OPEN ACCESS
EURASIA Journal of Mathematics Science and Technology Education
ISSN: 1305-8223 (online) 1305-8215 (print)
2017 13(8):4491-4512
DOI: 10.12973/eurasia.2017.00943a

A quantitative analysis of Educational Data through the Comparison between Hierarchical and Not-Hierarchical Clustering

Onofrio Rosario Battaglia
Dipartimento di Fisica e Chimica, University of Palermo, ITALY
Bendetto Di Paola
Dipartimento di Matematica e Informatica, University of Palermo, ITALY
Claudio Fazio
Dipartimento di Fisica e Chimica, University of Palermo, ITALY

Received 15 December 2015 ▪ Revised 18 March 2016 ▪ Accepted 6 March 2017

ABSTRACT

Many research papers have studied the problem of taking a set of data and separating it into subgroups through the methods of Cluster Analysis. However, the variables and parameters involved in Cluster Analysis have not always been outlined and criticized, especially in the field of Science Education. Moreover, in the field of Science Education, a comparison between two different Clustering methods is not discussed in the literature. In this paper two different Cluster Analysis methods are described and the variables and parameters involved are discussed in order to clarify the information that they can supply. The clustering results obtained by using the two methods are compared and showed a good coherence between them. The results are interpreted and compared with the literature. More detail about the relationship between different student conceptions of modeling in physics was obtained.

Keywords: evaluation, hierarchical cluster analysis, modeling, not-hierarchical cluster analysis, science education

INTRODUCTION

Extensive research involving open-ended questionnaires and interviews as well as multiple choice tests 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, subdividing samples of students into intellectually similar subgroups to develop detailed models of students' reasoning (Redfors & J. Ryder, 2001; Mestre, 2002, Bao & Redish, 2006). Particularly, Bao & 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 clusters.

Correspondence: Onofrio Rosario Battaglia, *Dipartimento di Fisica e Chimica, University of Palermo, viale delle Scienze edificio 18, 90100 Palermo, Italy.*
onofriorosario.battaglia@unipa.it

State of the literature

- Quantitative analysis in Science Education is a quite common methodology to analyze different questionnaire. Extensive research involving open-ended questionnaires and interviews as well as multiple choice tests has provided instructors/teachers with tools to investigate their students' conceptual knowledge of various fields of physics. In particular, many research papers have studied the problem of taking a set of data and separating it into subgroups through the methods of Cluster Analysis. However, the variables and parameters involved in Cluster Analysis have not always been outlined and criticized, especially in the field of Science Education. Moreover, in the Science Education research literature a comparison between two different Clustering methods is not discussed.

Contribution of this paper to the literature

- In this paper two different Cluster Analysis methods are described and the variables and parameters involved are discussed in order to clarify the information that they can supply. The clustering results obtained by using the two methods are compared in order to study their coherence and their points of strength and weakness. A synergism between the two clustering methods allows us to obtain more detailed and robust information about the modelling concept.
- Looking at the content from a pedagogical point of view, our study allowed us to obtain more detail about the relationship between different student conceptions of modeling in physics.

The problem of separating a group of students 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 studied through the methods 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 (unsupervised classification, e.g. Battaglia & Di Paola, 2015; Dayan, 1999; Coates & Ng, 2012; Sathya & Abraham, 2013; Battaglia et al., 2016; Battaglia et al., 2017; Di Paola et al, 2016). *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 and Sneath (1963). The application of techniques related to *ClA* is common in many fields, including Information Technology, Biology, Medicine, Archeology, Econophysics and Market Research (Ott, 1999; Mantegna, 1999; Cowgill & Harvey, 1999); Allen & Goldstein, 2013). 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 that have a tendency to be homogeneous with respect to one or more characteristics. The results of the analysis should reveal a high homogeneity within each group (intra-cluster), and high heterogeneity between groups (inter-clusters), in line with the chosen criteria.

ClA techniques (e.g. Everitt et al., 2011) are exploratory and do not necessarily require a-priori assumptions about the data, but they do need actions and decisions to be taken after the analysis (for the interpretation of the results). The selection of variables, the choice of the criteria of similarity between the data, the choice of clustering techniques, the selection of the number of groups to be obtained and the evaluation of the solution found, as well as the choice between possible alternative solutions, are particularly important. It is also important to bear in mind that the result of *ClA,* the subgroups of students, is dependent on the criteria used for the analysis of data as it is typical in all the processes of reduction and controlled simplification of information.

Some studies using *ClA* methods are found in the literature concerning research in education. They group and characterize students' responses by using open-ended questionnaires (Springuel, Wittmann & Thompson, 2007; Fazio et al., 2012; Fazio et al., 2013) or multiple-choice tests (Ding & Beichne, 2009). These papers show that the use of cluster analysis leads to identifiable groups of students that make sense to researchers and are consistent with previous results obtained using more traditional methods. Particularly, Springuel et al. (2007) identify by means of cluster analysis groups of responses in open-ended questions about two-dimensional kinematics. These groups show striking similarity to response patterns previously reported in the literature, and also provide additional information about unexpected differences between groups. Fazio et al. (2012, 2013) analyze students' responses to specially designed questionnaires using researcher-generated categories of reasoning, based on the physics education research literature on student understanding of relevant physics content. Through cluster analysis methods groups of students showing remarkable similarity in the reasoning categories are identified and the consistency of their deployed mental models is validated by comparison with researcher-built ideal profiles of student behavior known from previous research. Ding & Beichner (2009) study five commonly used approaches to analyzing multiple-choice test data (classic test theory, factor analysis, cluster analysis, item response theory and model analysis) and show that cluster analysis is a good method to point out how student response patterns differ so as to classify student.

In a recent paper, Stewart et al. (2012) analyze the student responses to contextually different versions of 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. The authors conclude that *ClA* is an effective method to extract the underlying subgroups in student data and that additional insight may be gained from a further 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.

In this paper, we want to apply cluster analysis in educational research and discuss in detail the use of different ClA methods. Moreover, we want to compare them in order to highlights their coherence and find a pedagogical interpretation of the results. Our paper deals with the application of *ClA* to data coming from an open-ended questionnaire administered to a sample of university students. Particularly, it focuses on the comparison of the quality of the solutions to the clustering problem obtained by applying two different clustering techniques. The comparison involves an analysis of the accuracy of the different algorithms, as well as the possibility to give meaning to the different solutions.

A deep analysis of the *ClA* procedures applied is needed, because they often include approximations strongly influencing the interpretation of results.

CONTEXT AND RESEARCH QUESTIONS

Our research aims at studying the use of *ClA* to analyze the understanding of scientific model concept by our sample of university students. This is done through the analysis of student answers to an open-ended questionnaire investigating the definition of scientific model, its main constituents and its functions. The questionnaire is made up of four open-ended questions, which focus on the understanding of the modeling concept (see **Appendix 1**). They are part of a more complex questionnaire, which has already been validated and used in previous research (Fazio & Spagnolo, 2008; Fazio et al., 2012). We chose to analyze a questionnaire with a low number of questions in order to easily describe the process of clustering data interpretation in terms of the answering strategies related to the questions.

The questionnaire was administered to 124 freshmen of the Information 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. The methodological approach aims at highlighting clusters of students that share representations of scientific model making sense to the researcher. Here, "to make sense to the researcher" means that such representations present a logical coherence and/or have been already described in the literature.

The study of the understanding of scientific model concept by means of *ClA* is performed by starting from a discussion of the cluster parameters that can help us to define the optimal cluster solution. Then, the results obtained by using two different *ClA* methods are compared and interpreted from a Physics Education Research point of view.

