. In present study were tested various values of k , ranging from 2 to 7. For k greater than 4 the number of respondents in some clusters was very low, less than 0.1% of the total number of cases. For this reason these models were discarded. Table 2 presents the results of the various models, with different number of clusters, considered in this study. Regarding the responses obtained in the questionnaire linked with the 10th grade, the analysis of Table 2 shows that the $k = 3$ and the $k = 4$ clustering models are quite similar. The main difference is the division of cluster 2 of the $k = 3$ clustering model into cluster 2 (with 346 objects) and cluster 4

(with 425 objects), into the $k = 4$ clustering model. For both models, Table 2 further reveals that cluster 1 includes only students who claim that the lab classes are developed from tentative situations Regarding cluster 3, this is made upon students that assert that the lab classes are done exclusively by themselves The splitting of cluster 2 (model of three clusters) into two clusters (model of four clusters), enable to group a part of the students that reported that the lab classes are done sometimes by them and occasionally by the teacher, into cluster 4. With respect to the model of two clusters, a glance of Table 2 shows that cluster 1 was formed by the students that claim that the lab classes are done always by themselves, while cluster 2 comprises the students that reported the opposite.

Table 2 Answers obtained on the questionnaire with respect to the 10th grade, split by issues and by clustering models

Who does the lab work?										
		$k = 2$		$k = 3$			$k = 4$			
		cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4
Students		821	0	156	0	665	156	0	665	0
Students and teacher		0	797	117	680	0	117	255	0	425
Teacher		0	118	27	91	0	27	91	0	0
How are the students organized in the lab classes?										
		$k = 2$		$k = 3$			$k = 4$			
		cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4
Groups of 3 students		232	195	78	167	182	78	75	182	92
Groups of 4 students		409	442	132	382	337	132	171	337	211
Groups with another number of students		180	278	90	222	146	90	100	146	122
Which is the basis of the lab work?										
		$k = 2$		$k = 3$			$k = 4$			
		cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4
Experimental guidelines		665	771	0	771	665	0	346	665	425
Experimental problems		156	144	300	0	0	300	0	0	0
What type of post-lab work is done?										
		$k = 2$		$k = 3$			$k = 4$			
		cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4
Worksheets		80	119	42	96	61	42	12	61	84
Written reports		651	685	197	604	535	197	325	535	279
Worksheets and written reports		90	111	61	71	69	61	9	69	62

Regarding the responses obtained in the questionnaire related with the 11th grade, an examination of Table 3 reveals that the $k = 2$ and the $k = 3$ clustering models are fairly analogous. The main difference is the separation of cluster 2 from the $k = 2$ clustering model into cluster 2 (with 835 objects) and cluster 3 (with 83 objects), into the $k = 3$ clustering one. For both models, cluster 1 was formed by the students that declare that the lab classes are done always by the students, while the division of cluster 2 (model of two clusters) into two clusters (model of three clusters) enabled to differentiate the students that reported that the lab classes are done always by the teacher (cluster 3), of those that state that the classes are done sometimes by them and sometimes by the teacher (cluster 2). The $k = 3$ and the $k = 4$ clustering models are quite similar too. In this case, cluster 4, which comprises 713 objects, was formed from cluster 1 (lost 480 cases) and from cluster 2 (lost 233 objects) of the three clusters model.

Once presented the various models of segmentation and set the main differences among them, it is necessary to define criteria to evaluate them. Since there is no theoretical reason to judge the models' performance their value is determined by the models' ability to provide useful descriptions of the data, taking into account the goals set. Having in mind that the study object is to investigate the influence of the methodology used in chemistry lab classes, on the weight attributed by the students to the lab work on learning and own motivation, it is intended that the clusters obtained should be as homogeneous as possible in terms of methodology used in chemistry lab classes. Thus, for both grades, the models of three clusters were selected, since the models of two clusters contain, in cluster 2, two distinct answers to the question *Who does the lab work?*. Conversely, the model of four clusters seems to bring no improvement, once the introducing of a new cluster does not result in a gain of homogeneity.

Table 3 Answers obtained on the questionnaire with respected to the 11th grade, split by issues and by clustering models

Who does the lab work?										
	$k = 2$		$k = 3$			$k = 4$				
	cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4	
	Students	793	0	793	0	0	313	0	0	480
Students and teacher	0	835	0	835	0	0	602	0	233	
Teacher	0	83	0	0	83	0	0	83		
How are the students organized in the lab classes?										
	$k = 2$		$k = 3$			$k = 4$				
	cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4	
	Groups of 3 students	278	260	278	240	20	278	240	20	0
Groups of 4 students	247	303	247	268	35	0	35	35	480	
Groups with another number of students	268	355	268	327	28	35	327	28	233	
Which is the basis of the lab work?										
	$k = 2$		$k = 3$			$k = 4$				
	cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4	
	Experimental guidelines	689	769	689	696	73	226	463	73	696
Experimental problems	104	149	104	139	10	87	139	10	17	
What type of post-lab work is done?										
	$k = 2$		$k = 3$			$k = 4$				
	cluster 1	cluster 2	cluster 1	cluster 2	cluster 3	cluster 1	cluster 2	cluster 3	cluster 4	
	Worksheets	17	108	17	84	24	4	35	24	62
Written reports	630	619	630	570	49	63	400	49	581	
Worksheets and written reports	146	191	146	181	10	26	167	10	70	

Explanatory models of segmentation

The cluster analysis by itself is insufficient since it is not known how the clusters were formed. Just allows to state “ $X\%$ of the elements of cluster i assert that the importance of lab work in chemistry learning is Y ”. In order to generate explanatory models of segmentation (i.e., to know how the clusters were formed), Decision Trees (DTs) were used. The input variables were the answers to the questions related with the methodology (issues Q1, Q2, Q3 and Q4) and the output variables were the clusters formed.

