UNIVERSITÀ DEGLI STUDI DI CAMERINO

School of Advanced Studies

DOCTORAL COURSE IN

SCIENCES AND TECHNOLOGY - THEORETICAL AND EXPERIMENTAL PHYSICS

XXXII cycle

METHODS AND TECHNIQUES IN EXPERIMENTAL DETERMINATION OF LABORATORY ACTIVITIES' EDUCATION IMPACT: THE PHYSICS LABORATORY ACTIVITIES IN UNIVERSITIES COURSES.

PhD Student
Dr. Ido Borsini

Supervisors
Prof. David Vitali

Co-supervisors
Dr. Daniela Amendola

Abstract.

From the point of view of the educational impact on university students, four different ways of carrying out physics experiments are compared. The didactical tools suitable for simulating, by means of an e-learning platform, the inquiry-based and the recipe-based laboratory activities are constructed and described. The tests are constructed, and validated by means of Item Response Theory and Factor Analysis techniques:

- for building the necessary statistical samples,
- for determining the educational impact of the four different didactical paths.

The statistical techniques adopted for evaluating the educational impact of the different didactical paths are described in their theoretical details, also providing, for each one, an explicit example of use.

The HTML-JavaScript algorithms needed to simulate the experiments in e-learning modality are described in detail. Furthermore, a computer tool is proposed that can help the teacher to produce simulations of thermodynamic systems subjected to thermal and adiabatic interactions.

The results of the statistical experiment are given as a functions of the statistical variables proficiency, awareness of knowledge acquired.

The statistical experiment confirms that:

- the inquiry-based didactic paths, especially if realized in the e-learning modality, reach the maximum educational impact from the point of view of the knowledge of the topics,
- the didactic paths carried out in the laboratory, especially for the inquiry-based experiments, reach the maximum educational impact from the point of view of the awareness of the acquired knowledge.

Finally, our study confirms recent finding in the literature, i.e., that when the didactic path is carried out in inquiry mode, the students can construct robust and original knowledge of the proposed subject as an evidence of the students' acquired agency.

Index

Introduction: from page 6 to page 12.

Bibliography: from page 13 to page 14.

Chapter 1 The theoretical Framework

1.1 Overview on theories of intelligence: from page 15 to page 17.
1.2 Learning as adaptation and experimentation: from page 18 to page 20.
1.3 The causal explanation: from page 21 to page 22.
1.4 The steps of causality: from page 23 to page 25.
1.5 Laboratory Experiments: Inquiry and Recipe Learning Activities: from page 26 to page 30.
1.6 The structure of the research: from page 31 to page 32.
1.7 Activity planning: from page 33 to page 34.
Bibliography: from page 35 to page 37.

Chapter 2 The statistical methodologies adopted.

2.1 Factor Analysis: from page 38 to page 41.
2.2 Principal Component Analysis (PCA): from page 42 to page 46.
2.3 Example of Explorative Factor Analysis (EFA): from page 47 to page 49.
2.4 The Item Response Theory (IRT): from page 50 to page 54.
2.5 Estimation of parameters: from page 55 to page 59.
2.6 Example of IRT: from page 60 to page 63.
2.6 The Analysis of Variance (ANOVA): from page 64 to page 68.
2.7 Example of ANOVA: from page 69 to page 70.
Bibliography: from page 71 to page 72.

Chapter 3 The Didactical Tools

3.1 E-learning structure: from page 73 to page 75.
3.2 A real simulator of thermal interactions: from page 76 to page 80.

3.3 Generalization of the real Simulator of Thermodynamic System: from page 81 to page 84.
3.4 The system reproduces the results of Classical Statistical Mechanics: from page 85 to page 86.
3.5 The real Brownian Motion simulator: from page 87 to page 88.
3.6 Generalization of the real system through an algorithm: from page 89 to page 92.
3.7 The system reproduces the results of Brownian Motion theory: from page 93 to page 95.
3.8 Computer tools: from page 96 to page 98.
3.9 The details of algorithms: from page 99 to page 103.
Bibliography: from page 104 to page 104.

Chapter 4 The Implementation and the results of the didactic experimentation.

4.1 The Placement Test: from page 105 to page 106.
4.2 The validation of Placement Test: from page 107 to page 111.
4.3 The equivalent work groups creation: from page 112 to page 113.
4.4 Multiple Knowledge and Abilities Test: from page 114 to page 116.
4.5 The analysis of scores: from page 117 to page 122.
4.6 The analysis of Awareness: from page 123 to page 127.
Bibliography: from page 128 to page 128.