The choice of the subject to analyze (the understanding of scientific model concept) is due to its relevance in many education related fields. To achieve an agreed definition of model is intrinsically problematic, since this issue can be tackled under very different points of view, shared, for example, by psychologists, scientists and philosophers of science. However, some aspects are considered relevant in all these fields. These are in our opinion found in the works by the philosopher of science Bunge (1973). According to him, the essential characteristics of a scientific model can be summarized in the following statement that makes explicit the ontological components, as well as the functions of a scientific model: *A scientific model is a representation of a real or conjectured system, consisting in a set of objects with its outstanding properties listed, and a set of law statements that declare the behaviors of these objects, the essential functions of a scientific model being predictions and explanations.*

The fundamental role played by models and modeling activities in the teaching/learning process in math and science education is widely recognized, and many studies present operative approaches in this direction. Particularly, the conceptual "framework" presented in the report of the Committee on Conceptual Framework for New K-12 Science Education Standards (2012) articulates the major practices that scientists employ in developing and using models. Such description also supplies an operational definition of what can be considered an "expert" point of view of the scientific model. Moreover, an extended body of knowledge in the field of science education research about models involved the understanding/conceptions of models' nature of students at different school levels (Grosslight et al., 1991; Pluta et al., 2011; Treagust, Chittleborough & Mamiala, 2002; Fazio et al., 2013) as well as of teachers (Van Driel & Verloop, 1999; Justi, & Gilbert, 2002; Justi, & Van Driel, 2005; Danusso, Testa & Vicentini, 2010). Such research makes explicit student epistemic criteria for evaluating scientific models and the student/teacher knowledge about scientific model, i.e. the role of the scientific model in the process of construction of knowledge, its components (that is the set of entities and laws that relate them) and its functions (i.e. prediction, explanation, testing (Danusso, Testa & Vicentini, 2010)).

The specific research questions that guided our study are:

- RQ1 - To what extent are two different *ClA* methods effective in partitioning students into groups that can be characterized by common traits in students' answers?
- RQ2 - How much are the results obtained by the two methods consistent with each other and with previous literature results?

DATA SETTING

Research in education that uses open-ended questions and presents quantitative analysis of student answers usually involves the development of coding procedures that, as it is well known, present inherent difficulties. Hammer & Berland (2014) point 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 propose guidelines for the presentation of research that quantifies individual student answers. Among such guidelines, they focus on the need to make explicit that: "*developing a coding scheme requires researchers to articulate definitions of categories well enough that others could interpret them and recognize them in the data*". Chi (1997) describes the process of developing a coding scheme in the context of verbal data such as explanations, interviews, problem-solving protocols, and retrospective reports. The method of verbal analyses is deeply discussed with the objective of formulating an understanding of the representation of the knowledge used in cognitive performances.

Following the approach previously described (Chi, 1997; Hammer & Berland, 2014), the logical steps that we use in our research to process data coming from student answers to an open-ended questionnaire can be synthesized by the flow chart represented in **Figure 1**.

The first and second steps (categorization and comparison by the *n* researchers involved in the study) involve the analysis of the records representing student answers (the data), in order to reveal patterns and trends, and to find common themes emerging from them. Through comparison and discussion among researchers, these themes are then developed and grouped in a number of categories whose definition take into account as much as possible the words, the phrase, the wording used by students (Chi, 1997). Such categories can be considered the typical "answering strategies" deployed by the *N* students when tackling the questionnaire. The three authors independently (Patton, 2001; Denzin, 2006) read the students' answers in order to empirically identify the main characteristics of the different student records (the raw data). Each author constructed a coding scheme consisting in the identification of keywords, which characterized student answers. During a first meeting, the selected

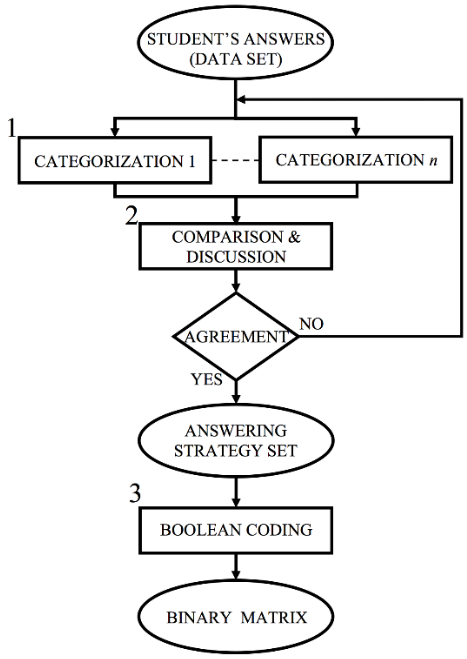

Figure 1. Flow chart of the steps that can be followed by n researchers when processing data coming from student answers to an open-ended questionnaire

keywords were compared and contrasted, and then grouped into "categories" (i.e. the typical answers given by students to the questions) based on epistemological and linguistic similarities[1]. These categories (see **Appendix 1**) were also re-analyzed through the authors' interactions with the data, and taking into account the existing literature about models and modeling (Grosslight et al., 1991; Treagust, Chittleborough & Mamiala, 2002; Pluta et al., 2011). To give an example of the categorization procedure we refer to question Q1 where the categories have been defined as follows:

- Category 1A groups all answers where the model conception reflects a confusion between model and example, or general law, or procedure, or rule, or experiment with the objective of describing phenomena.
- Student answers grouped in category 1B present a conception of model as a simple copy/replica of reality in the case objects (scale models) or phenomena (a simple experiment that models a phenomenon) are referred to. This category can be reported to Level 1 described by Grosslight et al. (1991) or to theme 2 (models as exact replicas) described by Treagust et al. (2002) and many answers supplied by our student sample are similar to those reported in such papers.
- Category 1C groups all answers where the model is clearly presented as a representation (pictures, mathematical expressions, diagrams…, of an entity, simple or complex, that displays a particular perspective or emphasis aimed at describing (or understanding how) such entity behaves.
- Student answers grouped in category 1D present a conception of model as a representation aimed at explaining (to understand why, to explain what happens, to supply a mechanism of functioning, to provide answers to a problem, to predict behaviors,…). Such conceptions of "models as explanatory tools" (Springuel, 2010) can be reported to characteristics of Level 3 conceptions reported by Grosslight et al. (1991), as well as to the primary epistemic criteria discussed by Pluta et al. (2011).

Once the categories have been shared and agreed among the researchers, as a third step of the process each researcher read again the student records and assigned each student answer to a given question to a specific category. Given the inevitable differences among the researchers' interpretations, the three lists were compared

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
...	...	...	...	...
AS_5	1	1	...	0
...	...	...	...	...
AS_M	0	1	...	0

and contrasted in order to get to a single agreed list. Discordances between researcher lists were usually a consequence of different researchers' 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 "not understandable statement". The complete list of 20 categories shared by researchers with respect to the four questions is reported in **Appendix 1**, where examples of specific student answers are also supplied.

Step 3 involves the binary coding of student answers[2], according to the defined categories, generating a binary matrix (as shown in **Table 1**). So, through categorization and coding (steps 1, 2 & 3), each student, i, can be identified by an array, a_i, composed of M components 1 and 0, where 1 means that the student used a given answering strategy to respond to a question and 0 means that he/she did not use it. Then, a M x N binary matrix (the "matrix of answering strategies") 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. As a result of the coding and categorization, we obtained 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 sample element (the student being analyzed).

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, column S_1 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.

DATA ANALYSIS

The matrix depicted in **Table 1** contains all the basic information needed to describe the sample behavior according to the previously described categorization. However, it needs some elaboration to be used for *ClA* (step 4 of **Figure 1**). Particularly, *ClA* requires the definition of new quantities that are used to build the grouping, a quantity to define the likeness between couple of students like the "similarity" or "distance" indexes. These indexes are defined by starting from the M x N binary matrix discussed above.

In the literature (Tyron, 1939; Mantegna 1999; Everitt et al., 2011) the similarity between two elements i and j (in our case between two students) 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).