To ensure statistical significance of the attained results, 20 (twenty) runs were applied in all tests. In each simulation, the available data is randomly divided into two mutually exclusive partitions, i.e., the training set, with two-thirds of the available data and used to construct the models, and the test set, with the remaining of the examples being used after training, in order to compute the accuracy values (Souza *et al.*, 2002). The DTs

obtained are shown in Fig. 5 and 6, respectively for the 10th and 11th grades. Concerning the 10th grade the rule regarding the cluster 1 is “the basis of lab work is Experimental Problems”. With respect to cluster 2 there are two rules. The former is “the basis of lab work is Experimental Guidelines and the lab work is done by Teacher”, while the second is “the basis of lab work is Experimental Guidelines and the lab work is done sometimes by the Students and sometimes by the Teacher”. These rules can be merged into “the basis of lab work is Experimental Guidelines and the lab work is **not done exclusively** by the Students”. Finally, the rule concerning the cluster 3 is “the basis of lab work is Experimental Guidelines and the lab work is done by the Students”.

Regarding the 11th grade, the rules concerning the clusters 1, 2 and 3 are respectively “the lab work is done by the Students”; “the lab work is done sometimes by the Students and sometimes by the Teacher”, and “the lab work is done by the Teacher”.

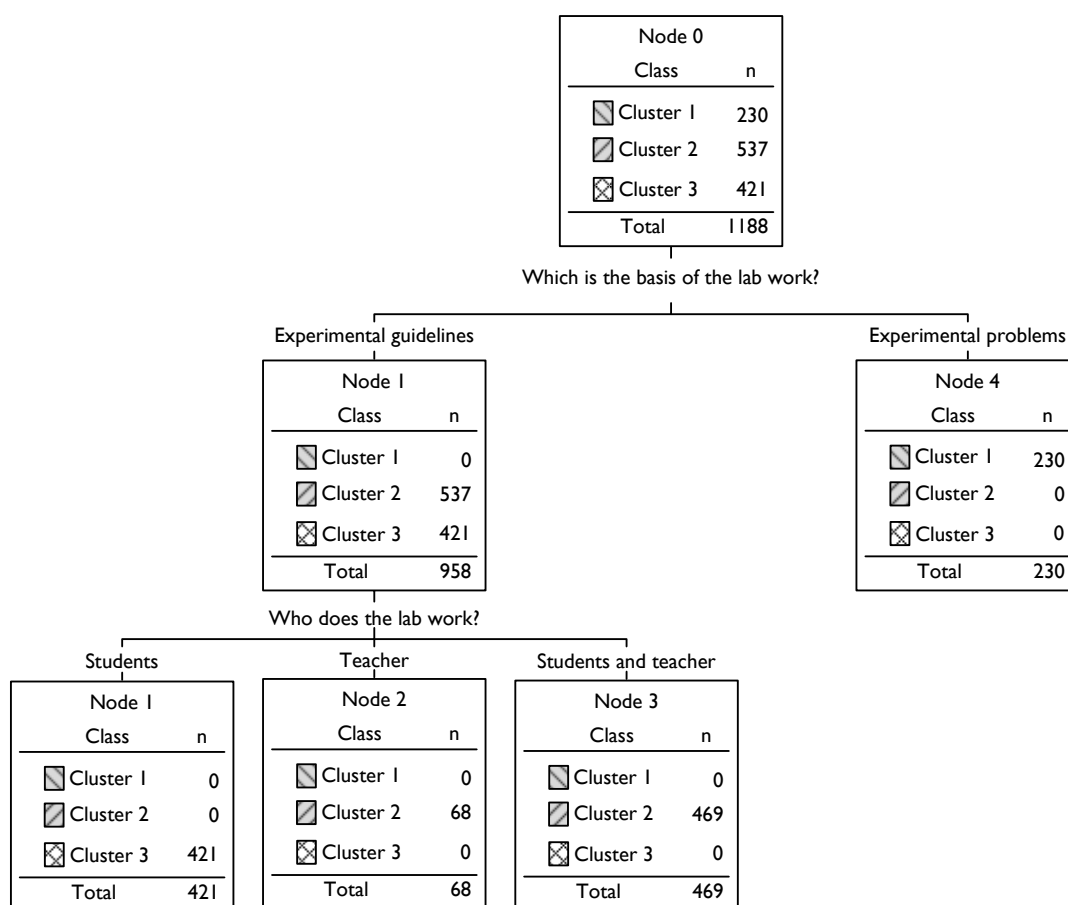


Fig. 5 Decision tree explanatory of the segmentation model obtained with the 10th grade sample

