Part 1
Strand 1
Learning science: Conceptual understanding

Co-editors: Odilla Finlayson & Roser Pinto

CONTENTS

INTRODUCTION TO STRAND 1

LEARNING SCIENCE: CONCEPTUAL UNDERSTANDING

Strand 1 focussed on Learning Science: conceptual understanding; within this strand, the research studies essentially address the process of learning science to develop understanding.

A central focus by all the researchers grouped under this strand is that learning science necessitates understanding science and this requires the comprehension of many ideas. As Bransford, Brown and Cocking (2000) point out: Learning with understanding suppose not to emphasize memory as often have been the case but to be able to use knowledge in different contexts and relate them. We have to take into account that a usable knowledge is not a mere list of disconnected facts. It needs connected and organized ideas around important nuclear concepts or models.

Therefore, researchers that have submitted their work on Strand 1 are aware that not only the facts are important in order to think and solve problems, but that students also (a) need to grasp how science has interpreted such facts and has built coherent models, and (b) have to insert the new knowledge into their pre-existing system of ideas and concepts.

So, as learning scientific concepts is not an easy process and as many teachers have experienced difficulties in this area, there has been much research devoted to addressing this problem, as shown by the large number of papers submitted to Strand 1. Also, the research shows different approaches to addressing this problem of learning with/for understanding.

Previous research in the 1980s and 1990s showed how students' conceptual difficulties were built on previous conceptions, e.g. compiled in the ICPE book*: Connecting Research in Physics Education with Teacher Education (1997)*. This research was useful to identify the obstacles that must be overcome in student learning of the main topics in school curricula. Currently, however, the analysis of student's difficulties for conceptual understanding takes on new perspectives. Now, we find research studies to determine student difficulties in topics addressed at higher educational levels such as: Quantum, Relativity, Astronomy, Mechanistic-chemistry, etc. Genetics and Theory of Evolution are also recurring subjects of attention, often related with personal beliefs.

Building on previous research, the current research reported goes beyond identifying gaps in students' conceptual knowledge and conceptual errors, but aims to propose and evaluate teaching strategies that should enable teaching towards conceptual understanding. To this end, different "Learning progressions" have been presented in this Strand 1 to trace students' ideas (for example: E. Osman, L. Wang, H. Hamdan, G. Ampatzidis, etc). Other studies in this Strand suggest that an understanding of the microscopic view of structure of matter can be a good teaching strategy (J.P. Burde, W. Wu, etc.) or that student use of modelling to connect ideas and learn (e.g. C. Fazio, A. C. Dindar,) give rise to in-depth learning.

We also highlight the research evaluating various strategies and resources that have been studied to help the process of modelling, such as:

(a) Use of conceptual maps as a teaching-learning strategy (for example, R. Grobler, F. Lombard)
(b) Solve problems in different contexts (A. Ferreira)
(c) Propose Peer discussions (C. Wagner)
(d) Use of educational games, ICTs, etc. (A. Guerra, A. Almeida)

It is also noted that there are a number of research papers dealing with the process of learning as influenced by the social and cultural conditions of students and their beliefs about their environment (for example, J. Weber, V. Vieira).

Finally, we should remark that the research papers under Strand 1 have two characteristics:

(a) The specific concepts to be learnt or the models to be built are clearly defined in all the studies. That is, there are no pedagogical or general reflections applicable in all. They are focused on particular specific scientific topics.

(b) A qualitative methodology is most frequently used in order to collect and analyse the data. This provides a rich data set that helps researchers to understand in depth what happens along the process of building conceptual models, rather than merely counting the frequency of some particular conceptual construction.

Odilla Finlayson and Roser Pinto

USING CLUSTER ANALYSIS TO STUDY THE MODELING ABILITIES OF ENGINEERING UNDERGRADUATE STUDENTS: A CASE STUDY

Claudio Fazio[1], *Onofrio R. Battaglia*[1], *Benedetto di Paola*[2] & *Dominique Persano Adorno*[1].
[1] Gruppo di Ricerca in Didattica della Fisica, Università di Palermo, Italy
[2] Gruppo di Ricerca sull'Insegnamento delle Matematiche, Università di Palermo, Italy

Abstract: In this contribution we discuss the application of a quantitative, non-hierarchical clustering method to make sense of the answers that 120 engineering undergraduates students at the University of Palermo, Italy, gave to four open-ended questions on the meaning of the modeling processes in Science. We will show that the use of non-hierarchical analysis allows us to easily separate students into groups that can be recognized and characterized by common traits in students' answers without any prior knowledge on the part of the researcher of what form those groups would take (unbiased classification).

Keywords: Cluster Analysis; Physics Education Research; Modeling

INTRODUCTION

Extensive qualitative research involving open answer questionnaires has provided instructors/teachers with tools to investigate their students' conceptual knowledge of various fields of physics. Some of these studies examined the consistency of students' answers in a variety of situations. Others looked at problems where the underlying physical systems are similar from the point of view of an expert. In recent years, some papers have tried to develop more detailed models of the consistency of students' reasoning, or to subdivide a sample of students into intellectually similar subgroups. Bao and Redish (2006) introduced model analysis as a framework for exploring the structure of the consistency of the application of student knowledge, by separating a group of students into intellectually similar subgroups.