A distance index can be defined by starting from the Pearson's correlation coefficient. It allows the researcher to study the correlation between students 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 a binary symbolic coding of the answers of students i and j, respectively), we proposed a modified form of the Pearson's correlation coefficient, R_{bin}, (Battaglia et al., 2016) similar to that defined by Tumminello et al. (2011).

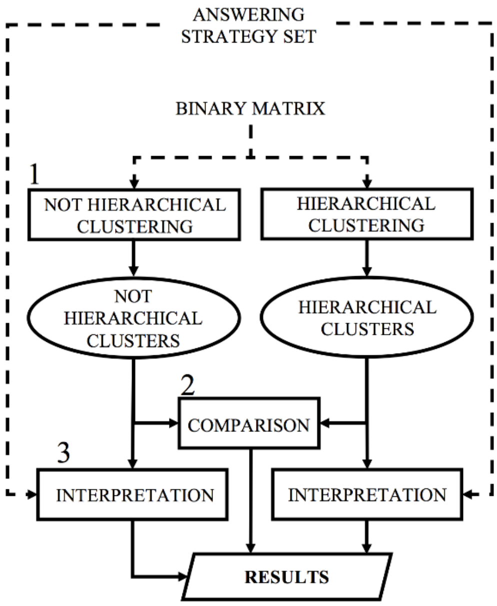

Figure 2. Flow chart of the steps involved in elaborating data coming from student answers to an open-ended questionnaire.

The similarity between students *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. As we want that two students, represented by arrays a_i and a_j, and negatively correlated (R_{bin}= -1) are more dissimilar than two simply uncorrelated (R_{bin} = 0) or positively correlated (R_{bin} = 1), a possible definition of the distance between a_i and a_j, making use of the modified correlation coefficient $R_{\mathrm{bin}}(a_i, a_j)$ is:

$$\boldsymbol{d_{i,j}} = \sqrt{\boldsymbol{2 \cdot \left(1 - R_{bin}(a_i, a_j)\right)}} \tag{1}$$

This function defines a Euclidean metric (Gower, 1966), which is required 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 using Eq. (1) we can, then build a new *N*x*N* matrix, D (the distance matrix), containing all the mutual distances between the students. The main diagonal of D is composed by 0s (the distance between a student and him/herself is zero). Moreover, D is symmetrical with respect to the main diagonal.

Clustering Analysis methods can be roughly distinguished in *Non-Hierarchical* (or *Centroid-Based*), and *Hierarchical* ones (also known as *connectivity based clustering* methods). The first category of methods basically takes to partitions of the data space into a structure known as a *Voronoi Diagram* (a number of regions including subsets of similar data). Non-hierarchical clustering methods are used to generate groupings of a sample of elements (in our case, students) by partitioning it and producing a smaller set of non-overlapping clusters with no hierarchical relationships between them. We will use here the well-known *k-means NH-ClA* algorithm, (MacQueen, 1967).

The second one is based on the core idea of building a binary tree of the data that are then merged into similar groups. This tree is a useful summary of the data that are connected to form clusters based on their known distance, and it is sometimes referred to as a *dendrogram*. Hierarchical clustering is a method of cluster construction and linkage that starts from the idea of elements (again students in our case) of a set being more related to nearby students than to farther away ones, and tries to arrange students representing them as being “above”, “below”, or “at the same level as” one another. This method connects students to form clusters based on the presence of common characteristics. As a *hierarchy* of clusters which merge with each other at certain distances is provided, the term "hierarchical clustering" has been used in the literature.

Similarly, to what we have done in **Figure 1**, we report in **Figure 2** the flowchart of the logical steps we are going to follow here in order to use the two methods of Cluster Analysis we outlined above. This figure should be considered as the continuation of the **Figure 1** flow chart and starts from the resources we will use to perform *ClA* and the subsequent result interpretation, i.e the Binary Matrix and the Answering Strategy Set shown in **Figure 1**.

Step 1 of **Figure 2** shows the two procedure (NH-ClA and H-ClA) of data analysis performed on our data that lead to two different sets of clusters, respectively. These two sets are compared each other to study their mutual coherence, i.e. to what extent the clusters found with NH-ClA are in accordance to the clusters found with H-ClA. Then, each set of clusters is interpreted on the basis of the answering strategy set (as explained in Section 4) and these interpretations, together with the comparison analysis, leads us to the final results of the study.

All the clustering calculations were made using a custom software, written in C language, for the *NH-ClA* (k-means method) as well as for *H-ClA*, where the weighted average linkage method was applied. The graphical representations of clusters in both cases were obtained using the well-known software MATLAB (version 8.6, 2015).

Non-hierarchical clustering results (NH-ClA)

In non-hierarchical clustering 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. The students are then grouped on the basis of the minimum distance between them and the seed points.

Starting from an initial classification, students are transferred from one cluster to another or swapped with students from other clusters, until no further improvement can be made. The students 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 defined as the cluster *centroids* (Leisch, 2006). This process is repeated and ends when the new centroids coincide with the old ones. A remarkable feature of the centroid C_k is that it contains the answering strategies most frequently given by students belonging to Cl_k, as shown in Battaglia et al. (2016) and Battaglia et al. (2017). In order to easily visualize the obtained partition of data, students can be represented in a Cartesian plane according to their mutual distances. As we said before, for each student, i, we know the N distances, d_{ij} between such a student and all the students of the sample (being $d_{ii} = 0$). 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 and Groenen, 1997).

In order to define the number q of clusters that best partitions our sample, the mean value of the *Silhouette function* (Rouseeuw, 1987) $\langle S(q) \rangle$, has been calculated for different numbers of clusters, from 2 to 10 (see Figure 3). 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 $\langle S(4) \rangle = 0.617$, with a 95% confidence interval[3] CI = (0.607, 0.625), indicates that a reasonable cluster structure has been found.

Figure 4 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**). The x- and y-axes simply report the values needed to place the points according to their mutual distance.

It is interesting to study how well a centroid geometrically characterizes its cluster. Two parameters affect this aspect: the cluster density and the number of its elements[4]. To take them into account in a previous paper (Battaglia et al., 2016) we proposed a specific coefficient, r_k, defined as the centroid *reliability*. The higher is r_k for a given centroid, the better the centroid describes the students composing the cluster.

Once the appropriate partition of data has been found, we characterize each cluster in terms of the most prominent answering strategies. Such characterizations will help us to compare clusters and relate our findings to the literature. To do this, we start by creating a 'virtual student' for each of the q clusters, Cl_k ($k = 1, 2,..., q$),

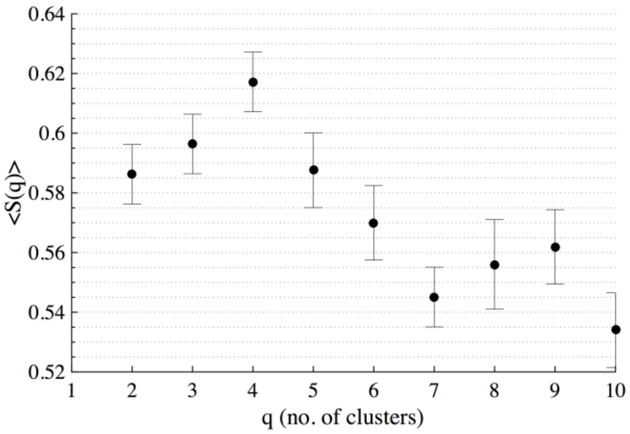

Figure 3. Average Silhouette values and related 95% confidence intervals (CI) for different cluster partitions of our sample. The two highest values are obtained for partitions in q = 4 clusters (<S(4)> = 0.617, CI = (0.607, 0.625)) and in q = 3 clusters (<S(3)> = 0.596, CI = (0.586, 0.603)).