Conclusions and future work: from page 129 to page 134.

Bibliography: from page 135 to page 135.

Introduction.

The construction of interpretation of inside and outside world and of its causal connections by children is a continuous, spontaneous, and very efficient experimental activity [1] [2]. From the sixties of the last century, this has stimulated an uninterrupted flow of didactic methodologies discussed by many authors into various disciplinary contexts at different degrees of school and university education [3] [4].

It has been noted that the feature of freedom, spontaneity, continuity, and efficiency of the construction of reality and causality through trials and experiments remain unchanged only for scientists [2], while educational institutions sometimes can even alienate those characteristics [5] [6]. In this regard, opposing conclusions are documented in the literature regarding the value of laboratory practices in science learning [7] [8] [9].

As scientific knowledge evolves to the extent that it supports its claims with experimental evidence, statement concerning student learning require the same standard and, if the results are not conclusive, or at least stable with respect to the conditions in which they are determined, it is difficult to draw definitive conclusion from them.

It must also be considered that the management of the laboratories is expensive, labour intensive, so that didactic effectiveness of laboratory becomes the condition to get rid of them [10].

The spread of Information and Communications Technology (ICT) and E-Learning techniques has helped to partly remove the economic constraints imposed on the didactic use of scientific laboratories. At the same time, the problem of assessing their educational impact has become a relevant topic in the literature in recent year [11] [12] [13] [14].

At the state of the art, comparing and evaluating the different ways of carrying out laboratory experiences in degree courses seems to be the most direct way to make the necessary educational choices.

Comparing and evaluating must be done:

- quantitatively i.e., measuring the resulting proficiency of the students,
- qualitatively i.e., measuring the agency [15] and the awareness of the knowledge acquired [16].

To achieve the purpose, it is necessary to answer the questions:

1) In what terms and forms is it possible to carry on Inquiry-Based and Recipe-Based laboratory experiments through laboratories and e-learning platforms?
2) In what terms and limits, can the comparison between different methods of carrying out laboratory experiments be made independent of the students subjected to the didactic stimuli and of the experiments proposed?
3) Does the level of awareness of the knowledge, and the students' agency acquired through the laboratory activity depend on the different methods of carrying out laboratory experiments?

Interesting even though non definitive answers to these questions are obtained and documented in this thesis.

In the first chapter three theories of intelligence are briefly classified, examined, and compared, extracting from each one the theoretical hypothesis and technical tools useful for this research. In particular:

- from Psychometric Theories of Intelligence: Factorial Analysis (FA), Item Response Theory (IRT), Analysis of Variance (ANOVA),
- from the Implicit Theories of Intelligence: the necessity to measure the level of awareness of the competences, acquired through the teaching activity,
- from the Functional Theories of Intelligence: the theoretical framework which we think is the more proper one to place the entire research.

The theoretical apparatus adopted leads us to consider experimental activity as a general form of "search for causality", which consists in verifying that one's ideas can explain in a consistent way a class of phenomena. This function is part of a more general process of adaptation to the relevant environment. In this setting, the influence of Piaget's thought is recovered and resumed in the description of the different degrees of acquisition of causality, from infancy to maturity [17].

Subsequently, the Recipe Based Laboratory Activity (RBLA) and the Inquiry Based Laboratory Activity (IBLA) are detailed. Based on the literature, these two didactic activities are characterized and compared, their limits and criticalities are highlighted, some strategies are indicated and adopted to overcome those difficulties. The RBLA and IBLA methods combined with two modalities of carrying out the experiment, in laboratory (Lab) or through e-learning technologies (E-learning), form the set of laboratory activities that will be compared.

Once the theoretical context has been described, the proper strategy suitable for answering the research questions is identified:

- the procedure suitable for establishing equivalent statistical units in the form of equivalent student work groups is described,
- four laboratory experiments suitable for reproducing the four combinations of methods and modalities are identified,
- the program of didactical activities that lead to the formulation of the answers to the previously mentioned research questions is indicated.

The second chapter is devoted to the theoretical and critical examination of the statistical tools used in the thesis.

We review the birth and development of Factor Analysis, indicating the essential contributions of the Simple Factor Model and of the Simple Structure Model as the crucial points of the development of the theory.

The two main statistical techniques in which the Factor Theory is currently declined, Explorative Factor Analysis (EFA) and Confirmative Factor Analysis (CFA), are briefly resumed.