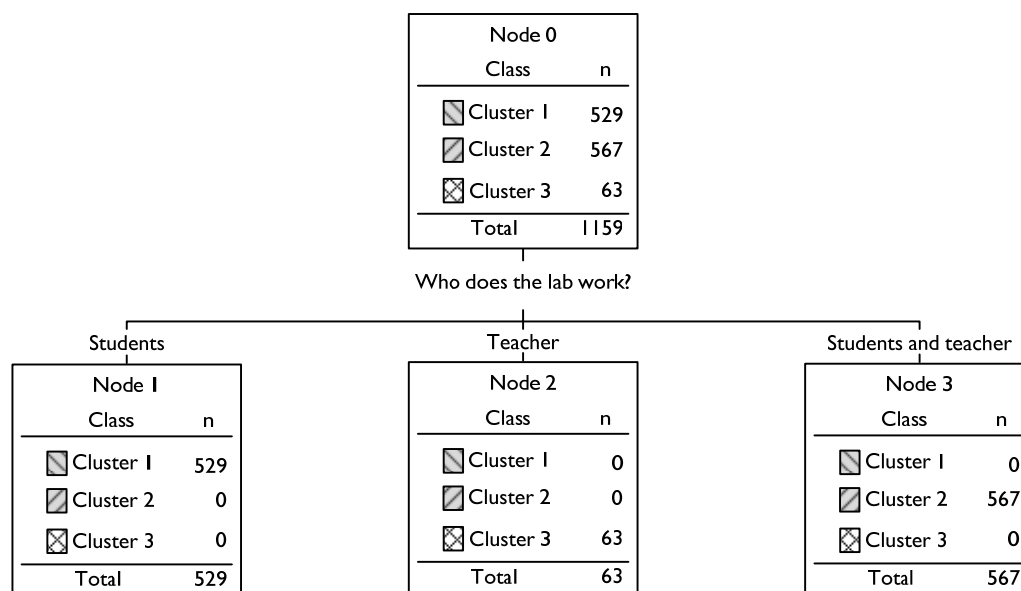


Fig. 6 Decision tree explanatory of the segmentation model obtained with the 11th grade sample

23
24
25
26
27
28
29
30
31
32

A common tool to evaluate the results presented by the DTs models is the coincidence matrix, a matrix of size $L \times L$, where L denotes the number of possible classes. This matrix is created by matching the values predicted by the model (rows) with the target values (columns). The coincidence matrixes, presented in Table 4, denote the average of 20 (twenty) experiments and reveal that the accuracy of the DTs displayed in Fig. 5 and 6 are 100% for both training and test sets.

33 Influence of the methodology used in lab classes on the weight 34 attributed by the students to lab work on learning of chemistry 35 and own motivation

36
37
38
39
40
41
42
43
44
45
46

10TH GRADE. In order to evaluate the influence of the teaching methodology followed in the lab classes on the weight attributed by the respondents to lab work in chemistry, the graph presented in Fig. 7a) was conceived. The strength of the relationships between clusters and answers is visible on the type of connections. It shows that regardless of the cluster to which the respondents are assigned, the majority of applicants consider that the significance of lab work in chemistry learning is “*Very high*” or “*High*”. Other possible answers like

“*Moderate*”, “*Low*” or “*Very low*” are negligible, once to them are assigned less than 2% of answers. However, a further analysis shows that the highest percentage of respondents, assuming that the significance of lab work in chemistry learning is “*Very high*”, belongs to cluster 1, i.e., the respondents that state that lab classes are based on tentative situations form this cluster. Another interesting point is related to the fact that no respondents allocated in cluster 1 have endorsed the responses “*Very low*” or “*Low*”. Only a small percentage ($\leq 0.5\%$) answered “*Moderate*”. Concerning cluster 2, that comprise the respondents whose lab classes are built on experimental guidelines and the lab work is not done exclusively by the students, the percentage of positive responses (“*Very high*” and “*High*”) are quite similar, and the percentage of negative responses (“*Moderate*”, “*Low*” and “*Very low*”) is higher than in the other clusters. Regarding cluster 3, that includes the respondents whose lab classes are based on experimental guidelines and the lab work is done exclusively by the students, the results are similar to that obtained for cluster 1 in terms of the positive responses, although exhibiting lesser percentages.

47
48
49
50
51
52
53
54
55
56
57

Table 4 Coincidence matrixes conforming to the decision trees illustrative of the segmentation, obtained for the 10th and the 11th grades ^a

	10 th grade						11 th grade					
	Training set			Test set			Training set			Test set		
	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3	Cluster 1	Cluster 2	Cluster 3
Cluster 1	230	0	0	70	0	0	529	0	0	264	0	0
Cluster 2	0	537	0	0	234	0	0	567	0	0	268	0
Cluster 3	0	0	421	0	0	244	0	0	63	0	0	20

^a The values displayed denote the average of 20 experiments.

This result may be related with the development of higher level skills associated with the inquiry and the planning of the lab work, which are not present in the lab classes based on experimental guidelines. According Hofstein (2004) the appropriate laboratory activities can be effective in promoting cognitive skills, metacognitive skills, practical skills, and attitude and interest towards chemistry, learning chemistry, and practical work in the framework of chemistry learning.

Another feature to be exploited in the present work has to do with the influence of the methodology followed in lab classes to increase the student's motivation to learn chemistry. Fig. 7b) shows the strength of the relationships (given in percentages) between the clusters to which the respondents are assigned and the replies to the question *Does your motivation to study chemistry increase when you execute lab work?*. The analysis of Fig. 7b) shows that regardless of the cluster, the majority of partakers consider that the motivation to study chemistry increases "Very high" or "High" with the execution of lab work. Nevertheless, an examination of Fig. 7b) shows that the percentages of positive answers given by the respondents assigned to cluster 1 are similar and lower than 50.0%. Furthermore, a non-negligible percentage of respondents ($\geq 6.0\% \wedge < 9.0\%$) claim that the increase of motivation to study chemistry is "Moderate". Regarding cluster 2, the percentage of answers "High" is greater than the percentage of answers "Very high". The percentage of negative answers is higher than in the other clusters. Concerning cluster 3, the percentage of answers "Very High" is greater than the percentage of answers "High". None of the respondents

assigned to this cluster indicated the answer "Very low". Only a small percentage ($\leq 0.5\%$) answered "Moderate" or "Low".