The problem of taking a set of data and separating it into subgroups where the elements of each subgroup are more similar to each other than they are to elements not in the subgroup has been extensively studied through the statistical method of Cluster Analysis (ClA). ClA can separate students into groups that can be recognized and characterized by common traits in their answers, without any prior knowledge of what form those groups would take (unbiased classification).ClA, introduced in Psychology by R.C. Tyron in 1939, has been the subject of research since the beginning of the 1960s, with its first systematic use by Sokal e Sneath in 1963. The application of techniques related to ClA is common in many fields, including Information Technology, Biology, Medicine, Archeology, Econophysics and Market Research. For example, in market research it is important to classify the key elements of the decision-making processes of business strategies as the characteristics, needs and behavior of buyers. These techniques allow the researcher to locate subsets or clusters within a set of objects of any nature. These have a strong tendency to be homogeneous "in some sense". The results of the analysis should reveal a high homogeneity within the group (intra-cluster), and high heterogeneity between groups (inter-clusters), in line with the criteria chosen.

In the literature concerning research in education, some studies using ClA methods are found. They group and characterize student responses by using open-ended questionnaires (Wittmann & Scherr, 2002; Fazio et al., 2012; Fazio et al., 2013) or multiple-choice tests (Ding & Beichner, 2009). A recent paper (Stewart et al., 2012) analyses the evolution of student responses to seven contextually different versions of two Force Concept Inventory questions, by using a model analysis for the state of student knowledge and ClA methods to characterize the distribution of students' answers. This paper shows that ClA methods are an effective way to examine the structure of student understanding and can produce significant subgroups of a data sample. The authors conclude that the ClA method is an effective mechanism for extracting the underlying subgroups in student data and that additional insight may be gained from a careful, qualitative analysis of clustering results. In fact, each cluster is characterized by means of a careful reading of the typical trends in the answers of the individuals that are part of the cluster. It is well known that there are inherent difficulties in the classification of student responses in the studies mainly involving open-ended questionnaires. In fact, the problem of quantifying qualitative data has been widely discussed in the literature for many years (Green, 2001), and it has been pointed out that, very often, a small or even unconscious researcher bias means that the categories picked out tend to find those groups of students that the researcher is already looking for. A recent paper (Hammer & Berland, 2014) points out that researchers "should not treat coding results as data but rather as tabulations of claims about data and that it is important to discuss the rates and substance of disagreements among coders" and proposes guidelines for the presentation of research that quantifies qualitative data. Another paper (Chi, 1997) discussed the need to describe the process of developing a coding scheme, by outlining that in the process of quantifying qualitative data, data means the qualitative records supplied by students and not the result of the coding scheme. If we call these records "raw data" we have to take into account that the data being quantitatively analyzed, which is obtained through the process of data reduction (Hammer & Berland, 2014) contained in the coding scheme, is biased by the subjective interpretation of researchers. It is important for this to be taken into account in the interpretation of the results of the subsequent quantitative analysis.

In this paper we start from a description of the data coding needed in ClA, in order to discuss the meanings and the limits of the interpretation of quantitative results. Then a method commonly used in ClA is described and the variables and parameters involved are outlined and criticized. The application of this method to the analysis of data from an open-ended questionnaire administered to a sample of university students and the related quantitative results is presented. In the last section we discuss the meaning of our results for the physics education researcher and outline some points of strength and limits.

METHODS

Data setting

Research in education that uses open-ended questions and is aimed at quantifying qualitative data usually involves the development of coding procedures. This requires an accurate reading of student answers in order to reveal (and then examine) patterns and trends, and to find common themes emerging from them. These themes are then developed in a number of categories, which can be considered the typical "answering strategies" put into action by the N students tackling the questionnaire items. Therefore, it is possible to summarize the whole set of answers given to the questionnaire into a limited number, M, of answering strategies, making the subsequent analysis easier. Through coding and categorization we produce a set of M data (the answering strategies) for each of the sample subjects (the N students doing the questionnaire). As a consequence, each subject, i, can be identified by an array, ai, composed of M components 1 and 0, where 1 means that the subject used a given answering strategy to respond to an item, and 0 means that he/she did not use it. Then, a M x N binary matrix (the "matrix of answers") modeled on the one shown in Table 1, is built. The columns in it show the N student arrays, a_i, and the rows represent the M components of each array, i.e. the M answering strategies.

Table 1. Matrix of data: the *N* students are indicated as S_1, S_2, …,S_N., and the *M* answering strategies as AS_1, AS_2, …, AS_M.

Strategy	*Student*			
	S_1	S_2	…	S_N
AS_1	1	0	…	0
AS_2	1	0	…	1
…	0	…	…	…
AS_M	0	1	…	0