CFA will be used to verify the consistency of the tests employed for the creation of the equivalent statistical units. For "test consistency" is meant that, based on the score, the tests are really capable of ascertaining some individual cognitive abilities.

EFA will be used to analyze the behavior of the statistical variable awareness on the summative final test. To clarify to the reader the details of the use of this technique in the context of assessment activities, a school test example is provided.

In the second part of the chapter, the Item Response Theory (IRT) is described as a theory based on traits (proficiency) and parameters (difficulty, discrimination, guessing), which are a consequence of the invariance of the performance with respect to the test raw score. The aim of the theory is to calculate the probability of answering correctly to an item as a function of the subject's proficiency level and item parameters.

We show the mathematical steps that, from the formulation of the Maximum Likelihood principle on an appropriate probability function, lead to the joint estimation of traits and parameters of the theory, using the Newton-Raphson method [18]. The calculation of parameters will allow to verify the consistency of the single items of the tests used in the assessments. In particular, the combined use of the indexes of difficulty and discrimination will select the items to be eliminated because unable to provide an adequate measure of the objective they referred to.

We will also deal with the problem of establishing the performance of the model in reproducing the actual situation by means of the Fit of Model matrix and Infit, Outfit vectors, giving an example within the one-parameter Logistic Model (1PL) named Rasch Model.

In the last part of chapter, it is theoretically shown that the analysis of variance (ANOVA) generalizes the statistical technique of comparing two samples. To show this property, multiple samples from a normally distributed population are extracted and their behaviour is modified in a differential way, according to a structural linear model. The calculation retraces the procedure that leads to the definition of the random variable F which is associated with the Snedecor-Fisher distribution.

The statistical test on variable F will allow to provide the criterion for determining the equivalence of the statistical units (student work groups), at the same time, the statistical test on variable F will provide the criterion to ascertain the effect of didactic experiment on the statistical units, as it is required by the research questions.

The third chapter describes the didactic tools used in the eight educational paths that we will compare: two, Lab and E-Learning, for each of the four experiments:

- Gay-Lussac pressure and temperature law,
- linear thermal expansion coefficient of solids,
- Boltzmann statistics,
- Brownian motion.

We will argue that, while for the RBLA the experiment can be carried out even through a video, from which it is possible to extract the necessary measures, for the IBLA this possibility is obviously denied. For the IBLA it will be necessary to use real or analogous systems or integrate the video of one of the possible experiments, with software that creates the mathematical model of systems under study. This is the strategy that will be used for experiments carried out through the e-learning platform.

Two analogous real systems are designed and described.

The first analogous systems are a chessboard with pawns capable of carrying out a Markovian type process. Starting from a non-equilibrium state, the simulator converges to the results of the Bose-Einstein statistics and therefore, approximately, to the Maxwell-Boltzmann, as students will be experimentally able to prove. The system solves the Boltzmann problem i.e., the determination of the state with maximum entropy, it is able to simulate the real system, and its ergodic property will be experimentally proved by the students.

For the second analogous system, an analogy is established between the central point of hexagons and the instantaneous position of a Brownian Particle on a plan, with or without drift. The system, together with the application of the elementary Sampling Theory, solves the differential equation for the probability density of stochastic variable r (the distance from initial point) at time t, as students will be experimentally able to prove. In particular we will see the particular relationship, between the mean and mean square deviation of the stochastic variable r, for two-dimensional Brownian motion.

After discussing the advantages, that make analogous systems fundamental for the realization of educational objectives programmed through laboratory experiments, a didactic application capable to simulate thermodynamical interactions through statistics is discussed. A two-level software structure is identified, Teacher-Level and Student-Level, which allows to develop and use the analogous system, in relation to the teaching needs. The two levels are supposed to have HTML-JavaScript front-end and PHP back-end. The variables necessary to determine the thermodynamical state and how to simulate the adiabatic interaction are indicated.

The last chapter of the thesis is devoted to the description of the procedures followed and the results obtained.

The first part of chapter describes the construction, administration, and validation through Factor Analysis and IRT techniques, of the Differential Aptitude Test (DAT) based on three dimensions: Verbal Reasoning, Abstract Reasoning and Logical Arithmetic Reasoning. On the basis of the results of the DAT, 12 work groups (statistical units) are formed with the same mean and variance, each made up of 5 students. The units were grouped into 6 pairs A, B, so that the same experiment can be performed with different Modality by each pair. Once the conditions of applicability have been verified (through a Bartlett’s test), it is shown that the test ANOVA confirms the equivalence of the work groups.