The results presented above seem to indicate, on the one hand, the role played by the lab classes to increase the students' motivation to study chemistry and, on the other hand, also seem to show that the scholars reveal some resistance in executing the research and planning work required to perform open lab classes, based on tentative situations. According to Logar and Savec (2011) these results may be linked to the use of experimental skills of the learners (e.g. what to do, how and when), i.e., working with laboratory equipment and materials, use of laboratory manuals, use of theoretical basics of experimental work, terms, symbols, representations, working with classmates in groups.

11TH GRADE. In order to examine the weight that the respondents attribute to the lab work in chemistry learning, considering the methodology followed in lab classes, with respect to the 11th grade sample, the graph presented in Fig. 8a) was envisaged. In an appraisal with the results gotten to the 10th grade sample, the positive answers ("Very High" or "High") prevail over the negative ones ("Moderate", "Low" or "Very low") in all clusters, although some answers included in the latter group exhibit higher significance (up to 28%). A closer analysis of Fig. 8a) shows that 64.9% of the respondents, included in cluster 1, state that the weight of lab work in chemistry learning is "Very high". Indeed, this cluster is made on the respondents to whom the lab work is done exclusively by the students. None of the respondents assigned to this cluster indicated the answer "Low" or "Very low", and only a small percentage ($< 1\%$) answered "Moderate".

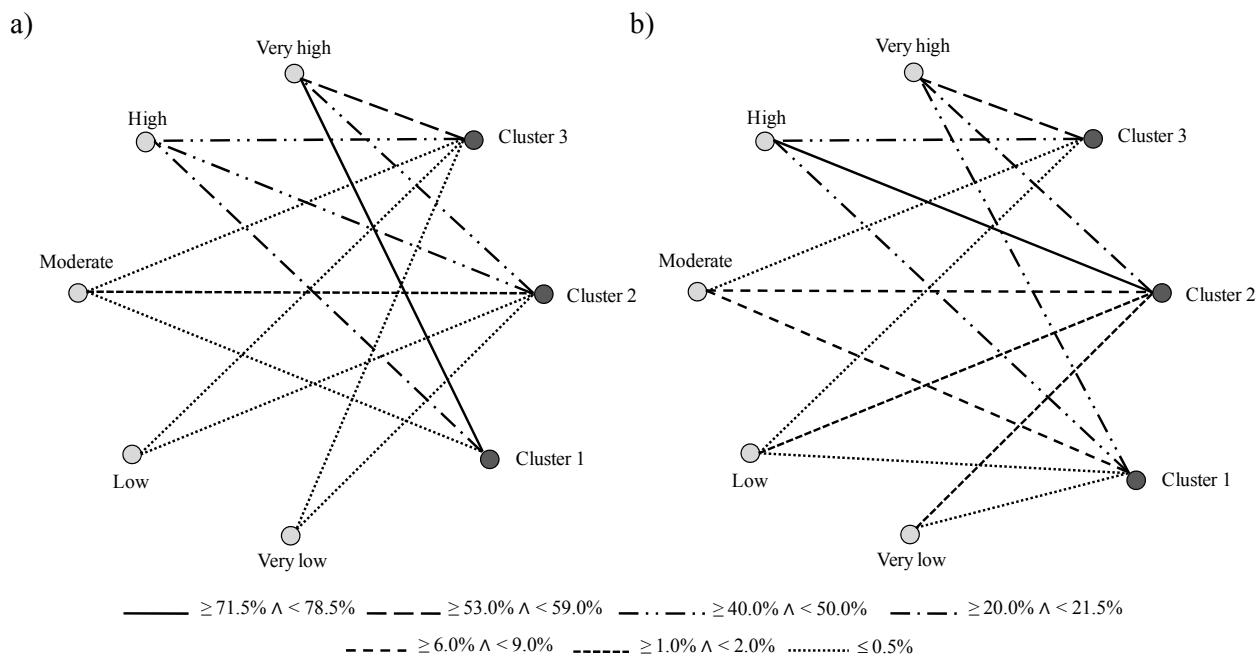


Fig. 7 Relationships between clusters and the percentages of answers to the questions related to students' perception about the importance of lab work in a) the learning of Chemistry and b) the increase of own motivation, for the 10th grade sample