For example, let us say that student S_1 used answering strategies AS_1, AS_2 and AS_5 to respond to the questionnaire questions. Therefore, S_1 column in Table 1 will contain the binary digit 1 in the three cells corresponding to these strategies, while all the other cells will be filled with 0.The matrix depicted in Table 1 contains all the information to describe the sample behavior with respect to the questionnaire items. However, it needs some elaboration in order to make this information understandable. ClA classifies subset behaviors in different groups (the clusters). These groups can be analyzed in order to deduce their distinctive characteristics and point out similarities and differences among them. ClA requires the definition of new quantities that are used to build the grouping, like the "similarity" or "distance" indexes. These indexes are defined by starting from the *M* x *N* binary matrix discussed above.In the literature the similarity between two elements*i* and *j* of the sample is often expressed by taking into account the distance, d_{ij}, between them (which actually expresses their "dissimilarity", in the sense that a higher value of distance involves a lower similarity). The distance index can be defined by starting from the Pearson's correlation coefficient. It allows the researcher to study the correlation between elements*i* and *j* if the related variables describing them are numerical. If these variables are non-numerical variables (as in our case, where we are dealing with the arrays a_i and a_j containing the binary coding of the answers of elements*i* and *j*, respectively), we propose a modified form of the Pearson's correlation coefficient, R_{mod}, similar to that defined by Tumminello et al. (2011) as,

$$R_{\mathrm{mod}}\left(a_i,a_j\right)=\frac{p\left(a_i\cap a_j\right)-\dfrac{p\left(a_i\right)p\left(a_j\right)}{M}}{\sqrt{p\left(a_i\right)p\left(a_j\right)\left(\dfrac{M-p\left(a_i\right)}{M}\right)\left(\dfrac{M-p\left(a_j\right)}{M}\right)}} \tag{1}$$

where $p(a_i)$ and $p(a_j)$ are the number of properties of a_i and a_j explicitly present in our elements (i.e. the numbers of 1's in the arrays a_i and a_j, respectively), *M* is the total number of properties to study (in our case, the possible answering strategies) and $p(a_i \cap a_j)$ is the number of properties common to both elements, *i* and *j* (the common number of 1's in the arrays a_i and a_j). By following eq. (1) it is possible to find for each student, *i*, the *N-1* correlation coefficients R_{mod} between him/her and the others students (and the correlation coefficient with him/herself, that is, clearly, 1). All these correlation coefficients can be placed in a *N* x *N* matrix that contains the information we need to discuss the mutual relationships between our students. The similarity between subjects *i* and *j* can be defined by choosing a type of metric to calculate the distance d_{ij}. Such a choice is often complex and depends on many factors. If we want two subjects, represented by arrays a_i and a_j and negatively correlated, to be more dissimilar than two positively correlated subjects (as is often advisable in research in education), a possible definition of the distance between a_i and a_j, making use of the modified correlation coefficient, $R_{mod}(a_i, a_j)$, is:

$$d_{ij}=\sqrt{2\left(1-R_{\mathrm{mod}}\left(a_i,a_j\right)\right)} \tag{2}$$

This function defines a Euclidean metric (Gower, 1966), which is required in order to use it for the following calculations. A distance d_{ij} between two students equal to zero means that they are completely similar, while a distance $d_{ij} = 2$ shows that the students are completely dissimilar. By following eq. (2) we can, then build a new N x N matrix , $\mathcal{D}$, containing all the utual distances between the students. The main diagonal of $\mathcal{D}$ is composed by 0s (the distance between a student and him/herself is zero). Moreover, $\mathcal{D}$ is symmetrical with respect to the main diagonal.

Clustering technique

In this paper we use a technique known as Non-Hierarchical Clustering (NH-ClA), that basically allows us to partition the data space into a structure known as a Voronoi diagram (a number of regions including subsets of similar data). Among the many NH-ClA algorithms, we use here the k-means, which was first proposed by MacQueen (MacQueen, 1967). In this method, the final result is a bi-dimensional Cartesian plane containing points that represent the students of the sample placed in the graph according to their mutual distances. As said before, for each student, i, we know N distances. It is, then, necessary to define a procedure to find two Cartesian coordinates for each student, starting from these N distances. This procedure consists in a linear transformation between a N-dimensional vector space and a 2-dimensional one and it is well known in the specialized literature as multidimensional scaling (Borg & Groenen, 1997). The starting point is the choice of the number of clusters one wants to populate and of an equal number of "seed points", randomly selected in the bi-dimensional Cartesian plane representing the data. The subjects are then grouped on the basis of the minimum distance between them and the seed points. Starting from an initial classification, subjects are transferred from one cluster to another or swapped with subjects from other clusters, until no further improvement can be made. The subjects belonging to a given cluster are used to find a new point, representing the average position of their spatial distribution. This is done for each cluster and the resulting points are called the cluster centroids. This process is repeated and ends when the new centroids coincide with the old ones. The spatial distribution of the set elements is represented in a two-dimensional Euclidean space, creating what is known as the k-means graph (see Figure 2).

NH-ClA has some points of weakness and here we will describe how it is possible to overcome them. The first involves the a-priori choice of the initial positions of the centroids. This can usually be resolved by repeating the clustering procedure for several values of the initial conditions and selecting those that lead to the minimum values of the distances between each centroid and the cluster elements. Furthermore, at the beginning of the procedure, it is necessary to arbitrarily define the number of clusters. A method widely used to decide if the number of clusters, q, initially used to perform the calculations is the one that best fits the sample element distribution is the calculation of the so-called Silhouette Function, S. (Rouseeuw, 1987).