After determining the training objectives to be observed in order to evaluate the effect of the didactic activities, we have carried out, the tool used to measure the achievement of these objectives is identified in a typical Multiple Knowledge and Abilities Test (MKAT).

Then we describe the procedure which allows students' awareness of knowledge acquired to be assessed through MKAT scores.

The last part of the chapter is reserved for the analysis of the proficiencies and awareness acquired by the students in relation to the different didactical paths followed.

The last chapter will be devoted to concluding remarks and to the discussion of possible future research perspectives.

Bibliography.

[1] J. Piaget - La construction du réel chez l'enfant. - 1937 - psycnet.apa.org.

[2] A. Gopnik - The scientist as child - Philosophy of science, 1996 - journals.uchicago.edu.

[3] J. S. Bruner - The act of discovery - 1961 - Harvard Educational Review. 31 (1): 21-32.

[4] E. von Glasersfeld - Radical Constructivism, A Way of Knowing and Learning – 1995.

[5] Taylor, G., Jungert, T. Mageau, G. A. Schattke, K. Dedic, H. Rosenfield, S. & Koestner. - A self-determination theory approach to predicting school achievement over time: The unique role of intrinsic motivation - 2014 - Contemporary Educational Psychology, 39(4), 342–358.

[6] T. Gnambs, B. Hanfstingl - The decline of academic motivation during adolescence: an accelerated longitudinal cohort analysis on the effect of psychological need satisfaction - Journal - Educational Psychology - Volume 36, 2016.

[7] A. Hofstein, V. N. Lunetta - The Role of the Laboratory in Science Teaching: Neglected Aspects of Research – Review of Educational Research - 1982

[8] A. Hofstein, V. N. Lunetta - The laboratory in science education: Foundations for the twenty-first century – Science Education – 2003.

[9] Cole Walsh, Katherine N. Quinn, C. Wieman, N. G. Holmes1, Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking, PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH 15, 010135 (2019).

[10] S. Lowery Bretz - Evidence for the Importance of Laboratory Courses - J. Chem. Educ. 2019, 96, 2, 193-195.

[11] Jing Ma, J. V. Nickerson - Hands-on, simulated, and remote laboratories: A comparative literature review - ACM Computing Surveys - September 2006.

[12] Gheorghe SCUTARU, et al. Relazione finale sull'uso di esperimenti simulati e remoti nella formazione, Education and Culture, Socrates Minerva (2015)

[13] M. A. Bochicchio, A. Longo - Hands-On Remote Labs: Collaborative Web Laboratories as a Case Study for IT Engineering Classes - IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, vol. 2, n. 4, October-December 2009.

[14] R. H. L. De La Torre, D. Galan, F. J. Cabrerizo, E. Herrera-Viedma, S. Dormido - Virtual and remote labs in education: A bibliometric analysis - Computers & Education – vol 98 – July 2016.

[15] BANDURA, Albert. Toward a psychology of human agency. *Perspectives on psychological science*, 2006, 1.2: 164-180.

[16] MORIN, Alain. Self-awareness part 1: Definition, measures, effects, functions, and antecedents. *Social and personality psychology compass*, 2011, 5.10: 807-823.

[17] J. Piaget, E. Duckworth, (1970). Genetic epistemology. *American Behavioral Scientist*, *13*(3), 459-480.

[18] Rossi, Richard J. Mathematical Statistics: An Introduction to Likelihood Based Inference, New York: John Wiley & Sons, 2018.

Chapter 1: The theoretical framework.

1.1 Overview on theories of intelligence.

There are several theories of intelligence, that for our purposes we consider classified into Psychometric Theories, Implicit Theories, Functional Theories. Boundaries of each of these classes are not well defined because, trying to classify a complex area of human knowledge is always non-trivial, especially when the various concepts and approaches are closely related to each other. We will use this classification to place the cultural context of our research and extract from those areas all the theoretical and technical tools necessary for our analysis [1] [2].

Psychometric Theories of Intelligence or Explicit Theories of Intelligence [3] [4] are based on the answers given by subjects to specific tasks, aimed at detecting intelligent behaviour. For example, a test on mental abilities is carried out and based on the answers, the factors of the intelligence, influencing the performance, can be detected.

The theory is based on the evidence that, who answers a test correctly will have the same trend in other similar trials, as who answer incorrectly will continue to do the same. It is therefore hypothesized to be able to measure not only the factor of the intelligence, but also the differences in the intelligence of the subjects.