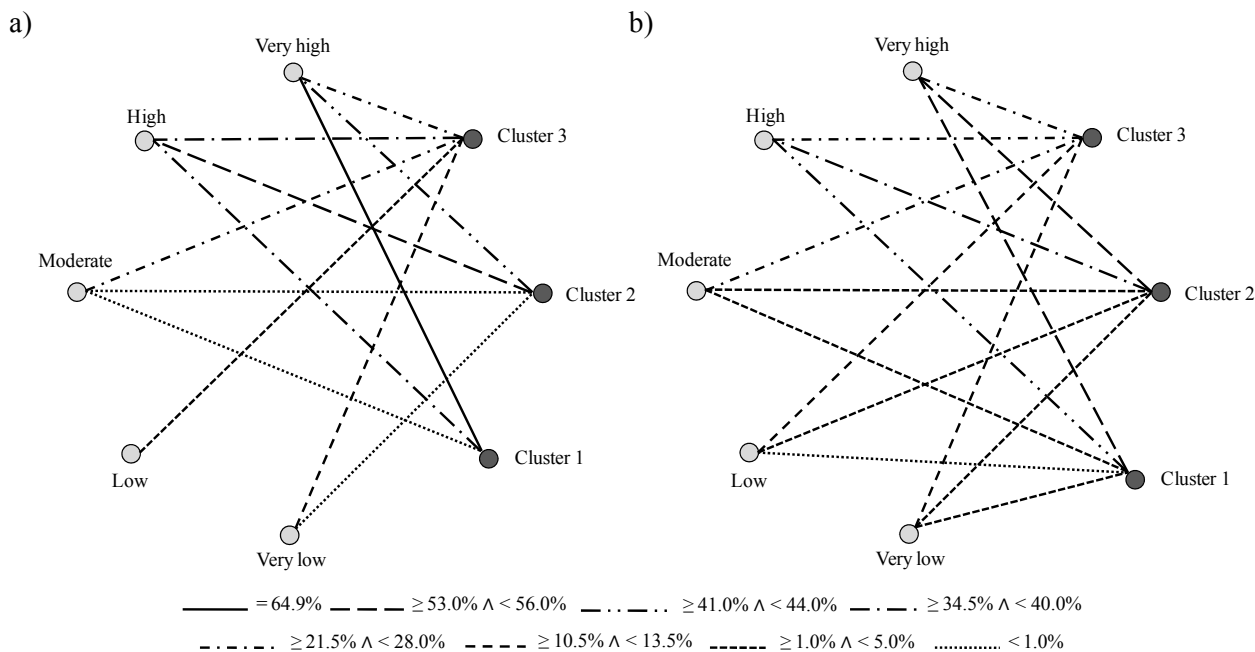


Fig. 8 Relationships between clusters and the percentages of answers to the questions related to students' perception about the importance of lab work in a) the learning of Chemistry and b) the increase of own motivation, for the 11th grade sample

Concerning cluster 2, moulded by the respondents that assert that the lab work is not done exclusively by the students, the percentage of answers "High" is greater than the percentage of answers "Very high". None of the respondents assigned to this cluster indicated the answer "Low", and only a small percentage ($< 1\%$) answered "Moderate" or "Very low". Regarding cluster 3, that includes the respondents to which lab work is done exclusively by the teachers, the percentage of answers "High" is greater than the one of answers "Very high". However, the overall percentage of the positive answers decreases, ranging between 56% and 67%. This result tells one that at least one third of the respondents included in this cluster have a negative opinion about the weight of lab work in chemistry learning.

These results show that the weight attributed by the students to lab work in chemistry learning is strongly dependent on their involvement. When lab classes are demonstrative (i.e. the lab work is carried out exclusively by the teacher), the weight of lab work in chemistry learning drops. Other researches (Cheung, 2007; Bennett *et al.*, 2010) confirm that the quality of learning based on lab classes increases when students have an active role in the process of adding knowledge. Hofstein and Lunetta (2004) emphasize that the laboratory experiences raise the interest and the students' motivation, and also provide the development of practical skills and the capability of solving problems that may contribute to understand the nature of Science.

Regarding the influence of the methodology followed in lab classes to upturn the students' motivation of the 11th grade to study chemistry, the graph showed in Fig. 8b) was conceived.

The positive answers overcome the negative ones in all clusters, although in some cases there is a relatively high percentage of negative answers (up to 28%). However, a glance to Fig. 8b) shows that the percentage of answers "Very High" given by the respondents assigned to cluster 1 is greater than the percentage of answers "High". Only a small percentage ($\leq 5\%$) answered "Moderate", "Low" or "Very low". Concerning cluster 2 the results are similar to those presented for cluster 1. Regarding cluster 3, the percentage of answers "Very high" and "High" is quite similar. The overall percentage of positive answers ranges between 43% and 56%. This result reveals that about half of the respondents in this cluster have a negative opinion with respect to see lab work as the mean to increase the students' motivation to study chemistry. Keeping in mind that cluster 3 includes the respondents that state that the lab work is done exclusively by the teachers, these results suggest that the increase of the students' motivation to study chemistry is not relevant when the lab classes are demonstrative. Conversely, when lab work is carried out by the students the results obtained with this sample suggest that their motivation to study chemistry increases. These results are in agreement with those obtained by Hofstein (2004). This author refers that the appropriate laboratory activities providing students with authentic and practical learning experiences has the potential to adjust the classroom learning environment and thus to enhance students' motivation to study chemistry.

Conclusions

This study focuses on the use of data mining tools in the educational context and it illustrated some of the potentialities

of this methodology of data analysis. The example chosen was the study of the methodology used in chemistry lab classes on the weight attributed by the students to the lab work on learning and own motivation. This approach allowed a deeper analysis of the results once the answer frequency analysis was unable to discriminate the opinions expressed by the respondents according to the type of the teaching methodology used in the lab classes. Indeed, the answer frequency analysis showed that the overwhelming majority of the respondents claim, on the one hand, that the lab work is important for chemistry learning and, on the other hand, leads to an enhancing in the students' motivation to study this subject. Therefore, it was not possible to conclude about the methodology that, in the opinion of the respondents, promotes chemistry learning and contributes to increase the students' motivation to study this subject.

Conversely, the data mining approach using k-means clustering models presented in this study, enabled one to identify the methodology to teach chemistry that, in the students' opinion, is important for learning chemistry and increasing their motivation. The results obtained with the data mining approach, based on students' opinion, showed that the methodology used in lab classes that most contributes for students' own motivation and for learning of chemistry is one that is based on the work of the students.