Several values of the function S are calculated once a value of the number of clusters, q, is fixed:

- the individual value, $S_{k,i}(q)$, with $k=1, 2,..q$, for each student, i, of the sample. It gives a measure of how similar student $i$is to the other students in its own cluster Cl_k, when compared to students in other clusters. It ranges from -1 to +1; a value near +1 indicates that student i is well-matched to its own cluster, and poorly-matched to neighboring clusters. If most students have a high silhouette value, then the clustering solution is appropriate. If many students have a low or negative silhouette value, then the clustering solution could have either too many or too few clusters (i.e. the chosen number, q, of clusters should be modified).
- The average silhouette value in cluster Cl_k, $< S_k(q)>$, with $k=1, 2,..q$. It gives the average value of $S_{k,i}(q)$, calculated on all the students belonging to cluster Cl_k and it is a measure

of the density of the cluster. Large values of $< S_k(q)>$ are to be related to cluster elements being tightly arranged in the cluster k, and vice versa (Rouseeuw, 1987).

- The total average silhouette value, $< S(q)>$ for the chosen partition in q clusters. It gives the average value of $S_{k,i}(q)$, calculated on all the students belonging to the sample. Large values of $< S(q)>$ are to be related to well defined clusters (Rouseeuw, 1987). It is, therefore, possible to perform several repetitions of the cluster calculations (with different values of q) and to choose the number of clusters, q, that gives the maximum value of $<S(q)>$.

Once the appropriate partition of data in q clusters Cl_k (with $k = 1,..q$) has been obtained, as well as their related centroids, C_k, (i.e. the coordinates in the 2-dimensional space of the q points that represent the average positions of the cluster spatial distributions), it is possible to transform such coordinates in terms of the same variables that represent the students in the plane. In particular, for each centroid, C_k, we find an array b_k with the same number M of components of the array, a_i, that identifies a generic real student i, (i.e. the number M of answering strategies to the questionnaire) and composed, as a_i, by 0 and 1 values. b_k can be considered as the array representing a *virtual student* in cluster Cl_k. By considering the meaning of cluster Cl_k centroid, we could use the answering strategies contained in array b_k to make sense of the features of the clusterreal students.

A remarkable feature of array b_k, that can validate our idea to use the centroid to characterize the features of the cluster Cl_k real students, is that it contains 1 values exactly in correspondence to the answering strategies most frequently given by students belonging to Cl_k. In fact, since a centroid is defined as the geometric point that minimizes the sum of the distances between it and all the cluster elements, by minimizing this sum the correlation coefficients between the cluster elements and the centroid are maximized and this happens when each centroid has the largest number of common strategies with all the students that are part of its cluster.

It is worth noting that if some answering strategies are only slightly more frequent than the other ones all those with similar frequencies should be also considered. In order to analyze how well each centroid characterizes its own cluster Cl_k, we propose a coefficient, r_k, defined as the centroid *reliability,* that relates the cluster density to its dimension. It is calculated as follows:

$$r_k = \frac{\langle S_k(q) \rangle}{1 - \langle S_k(q) \rangle} \frac{1}{n_k} \qquad (3)$$

where n_k is the number of students contained in cluster Cl_k and $< S_k(q)>$ is the average value of the *S-function* on the same cluster. High values of r_k indicate that the centroid characterizes well the cluster, as this happens for dense clusters or for clusters with a low number of students. In fact, considering two equally dense clusters, the one with a lower number of students involves smaller cluster dimensions, i.e. a lower variability of student properties.

Example of quantitative study

In this section we analyze the answer strategies to an open-ended questionnaire supplied by a sample of university students, using the techniques discussed above.

The questionnaire and the sample

The questionnaire is made up of four-items that focus on an understanding of the modeling concept (see Appendix). They are part of a more complex questionnaire, which has already been used, in previous research (Fazio et al., 2012). The selected four items refer to: I) the definition of a physics model, II) the subjects' beliefs about the representational modes of physics models, III) the main characteristics of models and IV) the student's beliefs about the modeling process. The questionnaire was administered to 124 freshmen of the Information Technology and Telecommunications Engineering Degree Course at the University of Palermo, during the first semester of the academic year 2013/2014. The students were given the questionnaire during the first lesson of general physics, before any discussion on the model concept had started.

Categorization of student answers

After the questionnaire had been submitted to our student sample, three researchers independently read the students' answers in order to identify the main characteristics of the different student records (the raw data). Then, they agreed to construct a coding scheme through the identification of keywords that were relevant for an understanding of these records. During the first meeting, the selected keywords were compared and contrasted, and then grouped into categories based on epistemological and linguistic similarities (for example, students that defined models as *simple phenomena* or *experiments* or *reproductions of an object on a small scale* have been put on the same category since the three definitions have been intended as giving a ontological reality to models.). These categories were also re-analyzed through the researchers' interactions with the data, and taking into account the existing literature about models and modeling (Grosslight et al., 1991; Van Driel & Verloop, 1999, Treagust et al., 2002; Pluta et al., 2011). As a third step, the researchers read the student records again and applied the new coding scheme, by assigning each student to a given category for each question. Given the inevitable subjectivity of the researchers' interpretations, the three lists were compared and contrasted in order to get to single agreed list. The inter-rater reliability of the analysis was good. Discordances between researcher lists were usually a consequence of the different personal decisions of the researchers to divide the student answers into a more or less restricted number of typologies. In some cases, discordances were due to different researcher interpretations of student statements. This happened 14 times when comparing tables of researchers 1 and 2, 9 times for researchers 2 and 3, and 12 times for researchers 1 and 3. Hence we obtained very good percentages of accordance (97%, or higher) between the analysis tables of each researcher pair. When a consensus was not obtained, the student answer was classified in the category "statement not understandable".