The theories of Differential Intelligence study the differences between qualities of individual intelligence. Their purpose is to identify intrinsic abilities or specific factors of intelligence that characterize and differentiate individuals [5].

Over time, powerful mathematical and statistical techniques have been developed to measure the multifactorial features of intelligence, to compare different samples and to validate the tools used in the measurements [6]. We will describe and use: Factorial Analysis (FA), Item Response Theory (IRT), Analysis of Variance (ANOVA).

The Implicit Theories of Intelligence [7] are based on the observation, that individual's beliefs about what intelligence is, influence her/his cognitive performance, seen as objectives, motivations, behaviours, and self-esteem.

Theorists have observed that raw intelligence, such as that measured by IQ, was unable to predict whether persons are able to face an unprecedented cognitive challenge, while knowledge of their implicit intelligence allowed to do it.

The various implicit intelligences can be collected in two classes related to two observable cognitive behaviours:

- Entity Theory of Intelligence: the individual considers the intelligence a fixed and innate trait of human behaviour, so that, faced with a problem to be solved, she/he will behave in relation to her/his perception of being able to solve it.
- Incremental Theory of Intelligence: the individual considers intelligence a trait of human behaviour that can be enhanced through commitment, study, and improvement of skills. The individual modulates the commitment in facing the challenge with the perception of being close or far from the solution of the problem.

Although these theories seem to depend on the theoretical conceptions of the scientists who formulate them, highlighting the connections between the personal beliefs and the individual cognitive performances suggests, that the value of the acquired knowledge should be measured, not only by having given the correct answer to a question, but also by the awareness of the correctness of the given answer [8] [9].

Since each answer to a dichotomous question can relate to the score, in order to measure the perception of the validity of the answer, we define the statistical variable Awareness, which measures the learning value of the teaching process leading to that answer on a sample of individuals.

It can be said that the individual shows a Functional Intelligence when it is expected that:

- she/he learns from the environment in which operate,
- she/he adapts her/his behaviour on bases of what she/he has learned

- she/he recognizes the adequacy or inadequacy of her/his behaviour, on the bases of knowledge further acquired,

to pursue her/his objectives.

Theories of Functional Intelligence consider learning the maximum manifestation of intelligence. The pedagogists’ efforts are oriented to detect and describe:

- the complex mechanisms through which knowledge is built,
- the relationships between the individual and relevant environment,
- the way in which knowledge is retained, modified, and used to solve new problems.

Within these theories we adopt Piaget's point of view, as a methodological tool. Piaget considers intelligence as the continuous capacity of adaptation to the relevant environment through “assimilation” and “accommodation” processes [10].

1.2 Learning as adaptation and experimentation.

Each learning theory of Constructivist inspiration is based on the hypothesis that Knowledge is a form of adaptation to the relevant environment [11][12].

By adaptation we mean the biological process through which the individual maintains equilibrium towards both the external world, in the domain of individual characteristics, actions and behaviours, and toward its own "internal" world, in the domain of mental operations and ideas [13]. These two domains depend on one another, and their interaction generates, the sensory-motor intelligence and the abstract rational thought, starting from the infant's genetic endowment [14].

Following Piaget's Genetic Epistemology [15], Knowledge is a construction due to the Abstraction Process, that everyone performs on objects and actions, consequently the value of the acquired knowledge derives from its actual functioning in the field of everyone's experience, and not from an imaginary function of representation, independent from the person who knows. The Behavioural schema "stimulus-action" must be replaced by the Constructivist Scheme [16]:

perceived situation --- action ---- expected result.

The concept of "scheme" is used in cognitive sciences to formalize an acquired knowledge [17] being the Scheme a unit of knowledge on a thing, a subject or an event, linked to other knowledge, which is based on experience, which is accessed to guide the recognition, the understanding, or the action.

The Action Scheme lends itself to formalizing the process of adaptation and hence, the process of building knowledge.

The perception in the Action Schema is a simplification of the stimulus, it tends to select the only features, that make the situation recognizable. In fact, situations never present themselves in the same way and in the same context of the perceived and the perceiver. Whenever we recognize that a situation is a repetition of a past situation, we say that Assimilation has taken place [18].

The Assimilation within the limits of perception is an equivalence relation on the set of situations and, as every equivalence relation, it performs two functions, it classifies our experiences, generates new abstract objects structuring rational thought.