The results obtained in this study, based on students' opinions, could be important for teachers. Indeed, the results show that the type of methodology that should be adopted in lab classes must involve the students' work. The 135 minutes session, planned in the curricula for the realization of practical work, may be determinant to engage students to proceed studies in the scientific area of chemistry and, in the future, to choose careers related with this science. To achieve such goal the practical work must be mainly lab work, in which the students must conduct by themselves all the stages of the lab work development (planning, execution and interpretation), i.e., the students should be involved in the process of gaining knowledge.

The encouraging results obtained in this work show that data mining approach can be very useful to identify the methodologies followed in lab classes that most contribute to increase the weight attributed by the students to the lab work on learning and their own motivation. However, it should be highlighted that this study is based on students' opinion. It is impossible to be assertive about the methodology that works best, since the study design did not consider the collection of data related with the learning assessment.

Notes and references

^a Departamento de Química, Centro de Investigação em Educação e Psicologia, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal. E-mail: mtf@uevora.pt

^b Departamento de Química, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal.

^c Algoritmi, Universidade do Minho, Braga, Portugal. E-mail: jneves@di.uminho.pt

^d Departamento de Química, Centro de Química de Évora, Escola de Ciências e Tecnologia, Universidade de Évora, Évora, Portugal. E-mail: hvicente@uevora.pt

Abrahams, I. and Reiss, M. J., (2012), Practical work: Its effectiveness in primary and secondary schools in England, *Journal of Research in Science Teaching*, **49**, 1035-1055.

Ames, C. and Archer, J., (1988), Achievement goals in the classroom: students' learning strategies and motivation processes, *Journal of Educational Psychology*, **80**, 260-267.

Baeten, M., Dochy, F. and Struyven, K., (2012), The effects of different learning environments on students' motivation for learning and their achievement, *British Journal of Educational Psychology*, **83**, 484-501.

Bell, J., (2010), *Doing your research project: A guide for first-time researchers in education, health and social science*, 5th ed., Maidenhead, United Kingdom: Open University Press.

Bennett, J., Hogarth, S., Lubben, F., Campbell, B. and Robinson, A., (2010), Talking science: The research evidence on the use of small group discussions in science teaching, *International Journal of Science Education*, **32**, 69-95.

Bennett, S. and O'Neale, K., (1998), Skills development and practical work in chemistry, *University Chemistry Education*, **2**, 58-62.

Bopegedera, A., (2011), Putting the laboratory at the center of teaching chemistry, *Journal of Chemical Education*, **88**, 443-448.

Bradley, P. S. and Fayyad, U. M., (1998), Refining Initial Points for K-Means Clustering, in Shavlik J. (ed.), *Proceedings of the 15th International Conference on Machine Learning (ICML98)*, San Francisco, United States of America: Morgan Kaufmann, pp. 91- 99.

Breiman, L., Friedman, J. H., Olshen, R. A. and Stone, C. J., (1984), *Classification and Regression Trees*, Boca Raton, United States of America: Chapman & Hall.

Changeiywo, J. M., Wambugu, P. W. and Wachanga, S. W., (2011), Investigations of students' motivation towards learning secondary school physics through mastery learning approach, *International Journal of Science and Mathematics Education*, **9**, 1333-1350.

Cheung, D., (2007), Facilitating chemistry teachers to implement inquiry-based laboratory work, *International Journal of Science and Mathematics Education*, **6**, 107-130.

Coe, R., Searle, J., Barmby, P., Jones, K. and Higgins, S., (2008), *Relative difficulty of examinations in different subjects*, Report for SCORE (Science Community Supporting Education), retrieved 17th February 2015 from <http://www.cem.org/attachments/score2008report.pdf>.

Cohen, L., Manion, L. and Morrison, K., (2011), *Research Methods in Education*, 7th ed., New York, United States of America: Routledge.

Couto, C., Vicente, H., Machado, J., Abelha, A. and Neves, J., (2012), Water Quality Modelling using Artificial Intelligence Based Tools, *International Journal of Design & Nature and Ecodynamics*, **7**, 299-308.

DeKetele, J. and Roegiers, X., (2009), *Méthodologie du Recueil d'Informations: Fondements des Méthodes d'Observation, de Questionnaire, d'Interview et d'Études de documents*, 4th ed., Paris, France: DeBoeck Université.

Dweck, C. S., (1986), Motivational processes affecting learning, *American Psychologist*, **41**, 1040-1048.