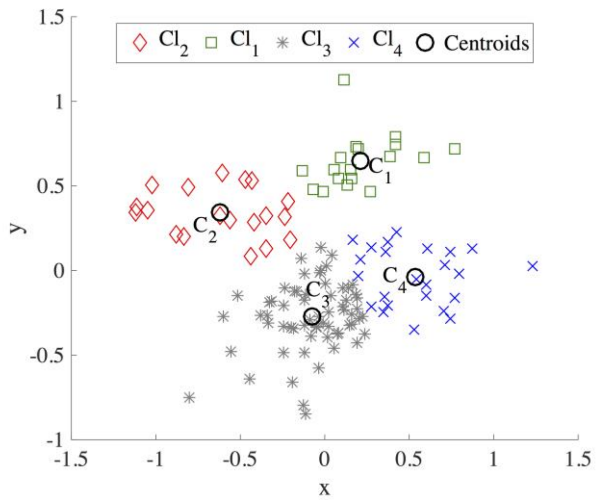

Figure 4. K-means graph. Each point in this Cartesian plane represents a student. Points labeled C_1, C_2, C_3, C_4 are the centroids

represented by the related centroids. Since each real student is defined by an M-dimensional array a_i composed by 0 and 1 values for each of the M answering strategies, the array for the virtual student, $\bar{a}_k$ should also contain M binary entries. It is possible to demonstrate that $\bar{a}_k$ always contains 1 values in correspondence to the answering strategies most frequently used by students belonging to Cl_k[6]. 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 virtual student are maximized and this happens when each virtual student has the largest number of common strategies with all the students that are part of its cluster. This is a remarkable feature of $\bar{a}_k$ that validates our idea to use it to characterize cluster Cl_k.

As discussed above, the four clusters Cl_k $(k=1,\ldots,4)$ can be characterized by their related centroids, C_k. They are the four points in the graph whose arrays, $\bar{a}_k$, contain the answering strategies most frequently applied by students in the related clusters (see **Table 2**). The codes used refer to the answering strategies for the questionnaire

Table 2. An overview of results obtained by NH-ClA

Cluster centroid	C_1	C_2	C_3	C_4
$\bar{a}_k$ (Most frequently given answers)	1B, 2C, 3B, 4A	1B, 2B, 3E, 4A	1C, 2B, 3A, 4A	1C, 2C, 3B, 4B
Number of students	18	19	63	24
$< S(4) >_k$	0.750, CI = (0.730, 0.763)	0.62, CI = (0.58, 0.64)	0.604, CI = (0.590, 0.616)	0.56, CI = (0.53, 0.58)
r_k^{norm}	1.40	-0.02	-0.92	-0.46

items described in **Appendix 1**. **Table 2** also shows the number of students in each cluster, the mean values of the *S-function* $< S(4) >_k$ (*k*=*1,..,4*) for the four clusters and the normalized reliability index r_k^{norm} of their centroids.

The $< S(4) >_k$ value indicates that cluster Cl_1 is denser than the others, and Cl_4 is the most spread out. Furthermore, the values of r_k^{norm} show that the centroid C_1 best represents its cluster, whereas C_3 is the centroid that represents its cluster the least.

Hierarchical clustering results (H-ClA)

In order to apply *H-ClA* to our data, we first have to choose what kind of *linkage* to use to build links among students. Several conditions can determine the choice of a specific linkage method (Everit et al., 2011; Battaglia et al., 2016). For instance, when the source data are in binary form (as in our case) the single and complete linkage methods do not give a smooth progression of the distances (Springuel, Wittmann and Thompson, 2007). For this reason, when the source data are binary, the viable linkage methods actually reduce to the average or weighted average ones.

Since we cannot use simple or complete linkages, we calculate the *Cophenetic Correlation Coefficient* (Sokal et al., 1962; Battaglia et al., 2016) for the average and weighted average linkages, which gives a measure of the accordance between the distances calculated by (1) and the *ultrametric* distances introduced by the linkage. We obtain the values 0.61 and 0.68 for average and weighted average linkages, respectively. We choose to use the weighted average link and **Figure 5** shows the obtained dendrogram of the nested cluster structure.

In this figure the vertical axis represents the ultrametric distance between two clusters when they are joined. The horizontal axis is divided in 124 ticks, each representing a student. Furthermore, vertical lines represent students, or groups of students, and horizontal lines represent the joining of two clusters. Vertical lines are always placed in the center of a student cluster and horizontal lines are placed at the height corresponding to the distance between the two clusters that they join.

By describing the cluster tree from the top down, as if clusters are splitting apart, we can see that all the students come together into a single cluster, located at the top of the figure. In this cluster, for each pair of students, *i* and *j*, the ultrametric distance is $\delta_{ij} \leq 1.8$. Since the structure of the tree shows that some groups of students are more closely linked, we can identify local clusters where students are linked with distances whose values are lower than the previous one. The problem is how to find a value of distance that involves significant links. By using the *Inconsistency Coefficient*, I_k (GhasemiGol et al. 2010; Battaglia et al., 2016), we can define a specific threshold and neglect some links because they are inconsistent. In fact, this coefficient characterizes each link in a cluster tree by comparing its height with the average height of other links at the same level of the hierarchy. The choice of the threshold is arbitrary and should be limited to the links in a specific range of distances (GhasemiGol, Yazdi and Monsefi, 2010), yet it allows us to compare all the clusters and to treat all links with the same criterion.

If we disregard the higher links ($\delta \approx 1.8$, red, dotted links in **Figure 5**) as their use would produce a unique, single cluster of our sample, or two big ones, and we also take into account a threshold for the Inconsistency Coefficients equal to 1.6 (i.e. we consider inconsistent all the links that have $I_k > 1.6$), we are able to accept all the links just below, including the green, dashed ones in **Figure 5** (that have I_k equal to 1.4 and 1.6, respectively). So, we find a partition of our sample into 4 clusters.

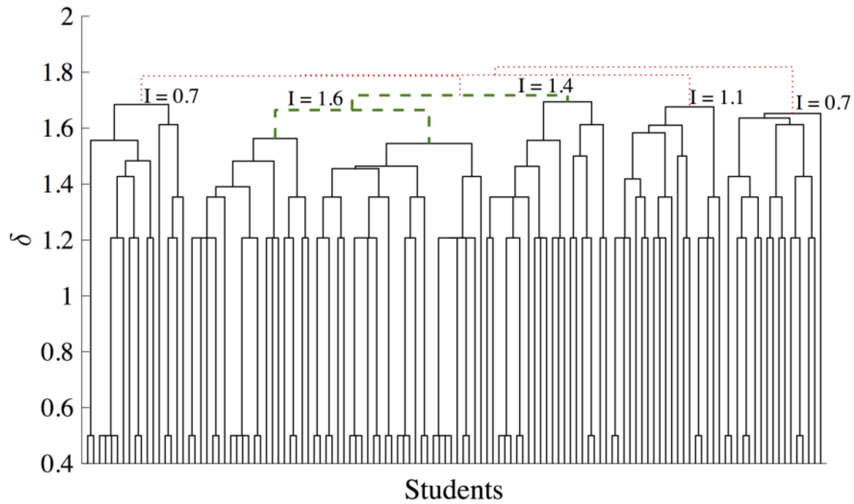

Figure 5. Dendrogram of our data. Horizontal and vertical axes represent students and ultrametric distances, respectively. Red, dotted links are at ultrametric distances of about 1.8. The Inconsistency Coefficients of the links just below these links are shown