Since Assimilation proceeds by simplifications of the situation, the action may not produce the expected result, generating an unpleasant state of individual’s imbalance. In the perception some unknown characteristics of the situation have prevented the action from leading to the expected result. The individual can commit to recognize the factors that have not been considered, she/he can try to remove them and ascertains whether the scheme of action is conserved, in which case it can be said that an Accommodation has been made to a new situation.

It also happens that, to restore equilibrium, the individual redefines the list of situation features that have to be taken into account, a new Action Scheme is produced, which leads to a different expected result: it is said that an Accommodation was made to a new Action Scheme [19].

In this context, knowledge is considered a personal acquisition, not an approximate representation of an ontological world and the fact, that a certain knowledge offers advantages to those who possess it, does not attribute to that knowledge the value of definitive acquisition, rather it determines the adaptive force of the community of people who possess that knowledge towards the relevant external environment.

It must also be considered that people expressing concepts, elaborating ideas, having opinions, relating to each other, and directing their lives in a conscious way, are the ones who are more suited to the environment in which they live. To induce Accommodation to new Action Schemes, as in learning, it is necessary that the new experiences disrupt equilibria and question the acquired knowledge. The process would not occurs in an unprotected environment in which the individual's confidence in the teacher is not expressed, for at least one recognized purpose.

1.3 The causal explanation.

The causal explanation [20] is the cognitive process characterizing the laboratory activity. To understand its nature, it is useful to recall the difference between empirical connection, legality in the Piaget' words, and causality. In legality the explanation is always observable, while in causality it is always deduced and depends only on mental operations, in a particular sense that we will describe [21].

Legality involves the repetition of operations, the reproduction of phenomena, the confirmation of the expected result. This does not mean that deductive schemes such as the Syllogism are not used to establish the Legal explanation of a phenomenon: "... we limit ourselves to insert particular laws in other more general ones, subsequently deriving the first by syllogism ..." [22]. In any case, to confirm the functional value of Legal Explanation, we do not need to demonstrate neither the correctness of reasoning, nor the existence of objects we have established through experience.

Unlike the legal one, the causal explanation is constructive. The objects, the particular and general laws that can be deduced must be found, to the desired degree of approximation. In other words, there is a causal explanation when the individual finds a correspondence between what he can establish in her/his deductions and what the objects actually do in reality: “.... if legality can be limited to the level of phenomena, without having to decide on the reality or uselessness of any supports, causality requires that "the object exists". Hence the continuous search for objects in all phases, whose historical beginnings date back to the time when, without any supporting experience or any suspicion of the experimental method, the Greeks arrived at the heroic hypothesis of a world of atoms which, by composing themselves, created the qualitative diversity of reality.” [23].

Resuming:

1) a causal explanation is given if there is an explicit relation between the domain of mental operations and ideas, the domain of actions and behaviours of the external world.
2) The connection between the two domains shows the constructive function of causal explanation in enrichment of individual’s cognitive structures and in enrichment of external experienced world.
3) In searching for the causal explanation, the expected result in the action scheme is to expect that, the relation between the two domains exists.

These features of the causal explanation are proper of and are realized in the laboratory activity.

1.4 The steps of causality.

We summarize the evolution of the search for causality by referring to the principal mathematical structures that are progressively and unconsciously activated in the cognitive growth of the individual up to adulthood.

An example of recognizable primitive form of causality is the patterns of the permanent object: the object that child looks for behind the screen that hides it [24].

Permanent objects and their movements are inseparable from a causal structure indeed, the characteristic of objects is of being the source, the scenario or the result of different actions, whose links constitute the category of causality. Manipulation of the permanent object builds a further level of causality, which concerns the sensorimotor knowledge of the Practical Group of Displacements:

- closure property is ascertained by the conduct of choosing different paths to move from a point to another
- associative property is identified by the conduct of turning around the obstacle,
- existence of the inverse is established because the opposite of a movement brings the object back to the starting point,
- existence of the null element is ascertained by the fact that the object can be left where it is.

The system of sensorimotor assimilation patterns induces equivalence relations and order relations on the sets of permanent objects, while the sequences of generated and observed movements contributes to the construction of functional relationships between space and time. At the end of this path the child can situate himself as a permanent object among permanent objects in space-time having laid the foundations for the subsequent developments.

Up to now the child refers only to concrete objects that can be classified, ordered, listed and therefore every judgment and every logical form of reasoning starts from concrete observations and not from hypotheses.

In the evolution of the child's rational thinking, it is possible to observe a progressive liberation of relations and classes from their practical content, a progressive ability to combine objects of different classes to form other classes and other relations. This great combinatorial freedom leads to the possibility of operating on ideas and hypotheses, no longer only on concrete and verifiable data. This revolution is completed when the adolescent shows that, she/he indifferently uses order and equivalence relations in her/his reasonings and deductions.