- 1 Elliot, A. J., McGregor, H. A. and Gable, S., (1999), Achievement goals,
2 study strategies, and exam performance: a mediational analysis,
3 *Journal of Educational Psychology*, **91**, 549-563.
- 4 Engle, R. A., (2012), The resurgence of research into transfer: an
5 introduction to the final articles of the transfer strand, *Journal of the*
6 *Learning Sciences*, **21**, 347-352.
- 7 Figueiredo, M., Vicente, L., Vicente, H. and Neves, J., (2014), School
8 Dropout Screening through Artificial Neural Networks based
9 Systems, in Mastorakis, N., Dondon, P. and Borne P. (ed.), *Advances*
10 *in Educational Technologies*, Educational Technologies Series, Vol. 12,
11 pp. 22-27.
- 12 Gee, B. and Clackson, S. G., (1992), The origin of practical work in the
13 english school science curriculum, *School Science Review*, **73**, 79-83.
- 14 Grant, H. and Dweck, C. S., (2003), Clarifying achievement goals and
15 their impact, *Journal of Personality and Social Psychology*, **85**, 541-
16 -553.
- 17 Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and
18 Witten, I. H., (2009), The WEKA Data Mining Software: An Update,
19 *SIGKDD Exploration*, **11**, 10-18.
- 20 Han, J., Kamber, M. and Pei, J., (2011), *Data Mining: Concepts and*
21 *Techniques*, 3th ed., San Francisco, United States of America:
22 Morgan Kaufmann Publishers.
- 23 Harackiewicz, J. M., Barron, K. E., Pintrich, P. R., Elliot, A. J. and
24 Thrash, T. M., (2002), Revision of achievement goal theory:
25 necessary and illuminating, *Journal of Educational Psychology*, **94**,
26 638-645.
- 27 Harter, S., (1981), A new self-report scale of intrinsic versus extrinsic
28 orientation in the classroom: Motivational and informational
29 components, *Developmental Psychology*, **17**, 300-312.
- 30 Heyman, G. D. and Dweck, C. S., (1992), Achievement goals and
31 intrinsic motivation: Their relation and their role in adaptive
32 motivation, *Motivation and Emotion*, **16**, 231-247.
- 33 Hodson, D., (1990), A critical look at practical work in school science,
34 *School Science Review*, **70**, 33-40.
- 35 Hodson, D., (1993), Re-thinking old ways: Towards a more critical
36 approach to practical work in school science, *Studies in Science*
37 *Education*, **22**, 85-142.
- 38 Hofstein, A. and Lunetta, V. N., (2004), The laboratory in science
39 education: foundations for the twenty-first century, *Science*
40 *Education*, **88**, 28-54.
- 41 Hofstein, A. and Mamlok-Naaman, R., (2007), The laboratory in science
42 education: The state of the art, *Chemistry Education Research and*
43 *Practice*, **8**, 105-107.
- 44 Hofstein, A., (2004), The laboratory in chemistry education: Thirty years
45 of experience with developments, implementation and evaluation,
46 *Chemistry Education Research and Practice*, **5**, 247-264.
- 47 Johnstone, A. H. and Al-Shuaili, A., (2001), Learning in the laboratory:
48 Some thoughts from the literature, *University Chemistry Education*,
49 **5**, 42-51.
- 50 Johnstone, A. H., (1993), The development of chemistry teaching: A
51 changing response to changing demand, *Journal of Chemical*
52 *Education*, **70**, 701-705.
- 53 Josephsen, J., (2003), Experimental training for chemistry students: does
54 experimental experience from the general sciences contribute?,
55 *Chemistry Education Research and Practice*, **4**, 205-218.
- 56 Keiler, L. S. and Woolnough, B. E., (2002), Practical work in school
57 science: The dominance of assessment, *School Science Review*, **83**,
83-88.
- 58 Kind, P. M., Kind, V., Hofstein, A. and Wilson, J., (2011), Peer
59 argumentation in the school science laboratory – Exploring effects of
60 task features, *International Journal of Science Education*, **33**, 2527-
-2558.
- Klosgen, W. and Zytkow, J., (2002), *Handbook of data mining and*
knowledge discovery. New York, United States of America: Oxford
University Press.
- Koballa, T. R. and Glynn, S. M., (2007), Attitudinal and motivational
constructs in science learning, in Abell S. K. and Lederman N. G.
(ed.), *Handbook of research on science education*, New Jersey,
United States of America: Lawrence Erlbaum Associates Publishers,
pp. 103-124.
- Lock, R., (1988), A history of practical work in school science and its
assessment, 1860-1986, *School Science Review*, **70**, 115-119.
- Logar, A. and Savec, V. F., (2011), Students' hands-on experimental
work vs lecture demonstration in teaching elementary school
chemistry, *Acta Chimica Slovenica*, **58**, 866-875.
- MacQueen, J. B., (1967), Some methods for classification and analysis of
multivariate observations, in *5th Berkeley Symposium on*
Mathematical Statistics and Probability, vol. 1, Berkeley, United
States of America: University of California Press, pp. 281-297.
- McMillan, J. and Schumacher, S., (2009), *Research in Education:*
Evidence-Based Inquiry, 7th ed., New York, United States of
America: Prentice Hall.
- Miguéns, M. and Garrett, R. M., (1991), Prácticas en la enseñanza de las
ciencias. Problemas y posibilidades, *Enseñanza de las Ciencias*, **9**,
229-236.
- Millar, R., (2002), Thinking about practical work, in Amos S. and
Boohan R. (ed), *Aspects of teaching secondary science: Perspectives*
on practice, London, United Kingdom: Routledge Falmer, pp 53-59.
- Millar, R., (2004), The role of practical work in the teaching and learning
of science, in *High school science laboratories: Role and vision*,
Washington DC, United States of America: National Academy of
Sciences, pp. 1-24.
- Millar, R., (2010), Practical work, in J. Osborne, J. and Dillon J. (ed.),
Good practice in science teaching: What research has to say, 2nd ed.
Maidenhead, United Kingdom: Open University Press, pp. 108-134.
- Neves, J., Figueiredo, M., Vicente, L., Gomes, G. Macedo, J. and Vicente,
H., (2015) Quality of Learning under an All-inclusive Approach, in Di
Mascio, T., Gennari, R., Vittorini, P. and De la Prieta, F. (ed.),
Methodologies and Intelligent Systems for Technology Enhanced
Learning, Advances in Intelligent and Soft Computing, Vol. 374,
Cham, Switzerland: Springer International Publishing, pp. 41-50.
- Nokes, T. J. and Belenky, D. M., (2011), Incorporating motivation into a
theoretical framework for knowledge transfer, in Mestre J. P. and
Ross B. H. (ed.), *The psychology of learning and motivation:*
Cognition and education, vol. 55, San Diego, United States of
America: Academic Press, pp. 109-135.
- Nokes-Malach, T. J. and Mestre, J., (2013), Towards a model of transfer
as sense-making, *Educational Psychologist*, **48**, 184-207.
- Nunes, J., Madeira, M., Gazarini, L., Neves, J. and Vicente H., (2012), A
Data Mining Approach to Improve Multiple Regression Models of