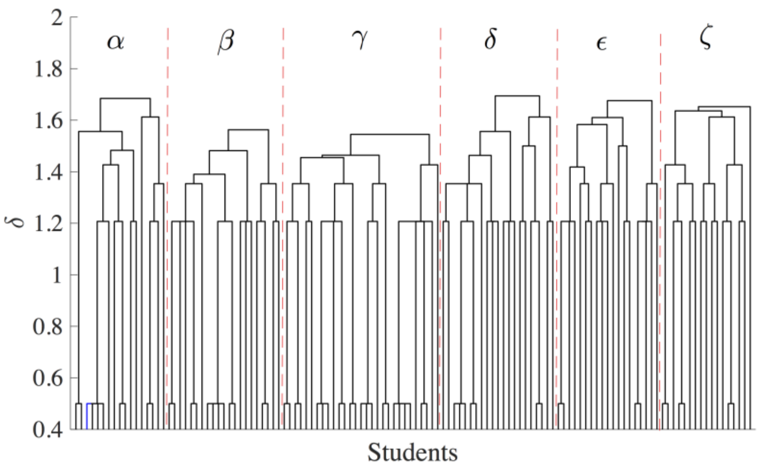

Figure 6. Dendrogram with six different clusters (α, β, γ, δ, ε, ζ), formed according to the Inconsistency Coefficient I_k=1.25

If, on the other hand, we introduce a lower threshold for the I_k value, still not producing a too high fragmentation, like for example $I_k > 1.25$, we must disregard the green, dashed links in the dendrogram in **Figure 5,** and obtain **Figure 6**, where 6 clusters are present. This last representation has a higher significance than the previous one since the links displayed are those that, at equal distances, show a higher consistency.

Figure 6 shows the 6 distinct clusters α, β, …, ζ above identified. They can be characterized by analyzing the most frequent answers to each of the four questions in the questionnaire (see Section 5).

In order to verify the validity of our choice we also used the *Variation Ratio Criterion* (*VRC*) (Calinski and Harabasz, 1974; Battaglia et al., 2016). **Figure 7** shows the *VRC* values for different numbers of clusters. The maximum value is obtained for q = 6.

Table 3 provides significant information concerning our *H-CIA* clustering results. By looking at the number of students, and at their identity, we can see that the main results of the new grouping are the redistribution

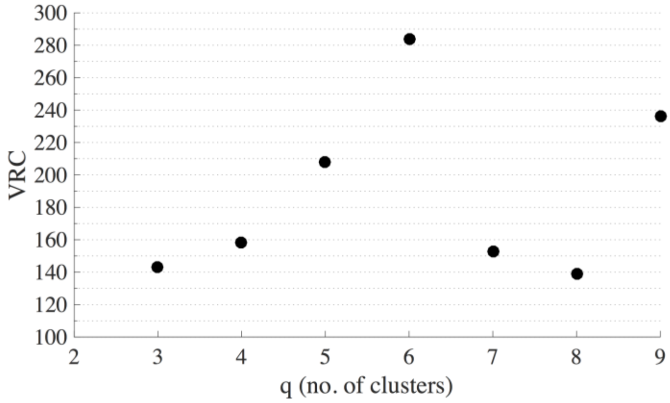

Figure 7. VRC values for some partitions of our sample in different numbers of clusters

Table 3. An overview of results obtained by H-ClA and comparison with those obtained by NH-ClA method

Cluster	α	β	γ	δ	ε	ζ
Most frequently given answers	1C, 2C, 3B, 4B	1B, 2C, 3B, 4A/4B	1B, 2B, 3E, 4A	1C,2B, 3D, 4A	1D, 2C, 3A, 4B	1A, 2A, 3A, 4D
Number of students	17	21	29	21	19	17
Characterization of students in cluster by the k-means method (*)	(14)Cl_4+(3) Cl_3	(18)Cl_1+(3)Cl_4	(19)Cl_2+(10)Cl_3	(19)Cl_3+(2)Cl_4	(14)Cl_3+(5) Cl_4	(17)Cl_3

(*) i. e. (14)Cl_4+(3)Cl_3, means that cluster α contains 14 students part of the cluster Cl_4 (in NH-ClA) + 3 students part of cluster Cl_3.

of the students, originally assigned to cluster Cl_3 by *NH-ClA*, into different sub groups, and a redistribution of students located on the edges of cluster Cl_4. Furthermore, the students in cluster Cl_1 are all located in cluster β and students in cluster Cl_2 are all located in cluster γ. This is in accordance with the high values of the r_k^{norm} coefficient (shown in **Table 2**) for Cl_1 and Cl_2 and the low value for clusters Cl_3 and Cl_4.

In conclusion, we can say that the two different partitions of our student sample are consistent, yet the characterization via the *H-ClA* method allows us to obtain more detail. This happens in particular, in the case of cluster Cl_3, which turns out to be very extensive, with a large number of students and a low value of r_k^{norm}.

In order to better compare the results obtained by *NH-ClA* and *H-ClA* methods, we applied the *Variation of Information* (*VI)* criterion (Meila, 2007; Battaglia et al., 2016), that measures the amount of information gained and lost when switching from one type of clustering to another. We calculated the value of *VI* to compare the 4-clustering results of *k-means* method with the 4-clustering, 5-clustering and 6-clustering results of *H-ClA* method and obtained the values of 0.34, 0.38 and 0.28, respectively. We can conclude that the best agreement can be found between the 4-clustering results of *k-means* method and the 6-clustering results of *H-ClA* method.

In the next section, we analyze the various clustering results from the point of view of the student answering strategies in order to give meaning to the found partitions.

DISCUSSION

Non-Hierarchical Clustering (NH-ClA)

Looking at *NH-ClA* results, four clusters (Cl_1, Cl_2, Cl_3, Cl_4) have been identified. They are characterized by the related centroids and each centroid is represented by one array $\bar{a}_k$, which identifies some answering strategies for each question. These strategies are defined as follows: $\bar{a}_1$ = (1B, 2C, 3B, 4A), $\bar{a}_2$ = (1B, 2B, 3E, 4A), $\bar{a}_3$ = (1C, 2B, 3A, 4A), $\bar{a}_4$ = (1C, 2C, 3B, 4B), where the codes in brackets refer to the questionnaire answer strategies reported in **Appendix 1**. We have already pointed out that the array describing each cluster centroid contains the answers most frequently supplied by the students belonging to 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 answering strategies in many aspects similar to those shared by scientific community (Grosslight et al., 1991; Treagust, Chittleborough & Mamiala, 2002). In fact, these students seem to 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 characteristics we are interested in (4B).* Such students show a conception of model similar to that defined by Grossslight et al. (1991) as "General Level 3", i. e. models as multiple representations, models as construction to test ideas or models as explanatory tools. Such ideas are also described by Treagust et al. (2002) as student relevant ideas in order to understand the role of scientific models in learning science. Pluta et al. (2011) identify such model characteristics as good epistemic criteria for scientific models, too. Clearly, we cannot quantitatively compare our results with the literature ones, since samples are different in terms of context and curriculum. Moreover, the results reported in the literature refer to percentages of individual characteristics (models as multiple representations or models as explanatory tools, …) while our Cl_4 students show a model conception that share all these characteristics.

Students in cluster Cl_2 show, in our opinion, a naive conception of scientific model. They refer to a model as *a simple phenomenon or the exemplification of a phenomenon through an experiment or a reduced scale reproduction of an object* (1B), and express the belief 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 matched to the Level 2 modelers of the classification scheme developed by Grosslight et al. (1991). Level 2 modelers see models as representations of real-world objects or events and not as representations of ideas about real-world objects or events, but they realize that there is a specific purpose that guides the way the model is constructed. Similar results have been also obtained in other studies, as for example the paper of Treagust et al. (2002), that group such conceptions in theme 2: scientific models as exact replicas. Also some studies involving teacher conception of scientific models (Justi & Van Driel, 2005; Danusso, Testa & Vicentini, 2010) identify the origin of such a realistic conception of scientific model from one side on the students' experiences of everyday models (a scale replica or a precise representation which has accuracy and details), from the other side on 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, Chittleborough & Mamiala , 2002). Students in cluster Cl_2, on the other hand, often held beliefs that correspond with a "naïve realist" epistemology (Treagust, Chittleborough & Mamiala, 2002; Pluta et al., 2011).