In the Bourbakist mathematical architecture [25] the lattice is considered as an order structure which is interpreted within the proposition logic and calculus of classes. The equivalence relations

$$(a \rightarrow b) \leftrightarrow (a \cup b = b)$$

$$(a \rightarrow b) \leftrightarrow (a \cap b = a)$$

are given which show the complete equivalence of the partial order relation and the equivalence relation. These relations are verified by segments, areas and volumes and are freed from their meaning, to give rise to a causality research structure, called by Piaget "of the two reversibility", very well studied on teenagers aged from 12 to 15. [26].

Let's consider the material implication $I = p \rightarrow q$, its inverse $N = p \wedge \neg q$, the reciprocal $R = q \rightarrow p$, the correlative $C = \neg p \wedge q$, being p, q proposition describing external events. It should be noted that $N=R \circ C$, $R=N \circ C$, $C=N \circ R$, $I=N \circ R \circ C$ and that the set $\{I; N; R; C\}$ with binary operator $\circ$ is an Abelian group [27]. It is observable the process of searching for the cause of a phenomenon through the following steps:

- If p is cause of q, i.e., $p \rightarrow q$, then it must be excluded that the inverse $p \wedge \neg q$ holds.
- If the reciprocal of $p \rightarrow q$, $q \rightarrow p$ is valid, i.e., q is cause of p, it must be excluded that its inverse $\neg p \wedge q$ is valid.
- $\neg p \wedge q$ is the correlative of $p \rightarrow q$ because if p is the cause of q however it may happen that q holds without p.
- $p \wedge \neg q$ is the correlative of $q \rightarrow p$ because if q is the cause of p however it may happen that p holds without q.
- Finally, $\neg p \wedge q$ is the inverse of $p \wedge \neg q$.

This conduct shows the unconscious use by adolescents of the INRC group in the search for the causes of a phenomenon.

1.5 Laboratory Experiments: Inquiry and Recipe Learning Activities.

Laboratory experiments are Inquiry-Based (IB) or Recipe-Based (RB).

The experimental activity generalizes a natural conduct presents, in analogous form, in all the phases of individual's intellectual development [27][28]. A model of this activity is the conduct, which leads a child from initially genetic-programmed body motions to the practical knowledge of Euclidean group of Isometries [29], or which leads scientists through the practices of authentic research processes.

Recipe-Based Laboratory Activities (RBLA) are used for the confirmation of physical laws.

Students are provided with all the knowledge they need to take, to complete the experiment, and while this will give them the chance to focus on technical expertise and analysis, it does not engage them in the construction of knowledge or in discovering new properties or relations.

Properties and relations of the objects of the experiment are a prerequisite for the activity to be performed. Furthermore, it is necessary to be sure that the students know the operating principles of the instruments used and how to use them.

Inquiry-Based Laboratory Activities (IBLA) [30] [31] incorporates the design process of experiment, builds knowledge, facilitate the emergence of independent thinking and creativity in problem-solving. This methodology exposes the students to an authentic research process, as they do not know the results of the experiment, training them to think like scientists:

1) students should go through the entire research process.
2) The results of the experiment should have some degree of value in terms of novelty, not only for the students.
3) Inquiry-based learning should be conducted independently by students.

The first point commits us to design laboratory experiences that, starting from legal explanations, lead the student towards the individual synthesis of abstract knowledge, connected with the behaviour of the system studied.

The tutoring function performed by teachers it is necessary. It should consist in suggesting the quantities to be determined, in proposing problems to be solved, in suggesting relations between quantities and foresights to be reached as confirmation of the existing causal explanation.

The second point implies a close connection between the teaching and research activities carried out by the teacher. Teachers are deeply immersed in the same educational environment as students and are subject to cognitive processes that are of the same nature as those of the learner. Under certain conditions, real, analogous, or simulated laboratory experiences derived from the teacher's personal research work can be used. We list some possibilities:

- In fundamental scientific research, data obtained from experiments conducted elsewhere are increasingly used, especially if complex and expensive instruments are used or if the experimental apparatus is spread on a big territory: we refer to these experiments with the expression Remote Experiment. These experiments provide large amounts of data that need to be structured and correlated by the causal explanation. The process presents a constructive nature, especially when data coming from different experiments are correlated, constituting the basis for new and original knowledge [32].