It is worth noting that very often the researchers' discussions while assigning each student to a given category produced a more refined definition of these categories. The complete list of 20 categories shared by researchers with respect to the four questions can be seen in Appendix A. As a result of the coding and categorization, we obtain a matrix like the one depicted in Table 1, where $N = 124$ and $M = 20$. This matrix of data represents a set of properties (the categories to which student answers have been assigned) for each subject (the student being analyzed).

RESULTS

All the clustering calculations were made using a custom software, written in C language. The graphical representations of clusters in both cases were obtained using the well-known MATLAB software.

In order to define the number q of clusters that best partitions our sample, the mean value of *S-function*, $\langle S(q)\rangle$, has been calculated for different numbers of clusters, from 2 to 10 (see Figure 1). The figure shows that the best partition of our sample is achieved by choosing four clusters, where $\langle S(q)\rangle$ has its maximum. The obtained value $\langle S(4)\rangle = 0.62$ indicates that a reasonable cluster structure has been found (Struyf et al., 1997).

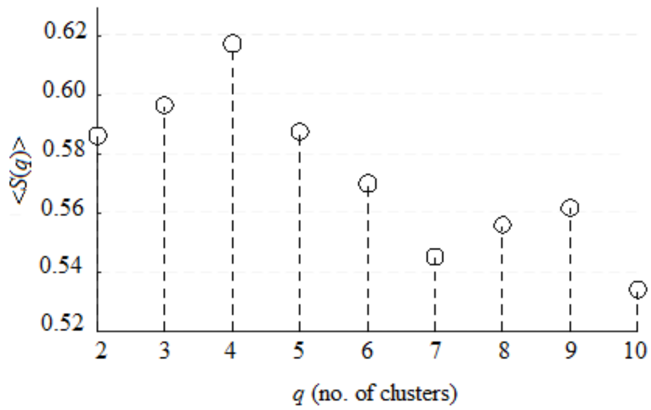

Figure 1. Average Silhouette values for different cluster partitions of our sample.

Figure 2 shows the representation of this partition in a 2-dimensional graph. The four clusters show a partition of our sample into groups made up of different numbers of students (see Table 2)

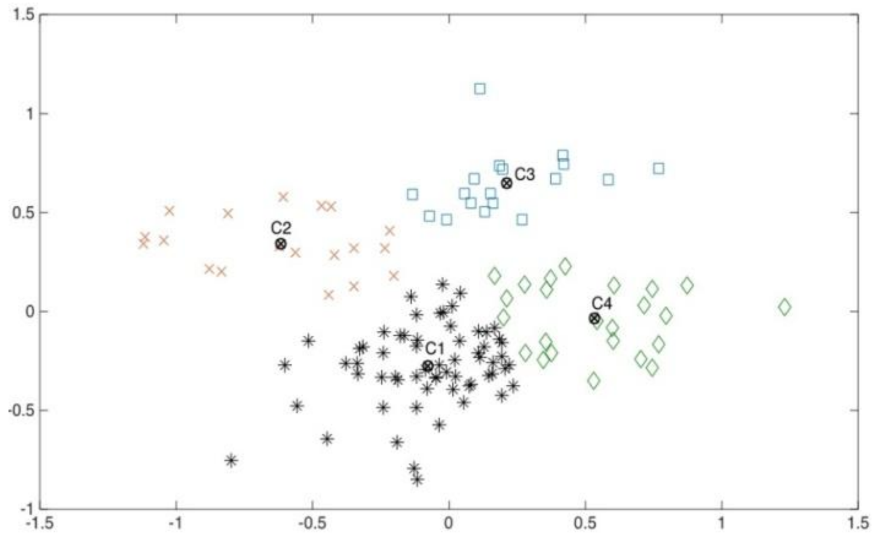

Figure 2. *K-means* graph. Each point in this Cartesian plane represents a student. Points labeled C_1, C_2, C_3, C_4 are the centroids.

The four clusters Cl_k $(k=1,...,4)$ can be characterized by their related centroids, C_k. They are the four points in the graph whose arrays, b_k, contain the answering strategies most frequently applied by students in the related clusters (see Table 2). The codesused refer to the answering strategies for the questionnaire items described in Appendix A. Table 2 also showsthe number of students in each cluster, the mean values of the *S-function* $<S_k(4)>$ $(k=1,..,4)$ for the four clusters andthe normalized reliability index r_k^{norm} of their centroids (in order to have comparable reliability values r_k, they have been normalized ($r_k^{norm} = [(r_k - <r_k>]/\sigma(r_k)$), where $<r_k>$ and $\sigma(r_k)$ are the mean value and the variance, respectively).

Table 2. An overview of the obtained results

Cluster centroid	C_1	C_2	C_3	C_4
b_k (Most frequently given answers)	1C, 2B, 3A, 4A	1B, 2B, 3E, 4A	1B, 2C, 3B, 4A	1C, 2C, 3B, 4B
Number of students	63	19	18	24
$<S_k(4)>$	0.60	0.62	0.75	0.56
r_k^{norm}	-0.92	-0.02	1.4	-0.46

We can see (from the value of $<S_k(4)>$) that cluster Cl_3 is denser than the others, and Cl_4 is the most spread out. Furthermore, the values of r_k^{norm} show that centroid C_3 best represents its cluster, whereas centroid C_1 is the least representative and characterizes less well the cluster.