Students in clusters Cl_1 and Cl_3 do not show a full coherence in their answers, although in different ways. Cl_1 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).

According to the literature that analyzes student and teacher conceptions on modeling in different physics fields (Van Driel & Verloop, 1999; Justi & Gilbert, 2002; Hrepic, Zollman & Rebello, 2005; Ding & Beichner, 2009; Fazio et al., 2013) we can infer that students belonging to this cluster share a ''hybrid'' (Ding & Beichner, 2009) or ''synthetic'' (Justi & Gilbert, 2002) conception of scientific model by referring to composite conceptions that unify different features of naïve conceptions and scientifically accepted ones.

Students in cluster Cl_3 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 to explain their sentences. For example, many of such students agree that motion without friction is a model, as well as the ideal gas, but do not consider motion with friction or the real gases as models, and explicitly mark them only as really existing situations.

On the other hand, it must be taken into account that the value of the reliability, r_k^{norm}, of the C_3 centroid is the lowest, showing that the array $\bar{a}_3$ is not very significant in representing the answering strategies of the cluster students. Moreover, looking in detail at the $\bar{a}_3$ array, the answering strategies are not easily understandable from the point of view of consistency. Although they represent the answers most commonly given by Cl_3 students, they are actually not very frequently used. 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_3 students. In our opinion, this may show that a substructure is present in cluster Cl_3, and this can be analyzed through results of *H-ClA*, which highlight a higher number of clusters than *NH-ClA*.

Finally, it is noteworthy that *NH-ClA* allows us to quantitatively express the different behavior of students in the different clusters by means of a distance parameter which supplies the distances between the cluster centroids. Looking at the distances between couples of centroids in **Figure 4** we can easily identify that C_4 and C_2 are the most far apart, and this corresponds to the maximum difference in the behaviors of students belonging to such clusters.

Hierarchical Clustering (H-ClA)

The six clusters obtained by *H-ClA* are characterized by the answer strategies reported in **Table 3**. It shows that clusters α, β and γ are closely related to clusters Cl_4, Cl_1 and Cl_2 (obtained through *NH-ClA*), respectively. Although containing a slightly different number of students, clusters α, β and γ include the majority of students of clusters Cl_4, Cl_1 and Cl_2, respectively. Cluster α includes more than half of Cl_4 students who mostly exhibit the same characteristics of Cl_4 centroid. Cluster β includes all students previously grouped in cluster Cl_1 and the few added students have not altered the cluster characteristics. Cluster γ includes ten students previously included in Cl_3 and all the ones previously grouped in cluster Cl_2. The answer sheets of all the students that have changed their placing (from cluster Cl_3 to 3 different clusters) were individually analyzed and their position in cluster Cl_3 was identified. Almost all were located on the border between different clusters and this fact may explain the new clustering highlighted by *H-ClA*.

Clusters δ, ε and ζ contain about 80% of students from cluster Cl_3 of *NH-ClA*, which are now divided into three groups. Students in cluster δ give a definition of what a model is analogous to that of Cl_4 students, but they are more focused on the concept of a physics model as mathematical model (*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 on the characteristics of models, *like simplicity and/or uniqueness and/or comprehensibility* (3D). In the modeling process, they seem to give priority to the process of simplification (*every natural phenomenon can be simplified in order to be referred to a given model* (4A)).

Cluster ε groups students with a good understanding of the concept of what a model is. In fact, this group is the only one that includes in the model definition the function of making predictions (*a model is a simplified representation describing a phenomenon aimed at the understanding of its mechanisms of functioning (or at explaining it or at making predictions (1D))*. Moreover, they describe the representational mode of models as *creations of human thought and their creation comes from continuous interaction with the "real" external world and from its simplification* (2C). Similar ideas are also involved in their definitions of model characteristics (*a model must contain all the rules or all the laws for a simplified description of reality and/or it must account for the features of reality)* (3A). The same can be said for their description of the modeling process (*a model can still contain errors or uncertainty connected with the possibility (or ability) to carefully reproduce the characteristics we are interested in* (4B)).

Cluster ζ groups students with a weak understanding of the model concept. Models are described as a *set of variables, rules, laws, experiments or observations that simplify reality and represent it on a reduced scale* (1A). Students in this cluster think that *models really exist and are simple, real life situations or simple experiments and humans try to understand them, sometimes only imperfectly* (2A). Among the models' characteristics they focus *on 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). Consistently with these ideas, they do not think that *all natural phenomena can be modeled. There are phenomena that still have not been explained, but perhaps they will be in the future* (4D).

Answers to the research questions

The analysis of answer strategies elicited by our student sample allows us to answer our two research questions:

- RQ1 - To what extent are two different *ClA* methods effective in partitioning students into groups that can be characterized by common traits in students' answers?
- RQ2 - How much are the results obtained by the two methods consistent with each other and with previous literature results?

Our results show that the *ClA* methods produce partitions of the student sample into groups that are characterized by common trends in questionnaire answers. However, some of such groups are clearly differentiated for their conceptions about the nature, characteristics and function of physics model, while other groups show conceptions only partially differentiated. Moreover, the two methods show a different "sensitivity" in the clustering procedure. In fact, two clusters (Cl_1 and Cl_2) resulting from *NH-ClA* almost completely maintain their individuality in *H-ClA*. The other two (Cl_3 *and* Cl_4) undergo a redistribution of their elements in a larger number of clusters. This was expected on the basis of the parameters characterizing such clusters, i.e. the spreading of Cl_4 and the reliability of Cl_3 centroid. *H-ClA* method reassigns some border line students of Cl_4 to other cluster and distribute the Cl_3 students to three different and smaller clusters. These students nonetheless show consistent answering strategies from the point of view of an expert.

We are aware that these results depend on the characteristics of our sample, our questionnaire and our initial empirical analysis. Moreover, although there is no way to determine whether one cluster analysis method is more accurate than another, we have shown that *H-ClA* supplied more details than *NH-ClA*.

To better answer to our second research question we compared the answering strategies of students belonging to different clusters with similar studies involving the scientific model concepts held by students (Grosslight at al., 1991; Pluta et al., 2011; Fazio et al., 2013) and teachers (Van Driel & Verloop, 1999; Justi, & Gilbert, 2002; Justi, & Van Driel, 2005; Danusso, Testa & Vicentini, 2010) reported in the literature. We can conclude that the results of cluster analysis agree with the results obtained by more common research methodologies. In fact, many of the response patterns showed by the groups of students identified by *ClA* methods bear remarkable similarity to those previously reported in the literature. In particular, we have identified in our data the conceptions characterizing the three general levels of thinking about models described by Grosslight et al. (1991), as well as answers characterized by the five factors/themes identified by Treagust et al. (2002) and the epistemic criteria for good scientific models (Pluta et al., 2011).

However, we point out that our analysis of answering strategies elicited by students grouped in different clusters provides additional information than those reported in previous research results obtained by using more traditional quantitative analysis methods. In fact, in each cluster we found that the meaning assigned by each student to the term model is related to its main characteristics, functions, as well as to procedures for model construction. For example, in our results it is not relevant how many students think models as "*set of variables, rules, laws, experiments or observations that simplify reality and represent it on a reduced scale* (1A)" but, it is relevant how many, among such students, also think that models are *real life situations* (2A) with particular characteristics (3A) and, for these reasons, not all natural phenomena can be modeled (4D) (cluster ζ). Moreover, data show that sometimes students give the same definition of model, but they differ for the other kinds of answers. This is mainly evident in two cases:

1. students belonging to clusters β and γ (that see models as real objects or events) supply the same definition of scientific model, but their interpretation of model functions is different. Students in cluster β see as relevant the problem of simplification of reality, whereas for the others the idea of model as a mathematical formula is considered more relevant;
2. students belonging to clusters α and δ supply the same definition of scientific model (model as mental representation), but the first group of students see such a representation as aimed at the understanding of real system behaviors, whereas the second one focuses on simplicity or uniqueness of models useful for the description of reality. In this second case, it is evident that a different meaning is given to the word representation.