- 1 Soil Nitrate Concentration Predictions in *Quercus rotundifolia*
2 “Montados” (Portugal), *Agroforestry Systems*, **84**, 89-100.
- 3 OECD, (2009), *Top of the class: High performers in Science in PISA*
4 *2006*, Paris, France: OECD Publishing.
- 5 OECD, (2014), *PISA 2012 Results: What Students Know and Can Do –*
6 *Student Performance in Mathematics, Reading and Science*, vol. I,
7 revised ed., Paris, France: OECD Publishing.
- 8 Oxford Economics, (2010), *The Economic Benefits of Chemistry*
9 *Research to the UK*, retrieved 17th February 2015 from
10 <http://www.oxfordeconomics.com/my-oxford/projects/129036>.
- 11 Perkins, D. N. and Salomon, G., (2012), Knowledge to go: a motivational
12 and dispositional view of transfer, *Educational Psychologist*, **47**, 248-
13 258.
- 14 Pinto, A., Fernandes, A., Vicente, H. and Neves, J., (2009), Optimizing
15 Water Treatment Systems Using Artificial Intelligence Based Tools,
16 in Brebbia, C. and Popov V. (ed.), *Water Resources Management V*,
17 WIT Transactions on Ecology and the Environment, Vol. 125,
18 Southampton, United Kingdom: WIT Press, pp. 185-194.
- 19 Pugh, K. J. and Bergin, D. A., (2006), Motivational influences on
20 transfer, *Educational Psychologist*, **41**, 147-160.
- 21 Quinlan, J. R., (1986), Induction of decision trees, *Machine Learning*, **1**,
22 81-106.
- 23 Quinlan, J. R., (1993), *C4.5 Programs for Machine Learning*, San Mateo,
24 United States of America: Morgan Kaufmann Publishers Inc.
- 25 Richey, J. E. and Nokes-Malach, T. J., (2013), How much is too much?
26 Learning and motivation effects of adding instructional explanations
27 to worked examples, *Learning and Instruction*, **25**, 104-124.
- 28 Romero, C., Ventura, S. and García, E., (2008), Data mining in course
29 management systems: Moodle case study and tutorial, *Computers &*
30 *Education*, **51**, 368–384.
- 31 Şen, B. and Uçar, E., (2012), Evaluating the achievements of computer
32 engineering department of distance education students with data
33 mining methods, *Procedia Technology*, **1**, 262-267.
- 34 Şen, B., Uçar, E. and Delen, D., (2012), Predicting and analyzing
35 secondary education placement-test scores: A data mining approach,
36 *Expert Systems with Applications*, **39**, 9468-9476.
- 37 Sevindik, T. and Cömert, Z., (2010), Using algorithms for evaluation in
38 web based distance education, *Procedia Social and Behavioral*
39 *Sciences*, **9**, 1777-1780.
- 40 Shachar, H. and Fischer, S., (2004), Cooperative learning and the
41 achievement of motivation and perceptions of students in 11th grade
42 chemistry classes, *Learning and Instruction*, **14**, 69-87.
- 43 Sharma(Sachdeva), R., Alam, M. A. and Rani, A., (2012), K-Means
44 Clustering in Spatial Data Mining using Weka Interface,
45 *International Journal of Computer Applications – Proceedings on*
46 *International Conference on Advances in Communication and*
47 *Computing Technologies 2012*, **ICACACT(1)**, 26-30.
- 48 Sherz, Z. and Oren, M., (2006), How to change students’ images of
49 science and technology, *Science & Education*, **90**, 965-985.
- 50 Souza, J., Matwin, S. and Japkowicz, N., (2002), Evaluating Data Mining
51 Models: A Pattern Language, in *Proceedings of the 9th Conference*
52 *on Pattern Language of Programs*, Illinois, United States of America,
53 pp. 1-23.
- 54 Stipek, D. (1996), Motivation and instruction, in Berliner, D. and Calfee,
55 R. (ed.), *Handbook of Educational Psychology*, New York, United
56 States of America: Macmillan, pp. 85-113.

Supplemental Materials

QUESTIONNAIRE

Dear Participant,

- This questionnaire aims to collect data for a study on the influence of the methodology used in chemistry lab classes, on the weight attributed by the students to the lab work on learning and own motivation.
- This questionnaire is anonymous and confidential.
- The collected data will not be taken into account for the assessment of discipline.
- In questions Q1 to Q4 should tick the box that **best describes your entire previous school career**.
- In questions Q5 and Q6 there is no right answer. **You must respond according to your opinion.**

Age: _____ Gender: M F Grade: 10th 11th District: _____

Q1. In lab classes who does the lab work?

- a) always the students
- b) sometimes the students and sometimes the teacher
- c) always the teacher

Q2. How are the students organized in the lab classes?

- a) groups of 3 students
- b) groups of 4 students
- c) groups with another number of students

Q3. Which is the basis of the lab work?

- a) experimental guidelines
- b) experimental problems

Q4. What type of post-lab work is done?

- a) worksheets
- b) worksheets and written reports
- c) written reports

Q5. How do you classify the importance of lab work on the learning of Chemistry?

- a) very high
- b) high
- c) moderate
- d) low
- e) very low

Q6. How do you classify the importance of lab work to increase your motivation to study Chemistry?

- a) very high
- b) high
- c) moderate
- d) low
- e) very low