DISCUSSION AND CONCLUSIONS

The four questions in our questionnaire mainly refer to: I) the definition of a physics model, II) the subjects' beliefs about ways of representing physics models, III) the main characteristics of models and IV) the subjects' beliefs about the modeling process. We have classified student answers into categories, also called answering strategies, that explain student reasoning strategies. Looking at our results, the four clusters identified are characterized by the related centroids and each centroid is represented by one array b_k, which describes the different answering strategies categorized for each question. These strategies are defined as follows: b_1: (1C, 2B, 3A, 4A), b_2: (1B, 2B, 3E, 4A), b_3: (1B, 2C, 3B, 4A), b_4: (1C, 2C, 3B, 4B), where the codes in brackets refer to the questionnaire answer strategies reported in the Appendix. We have already pointed out that the array describing the cluster centroid describes to the answers most frequently given by the students in the cluster, and in this sense we can identify at what frequency each answering strategy is shared by the cluster students.

In particular, cluster Cl_4 is mainly composed of students that use higher level answering strategies to deal with the concepts in the questionnaire. In fact, these students recognize that a model *is a mental representation of a real object or phenomenon,which takes into account the characteristics that are significant for the modeler* (1C). They also think that models*are creations of human thought and their creation comes from continuous interaction with the "real" external world and from its simplification* (2C) and that a model*must highlight the variables that are relevant for the description and/or explanation of the phenomenon analyzed (or the object studied) and their relationships* (3B). *The modeling process is seen as a construction where the model can still contain errors or uncertainty connected with the possibility (or ability) to carefully reproduce the characteristicswe are interested in* (4B). It is worth noting that only 19% of the students belong to cluster Cl_4 and show an informed view of physics models. Such low percentages are also found in the literature (Grosslight et al., 1991; Treagust et al., 2002), although quantitative comparisons cannot be performed, given the differences in the analyzed samples.

Students in cluster Cl_2 show the weakest understanding of the model concept. They refer to a model as *a simple phenomenon or the exemplification of aphenomenon through an experiment or a reduced scale reproduction of an object*(1B), and believe that *models are simple creations of human thought like mathematical formulas, or physics laws and/or they are what we call theories or scientific method* (2B), and give answers regarding the main characteristics of a model that are confused and unclear (3E). For these students *every natural phenomenon can be simplified in order to be referred to a given model* (4A).

Cl_2 students can be reported to the level II modelers based on the classification scheme developed by Grosslight et al. (1991). Level II modelers see models as representations of real-world objects or events and not as representations of ideas about real-world objects or events. They also see the use of different models as that of capturing different spatio-temporal views of the object rather than different theoretical views. Similar results have been obtained in other studies, as for example paper Treagust et al. (2002), that found a significant group of students with a narrow and naïve understanding of the concept of model as an exact replica: the scale replica, a precise representation, which has accuracy and detail; and the imprecise representation, which doesn't have the accuracy or detail, and may be nothing like the object, but can provide insight into why and how something works the way it does. Some studies involving teacher conception of scientific models (Justi& Van Driel, 2005; Danusso et al.,

2010) report conceptions related to such realistic view, mainly where teachers focus on the role of models as examples of objects/processes or their simplifications. To sum up, we can say that the students in cluster Cl_4 seem to share many conceptions connected with an epistemological constructivist view (Treagust et al., 2002). Students in cluster Cl_2, on the other hand, often held beliefs that correspond with a "naïve realist" epistemology, i. e. they usually considered models to be exact copies of reality, albeit on a different scale, or simplified representations (Treagust et al., 2002).

Students in clusters Cl_1 and Cl_3 do not show a full coherence in their answers, although in different ways. Cl_3 students seem to share with Cl_2 students the ideas concerning the definition of physics models and the modeling process, but they also share their beliefs about the function as well as the characteristics of physics models with the students from cluster Cl_4. In fact, they state that *physics defines models as a simple phenomenon or the exemplification of a phenomenon through an experiment or a reduced scale reproduction of an object* (1B). However, they also say that they *are creations of human thought and their creation comes from continuous interaction with the "real" external world and from its simplification* (2C). Furthermore, they seem to share the idea that in a modeling process it is important *to highlight the variables that are relevant for the description and/or explanation of the phenomenon being analyzed (or the subject being studied) and their relationships* (3B) and *that every natural phenomenon can be simplified in order to be referred to a given model* (4A). Such conception of physics model can be reported to literature findings (Fazio et al., 2013; Hrepic et al., 2005) that analyze students' reasoning in different fields and define some kinds of reasoning as ''hybrid models'' (Ding & Beichner, 2009) or ''synthetic models'' (Justi & Gilbert, 2002), by referring to composite mental models that unify different features of initial spontaneous models and scientifically accepted models. Research reveals (Bao & Redish, 2006; Hrepic et al., 2005) that a student can use different mental models in response to a set of situations or problems considered equivalent by an expert. In particular, Bao and Redish (2006) developed a way to deal with these composite mental models and define students' model states that can change with specific contextual features in different equivalent questions. Our data point out that such inconsistency is deployed in the elicitation of model constituents as well as of functions and characteristic of the modeling process.