Moreover, the *k-means* results allow us to quantify how the four clusters we identified are different and this gives us insights about how different the students' conceptions are. For example, the distance between clusters C_2 and C_4 reported in **Figure 4** is the highest of all distances between couples of clusters, and this is reflected in the categories expressed by the respective centroids, that are most different. In fact, these centroids represent completely different conceptions held by the students contained in these clusters with respect to characteristics and functions of scientific models, as well as to the modeling procedures.

CONCLUSIONS

In this paper, we discussed the problem of using *ClA* methods in Physics Education Research to study how to identify groups with common behavior, ideas, beliefs and conceptual understanding in a sample of students. We presented two methods of cluster analysis (*NH-ClA* and *H-ClA*) in order to understand the possibilities offered by such methods and their main characteristics. We proposed an example of their application in order to clarify the involved procedures and the different ways of interpreting results. The example is an analysis of the answers to an open-ended questionnaire given to a sample of university students. The results of this analysis indicate that the two methods are consistent, even if not completely, and that *H-ClA* supplies a more detailed partition of our sample into clusters. We discussed the discrepancies through the interpretation of answering strategies of students from different clusters.

It is well known that there is no way to decide whether one clustering method is more significant than another one, as the relevance of each clustering method is related to the research content (Meila, 2007). However, we think that in *H-ClA* the calculations of the consistence/inconsistence of each link (i. e. how relevant the link is in relation to other links in the same hierarchical order) can provide the necessary instrument to analyze clusters in a more detailed way than in *NH-ClA*. We have also outlined what choices are at the basis of each algorithm. For example, in the case of *H-ClA* we had to choose between *average and weighted average linkage* and we chose the second one, on the basis of the cophenetic correlation coefficient values. We chose the weighted average linkage since in this case the results were more consistent with those of *NH-ClA*. Furthermore, it is well known that the results of

ClA are only valuable if researchers are able to give meaning to them (Mantegna, 1999; Kenett et al., 2010), and we have found that our choice made it easier for us to make an interpretation.

It is worth remembering that data from an open-ended questionnaire that are quantitatively analyzed are often the result of an empirical categorization of raw data (the individual student answers). This reduction of the initial data is subject to the researchers' interpretation of student answers that influences the inference about the answering strategies related to such answers. Such an influence can be reduced through a process of categorization shared among the researchers and it should be taken into account when we try to infer typical students' reasoning strategies (Patton, 2001; Denzin, 2006).

Looking at the meaning of the concept of physics model as it is understood by the students in our sample, the results are consistent with those described in the literature, which illustrate a continuum of ideas/beliefs ranging from naive conceptions to constructivist ones. 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. Moreover, we have been able to quantitatively express the different behavior of students in the different clusters by means of a distance parameter related to the correlation among the student answers.

In conclusion, the results obtained with the two *ClA* methods, have provided us with more detailed information than that reported in the literature and obtained using more traditional quantitative methods, since in each cluster the meaning assigned by each student to the term model is related to its main characteristics, functions as well as to procedures for model construction. The characteristics of student clusters previously described allow us to draw this general conclusion: in addition to the group of students who exhibit a conception of scientific model basically consistent with a realistic epistemology, the majority of students use the term "representation" (of objects, events, reality) in their definition of scientific model, yet such a term seems used with different meanings. Some intend representation as merely "the way that something is shown and/or described", others go beyond this conception since they include the need to define the model constituents and their behaviors in order to make a representation able to explain the object/event that is represented. Only very few, that see the representation as a way to refer also to a conjectured system, are able to identify the predictive function as relevant among different functions of scientific models.

NOTES

1. 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 an ontological reality to models.
2. For the sakes of simplicity here we refer to the use of a two-level coding, where 1 means that a given answering strategy was used and 0 means that that strategy was not used.
3. The confidence intervals have been calculated by using the Bootstrap method (DiCiccio & Efron, 1996; Inkley, 1997)] as the distribution of Silhouette values is not a-priori known.
4. For example, two clusters with similar density but different student numbers (i.e. with different variability of student properties) are differently characterized by their centroids: the more populated cluster being less well characterized by its centroid than the other one.
5. The term $1-< S(q) >_k$ in (2) is needed to differently weight $< S(q) >_k$ and n_k because when $< S(q) >_k \rightarrow 1$ the r_k value must be independent of the value of n_k.

It is worth noting that if some answering strategies are only slightly more frequent than the other ones all those with similar frequencies should also be considered.

REFERENCES

Allen, D. N., & Goldstein, G. (eds.) (2013). *Cluster Analysis in Neuropsychological Research: 13 Recent Applications,* Springer Science+Business Media, New York.

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

Battaglia, O. R., & Di Paola, B. (2015). A quantitative method to analyse an open answer questionnaire: a case study about the Boltzmann Factor. *Il Nuovo Cimento, 38C*(3), id 87.

Battaglia, O. R., Di Paola, B., & Fazio, C. (2016). A New Approach to Investigate Students' Behavior by Using Cluster Analysis as an Unsupervised Methodology in the Field of Education. *Applied Mathematics*, 7, 1649-1673.

Battaglia, O. R., Paola, B. D., & Fazio, C. (2017). K-means Clustering to Study How Student Reasoning Lines Can Be Modified by a Learning Activity Based on Feynman's Unifying Approach. *EURASIA Journal of Mathematics, Science and Technology Education, 13*(6), 2005-2038.

Bunge, M. (1973). *Method, Model and Matter*, Springer Netherlands, Dordrecht.

Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in statistics. Theory and methods, 3*(1), 1-27.

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

Coates, A., & Ng, A. Y. (2012). Learning Feature Representations with K-Means, in Neural Networks: Tricks of the Trade, edited by G. Montavon, G. B. Orr, K.R. Muller, 2nd edn. Springer LNCS 7700, Berlin Heidelberg, pp. 561-580.

Cowgill, M. C., & Harvey, R. J. (1999). A Genetic Algorithm Approach to Cluster Analysis. *Computers and Mathematics with Applications, 37*, 99-108.

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

Dayan, P. (1999). Unsupervised Learning in The MIT Encyclopedia of the Cognitive Sciences Wilson, edited by Wilson, R.A. & Keil, F. The MIT Press, London, 1-7.

Denzin, N. (2006). Sociological Methods: A Sourcebook, 5th edition (Aldine Transaction).

Di Paola, B., Battaglia, O. R., & Fazio, C. (2016). Non-Hierarchical Clustering to Analyse an Open-Ended Questionnaire on Algebraic Thinking. *South African Journal of Education, 36*, 1-13.

DiCiccio, T. J., & Efron, B. (1996). Bootstrap confidence intervals. *Statistical Science, 11*(3), 189-228.

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

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

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., & Spagnolo, F. (2008). Conceptions on modelling processes in Italian high school prospective mathematics and physics teachers. *S. Afr. J. Educ., 28*, 469.

Fazio, C., Battaglia, O. R., & 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.

GhasemiGol, M., Yazdi, H. S., & Monsefi, R. (2010). A new Hierarchical Clustering Algorithm on Fuzzy Data (FHCA). *International Journal of Computer and Electrical Engineering, 2*(1), 134-140.

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

Grosslight, L., Unger, C., Jay, E., & Smith C. L. (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. K. (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

Inkley, D. V. (1997). Bootstrap methods and their applications, Cambridge Series in Statistical and Probabilistic mathematics. Cambridge University Press, Cambridge.

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

Justi, R., & Van Driel, J. K. (2005). The development of science teachers' knowledge on models and modelling: Promoting, characterizing, and understanding the process. *Int. J. Sci. Educ., 27*(5), 549–573.