Students in cluster Cl_1 share the idea that a model is a *mental representation aimed at describing a real object or a phenomenon, which takes into account the characteristics that are significant for the modeler* (1C). However, they also think that *models are simple creations of human thought, like mathematical formulas or physics laws, and/or they are what we call theories or scientific method* (2B). These ideas are not completely consistent with the characteristics assigned to the model or with the students' ideas about the modeling process. In fact they declare that a model *must contain all the rules or all the laws for a simplified description of reality and/or it must account for all the features of reality* (3A) and *that every natural phenomenon can be simplified in order to be referred to a given model* (4A). Their focus on the process of *"simplification"* is also made explicit in the examples they report in order the explain their sentences. For example, for many of such students "the motion without friction is a model as well as the perfect gas (not the motion with friction or real gases".

On the other hand, it must be taken into account that the value of the reliability, r_k^{norm}, of the C_1 centroid is the lowest, showing that array b_1 is not very significant in representing the answering strategies of thecluster students. Also, looking in detail at b_1 array, the answering strategies are not easily understandable from the point of view of consistency and although they represent the answers most commonly given by Cl_1 students, these do not have very high frequencies. For example, no more than 38% is assigned to category 1C. Other answers were also given by a large number of students; for example answering strategy 1B (*A physics model is a simple phenomenon or the exemplification of a phenomenon through an* *experiment or a reduced scale reproduction of an object)* was selected by 30% of Cl_1 students. In our opinion, this may show that a substructure is present in cluster Cl_1, and this should be analyzed through different analysis methods, like, for example Hierarchical Cluster Analysis (Everit et al., 2011), that can point out a higher number of clusters and help to make sense of them.

In conclusion, in this paper, we discussed the problem of quantifying qualitative data in order to analyze how to identify groups with common behavior, ideas, beliefs and conceptual understanding in a sample of students. We presented a method of cluster analysis and analyzed definitions, variables and algorithms in detail, in order to understand the possibilities offered by such a method and its limits. We gave an example of their application in order to demonstrate the necessary approximations and the different ways of interpreting results. The example is an analysis of the answers to a questionnaire given to a sample of university students. It is worth remembering that data that are quantitatively analyzed are the results of a categorization of raw data (the individual student answers) and this reduction of the initial data can be subject to errors, which obviously influences the final evaluation and the inference about the reasoning strategies supporting students' answers. Such errors can only be reduced (through a clear process of coding and subsequent categorization) and not eliminated, and this must be taken into account when we try to infer typical students' reasoning strategies.

Looking at the meaning of the concept of a physics model as understood by the students in our sample, our results are consistent with those described in the literature, which illustrate a continuum of ideas/beliefs ranging from naive conceptions to constructivist ones (Grosslight et al., 1991; Van Driel & Verloop, 1999, Treagust et al., 2002; Pluta et al., 2011) . Our analysis gives details of student conceptions about the function of a physics model and its properties, by identifying features of intermediate conceptions as well as groups of students sharing such conceptions, in a continuum of this type. Furthermore, the results of this study provide important hints and insights for teaching methods that may improve students' model-based reasoning, and provide teachers with information about their students' level of understanding, with which they can make instructional decisions.

REFERENCES

Bao, L. & Redish, E. (2006). Model analysis: Representing and assessing the dynamics of student learning, *Phys. Rev. ST Phys. Educ. Res.* **2**, 010103.

Borg, I. & Groenen, P. (1997). *Modern multidimensional scaling*. New York: Springer Verlag.

Chi, M. (1997). Quantifying Qualitative Analyses of Verbal Data: A Practical Guide, *The Journal of the Learning Sciences* **6**(3), 271-315.

Danusso, L., Testa, I. & Vicentini, M. (2010). Improving Prospective Teachers' Knowledge about Scientific Models and Modelling: Design and evaluation of a teacher education intervention, *International Journal of Science Education* **32**(7), 871-905.

Ding, L. & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions, *Phys. Rev. ST Phys. Educ. Res.* **5**, 020103.

Everitt, B., Landau, S., Leese, M. and Stahl, D. (2011). *Cluster Analysis*. Chichester, UK: John Wiley & Sons, Ltd.

Fazio, C., Di Paola, B. & Guastella, I. (2012). Prospective elementary teachers' perceptions of the processes of modeling: A case study, *Phys. Rev. ST Phys. Educ. Res.* **8**, 010110.

Fazio, C., Battaglia, O. & Di Paola, B. (2013). Investigating the quality of mental models deployed by undergraduate engineering students in creating explanations: the case of thermally activated phenomena, *Phys. Rev. ST Phys. Educ. Res.* **9**, 020101.

Gower, J. (1966). Some Distance Properties of Latent Root and Vector Methods Used in Multivariate Analysis, *Biometrika Trust* **53(**3/4), 325-338.

Green, E. (2001). Can Qualitative Research Produce Reliable Quantitative Findings? *Field Methods* **13**(1), 3–19.

Grosslight, L., Unger, C., Jay, E.& Smith, C. (1991). Understanding models and their use in science: Conceptions of middle and high school students and experts, *Journal of Research in Science Teaching* **28**, 799–822.