Kenett, D. Y., Tumminello, M., Madi, A., Gur-Gershgoren, G., Mantegna, R. N., & Ben-Jacob, E. (2010). Dominating Clasp of the Financial Sector Revealed by Partial Correlation Analysis of the Stock Market. *Plos One, 5*(12), e15032.

Lerman, I. C., Gras, R., & Rostam, H. (1981). Elaboration et évaluation d'un indice d'implication pour des données binaires I. *Math. Sci. Hum., 74*, 5.

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations, in Proc. 5th Berkely Symp. Math. Statist. Probab. 1965/66 vol. I, edited by L.M. LeCam & J. Neyman, Univ. of California Press, Berkeley, 281- 297.

Mantegna, R. N. (1999). Hierarchical structure in financial markets. *Eur. Phys. J., B11*, 193–197.

MATLAB version 8.6 (2015). Natick, Massachusetts: The MathWorks Inc., www.mathworks.com/products/matlab/

Meila, M. (2007). Comparing clusterings—an information based distance. *Journal of Multivariate Analysis, 98*, 873–895.

Mestre, J. P. (2002). Probing adults' conceptual understanding and transfer of learning via problem posing. *Journal of Applied Developmental Psychology, 23*, 9-50.

NRC (2012). A Framework for K–12 Science Education: Practices, Crosscutting Concepts and Core Ideas. (National Academies Press, Washington,). (http://www.nap.edu/catalog/13165/a-framework-for-k-12-science-education-practices-crosscutting-concepts).

Ott, J. (1999). *Analysis of Human Genetic Linkage.* 3rd Edition. Johns Hopkins University Press, Baltimore, London.

Patton, M. Q. (2001). *Qualitative Research and Evaluation Methods*, 3rd edition. Sage Publications, Thousands Oaks.

Pluta, W. J., Clark, A., Chinn, C. A., & Duncan, R. G. (2011). Learners' Epistemic Criteria for Good Scientific Models. *Journal of Research in Science Teaching, 48*, 486-511.

Redfors, A., & Ryder, J. (2001). University physics students' use of models in explanations of phenomena involving interaction between metals and electromagnetic radiation. *Int. J. Sci. Educ., 23*(12), 1283- 1301.

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

Saracli, S., Dogan, N., & Dogan, I. (2013). Comparison of hierarchical cluster analysis methods by cophenetic correlation. *Journal of Inequalities and Application, 203*, 1-8.

Sathya, R., & Abraham, A. (2013). Comparison of Supervised and Unsupervised Learning Algorithms for Pattern Classification. *International Journal of Advanced Research in Artificial Intelligence, 2*(2).

Saxena, P., Singh, V., & Lehri, S. (2013). Evolving efficient clustering patterns in liver patient data through data mining techniques. *International Journal of Computer Applications, 66*(16), 23-28.

Sokal, R. R., & Rohlf, F. J. (1962). The Comparison of Dendrograms by Objective Methods. *International Association for Plant Taxonomy, 11*(2), 33-40.

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

Springuel, R. P. (2010). *Applying cluster analysis to physics education research data*, PhD Thesis (available from https://www.academia.edu).

Springuel, R. P., Wittmann, M. C., & Thompson, J. R. (2007). Applying clustering to statistical analysis of student reasoning about two-dimensional kinematics. *Phys. Rev. ST Phys. Educ. Res., 3*, 020107.

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. J. (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. *Int. J. Sci. Educ., 24*(4), 357-368.

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

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

Van Driel, J. H., & Verloop, N. (1999). Teachers' knowledge of models and modelling in science. *Int. J. Sci. Educ. 21*(11), 1141-1153.

APPENDICES

Appendix 1

Questions and typical answering strategies and examples of specific student answers

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

Answering strategy subset for question 1	Examples of real student answers
1A) A set of variables or rules or laws or experiments and observations that simplify reality and represent it in a reduced scale.	*A model is a general law or an abstract method, used to represent reality in a reduced way.* *A model is a set of conventional rules, experiments and observations aimed to simplify and describe Nature*
1B) A simple phenomenon or the exemplification of a phenomenon through an experiment or a reduced scale reproduction of an object.	*As model we intend a physical phenomenon simplified by means of an experiment.* *A model is an object that copy in a small scale a real one.*
1C) A mental representation aimed at describing a real object or a phenomenon, which takes into account the characteristics significant for the modeler.	*A model is an idealization of a phenomenon that allows the researcher to describe what he thinks about its characteristics.* *A model is a product of the researcher mind. It is aimed to describe an object or a phenomenon and its features.*
1D) A simplified representation describing a phenomenon aimed at the understanding of its mechanisms of functioning (or at explaining it or at making prediction).	*A model is an ideal representation of a phenomenon that allows us to explain how the phenomenon works.* *A model is a representation of reality based on the scientific method that allows us to explain what happens and also to make predictions.*
1E) No answer or not understandable answer	

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

Answering strategy subset for question 2	Examples of real student answers
2A) Models really exist and are simple, real life situations or simple experiments and humans try to understand them, sometimes only imperfectly.	*A model is a natural phenomenon that is reproduced in laboratory to be studied.* *A model is a simple experiment that we do to reproduce a physical situation and to try to understand it, often only roughly.*
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.	*Models are creations of human mind, expressed in a mathematical form.* *A model is part of the scientific method. It is created by human thought and it is resumed in laws and theories.*
2C Models are creations of human thought and their creation comes from continuous interaction with the ''real'' external world and from its simplification.	*Models are the creation of human thought. They are abstractions coming from the real phenomena in order to simplify them.* *A model is the creation of human mind drown from scientists' observation of natural phenomena.*
2D) Models are creations of human thought aimed at explaining natural phenomena and making predictions.	*A model is an artificial creation of man. It is based on the observation of Nature and is aimed at explaining it.* *Models are created by the human mind and are aimed at explaining Nature and making predictions.*
2E) No answer or not understandable answer	

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

Answering strategy subset for question 3	Examples of real student answers
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.	*A physical model is characterized by a mathematical formulation that allows us to completely describe the variables we really observe.* *The main characteristics of a model are the laws that simplify the description 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.	*A model is useful if it puts in evidence the variables relevant to understand the phenomenon.* *A model must set the relationships among the variables that we measured during the observation/experimental phase.*
3C) Their characteristics can classify models as descriptive or explicative or interpretative.	*A model must allow us to describe and explain what's happening in nature.* *A model is a way to understand the nature.*
3D) Their main characteristics are simplicity and/or uniqueness and/or comprehensibility.	*A model should mainly be comprehensible, so to be used by everyone.* *A model must be clear and unique, so to not give ambiguous answers to our questions.*
3E) No answer or not understandable answer.	

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

Answering strategy subset for question 4	Examples of real student answers
4A) Yes, every natural phenomenon can be simplified in order to be referred to a given model.	*Yes. We can always find a simple model for each natural phenomenon.* *Yes, because we can choose a simplification level for the model. So, it is always possible to build a model for a phenomenon.*
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.	*Yes, but all depends on the complexity of the phenomenon. In some cases we cannot completely reproduce a phenomenon.* *Yes, but we will always be able to only partially reproduce a phenomenon, due to uncertainty in measurements and in their description.*
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.	*Probably not, because we cannot always find variables for all natural phenomena. An example is biology, where we cannot simplify the functioning of the nucleus of the cell with precise variables, but take all into account.* *No, because we cannot know all the variables relevant for the description of the phenomenon*
4D) No. There are phenomena that have not been still explained and these, perhaps, will be in the future.	*No. There are phenomena that are not explained due to our limitations and technology.* *No, because in some cases we don't have the right mathematical tools to build a model to the level we want. But in the future we will probably have them.*
4E) No answer or answer not understandable	