Hammer, D. & Berland, L. (2014). Confusing Claims for Data: A critique of Common Practices for Presenting Qualitative Research on Learning, *Journal of the Learning Sciences* **23,** 37-46**.**

Hrepic, Z., Zollman, D. & Rebello, S. (2005). Eliciting and Representing Hybrid Mental Models. *Annual International Conference of the National Association for Research in Science Teaching, April 4-7, Dallas, TX.* http://ldsp01.columbusstate.edu:8080/xmlui/handle/11075/723

Justi, R. & Gilbert, J. (2002). Science teachers' knowledge about and attitudes towards the use of models and modelling in learning science. *International Journal of Science Education* **24**(12), 1273–1292.

Justi, R. & Van Driel, J. (2005). The development of science teachers' knowledge on modelsand modelling: Promoting, characterizing, and understanding the process. *International Journal of Science Education* **27**(5), 549–573.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In: L. Le Cam, J. Neyman (Eds.): *Proc. 5th BerkelySymp. Math. Statist. Probab. 1965/66*, (vol. I, pp. 281-297). Berkely: Univ. of California Press.

Pluta, W., Clark, A.,Chinn, A., Ravit Golan Duncan, R. (2011). Learners' Epistemic Criteria for Good Scientific Models, *Journal of Research in Science Teaching* **48**, 486–511

Rouseeuw, P. (1987). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis, *Journal of Computational and Applied Mathematics* **20,** 53-65**.**

Sokal, R. & Sneath, P. (1963). *Principles of Numerical Taxonomy*. San Francisco and London: W. H. Freeman & Co.,)

Stewart, J., Miller, M., Audo, C. & Stewart, G. (2012). Using cluster analysis to identify patterns in students' responses to contextually different conceptual problems, *Phys. Rev. ST Phys. Educ. Res* .**8**, 020112

Struyf, A., Hubert, M. & Rousseeuw, P. (1997). Clustering in an Object-Oriented Environment. *Journal of Statistical Software* **1**(4), 1-30.

Treagust, D., Chittleborough G. & Mamiala, T. (2002). Students' understanding of the role of scientific models in learning science, *International Journal of Science Education* **24**(4),357-368.

Tryon, R. (1939). *Cluster Analysis: Correlation Profile and Orthometric (factor) Analysis for the Isolation of Unities in Mind and Personality*.Ann Arbor: Edwards Brothers.

Tumminello, M., Micciché, S., Dominguez, L., Lamura, G., Melchiorre, M., Barbagallo, M. & Mantegna, R. (2011). Happy Aged People Are All Alike, While Every Unhappy Aged Person is Unhappy in Its Own, *PLoS ONE* **6**(9): e23377.

Wittmann, M. & Scherr, R. (2002). Student epistemological mode constraining researcher access to student thinking: An example from an interview on charge flow. In K. Cummings, J. Marx, & S. Franklin (Eds), *Physics Education Research Conference Proceedings 2002* (p. 43–46). Boise (USA): PERC Publishing.

Van Driel, J. & Verloop, N. (1999). Teachers' knowledge of models and modelling in science, *International Journal of Science Education* **21**(11), 1141-1153.

APPENDIX. QUESTIONNAIRE AND ANSWERING STRATEGIES

Q1. The term "model" is very common in scientific disciplines, but what actually is the meaning of "model" in physics?

1A) A set of variables or rules or laws or experiments and observations that simplify reality and represent it in a reduced scale.
1B) A simple phenomenon or the exemplification of a phenomenon through an experiment or a reduced scale reproduction of an object.
1C) A mental representation aimed at describing a real object or a phenomenon, which takes into account the characteristics significant for the modeler.
1D) A simplified representation describing a phenomenon aimed at the understanding of its mechanisms of functioning (or at explaining it or at making prediction).
1E) No answer or not understandable answer

Q2. Are the models creations of human thought or do they already exist in nature?

2A) Models really exist and are simple, real life situations or simple experiments and humans try to understand them, sometimes only imperfectly.
2B) Models are simple creations of human thought like mathematical formulas, or physics laws and/or they are what we call theories or scientific method.
2C) Models are creations of human thought and their creation comes from continuous interaction with the ''real'' external world and from its simplification.
2D) Models are creations of human thought aimed at explaining natural phenomena and making predictions.
2E) No answer or not understandable answer

Q3. What are the main characteristics of a physical model?

3A) It must contain all the rules or all the laws for a simplified description of reality and/or it must account for all the features of reality.
3B) It must highlight the variables that are relevant for the description and/or explanation of the phenomenon analyzed (or the object studied) and their relationships.
3C) Their characteristics can classify models as descriptive or explicative or interpretative.
3D) Their main characteristics are simplicity and/or uniqueness and/or comprehensibility.
3E) No answer or not understandable answer.

Q4. Is it possible to build a model for each natural phenomenon?

4A) Yes, every natural phenomenon can be simplified in order to be referred to a given model.
4B) Yes, but the model can still contain errors or uncertainty connected with the possibility (or ability) of carefully reproducing the characteristics we are interested.
4C) No. There are phenomena that cannot be described or explained with a model and/or that cannot be defined in terms of precise physical quantities.
4D) No. There are phenomena that have not been still explained and these, perhaps, will be in the future.
4E) No answer or answer not understandable