- The analogy keeps together the systems having the same mathematical model. There are cases in which laboratory experiment of physical systems are difficult, expensive, or simply not possible, but within the set of analogous systems it can be identified the one for which it is possible to carry out the laboratory experiment. The Analogous Experiment has the same complexity of the experiments connected in the analogy and it is source of potential discoveries [33].

- Experimental data can be the result of numerical elaboration of mathematical models, we call these experiences Simulated Experiments. A plenty of software has been produced and used for educational purposes, this software implements mathematical models, giving the student the possibility to infer the properties of the studied physical system. Another less tested and far more powerful possibility is use of professional software for educational purposes. The practical advantages offered by this possibility are evident for what concern the scientific rigor of the model and the time saved for the design of the computer tool.

The third point concerns the creation of a structured relationship environment to stimulate the collaborative skills between the work groups. This operation cannot be extemporaneous and limited to a short period, it must encourage the spontaneous construction of interest groups based on the tasks to perform, and it must support the transformation of groups and the ability to produce and exchange specific knowledge.

This approach also involves redefining students' assessment. Collaborative interactions, past grading performance, goals achieved must be evaluated from log data, together with disciplinary competences, using advanced grading methods inside a learning environment [34].

The IBLA have been subjected to criticisms in the field of learning; two of them:

1) It could create cognitive overload,
2) It could create potential misconceptions.

We are referring to learning as a biological process in which Assimilation and Accommodation are the fundamental step. Every biological process has its specific duration. Respecting the natural times is the essential element for the success of teaching, in facing cognitive overload. The strategy adopted is to increase students' educational contact time with the experiment, designing easily achievable experiments in environments other than the laboratory. In this thesis, we obtained this goal using two analogous systems. The first one is analogous to a thermodynamical system subject to thermal interaction, the second one is analogous to a two-dimensional Brownian motion.

The e-learning platform is another powerful tool in facing cognitive overload, being able to asynchronously provide videos, algorithms of experiments and all the necessary didactical supports.

The second point of criticism is concerned with misconceptions [35].

The term misconception is not appropriate if we refer to incorrect knowledge. A knowledge has its own domain of validity and works for that precise domain, if this does not happen, the knowledge does not survive.

The definition of a valid knowledge domain is an epistemological problem, we gave an idea of our point of view on this topic in the previous paragraph, now we want to argue of the direct educational consequences of that definition.

In Science, the correctness of a knowledge depends on the sensitivity of the measure, on the experimental context in which it is established, on the state of development of the discipline, and on the validity of the deductible consequences of that knowledge (causality). Therefore, if it seems convenient to talk about knowledge in relation to its domain of validity, then we must consider a mistake as the fruit of a valid knowledge, in the context where it is originated, and no longer valid in a new context. This evidence suggests the adoption of a strategy of conducting the educational process, that is more consistent with the awareness, that knowledge is a form of adaptation to the relevant environment.

The necessity to understand the roots of the incorrect knowledge, the effort to take the point of view of the bearer of the incorrect knowledge and the acceptance of its reasonableness, the necessity to make the learner aware of limits of the incorrect knowledge as a prerequisite to change it, the opportunity to confirm the new knowledge by testing more contingent or more general facts are points through which the educational process must grow and establish itself.

1.6 The structure of the research.

In the previous paragraphs, we identified two methods of didactic use of experiments: we talked about Inquiry-Based (IB) and Recipe-Based (RB). We superimpose on these methods, two modalities of carrying out the experiments: we used the word "Lab" for laboratory experiments, "E-Learning" for those experiments carried out through the e-learning platform.

The objective of this research is to determine and compare the impact of laboratory experiences on the learnings, in relation to the four combinations of methods and modalities. To achieve this goal, a first condition is that the sample on which to carry out the survey is made of equivalent statistical units, from the point of view of learning capacities. A second condition is that the sample shows a similar behaviour for different experiences carried out with the same modality and method.

To ensure that the two properties are verified, it is necessary to compare equivalent work groups on the same experience carried on in RB and IB teaching methods and in all modalities i.e., Lab and E-learning. This result can be achieved using at least four different experiments (Fig 1).

What does "equivalent work groups" mean in this research? As a working hypothesis, the equivalent statistical unit is the work group that have the same mean and variance of the other work groups of the sample, based on a battery of calibrated psycho-aptitude tests carried out on every student constituting groups, on three factors:

- Verbal reasoning,
- Abstract reasoning,
- Logical-Arithmetic reasoning.

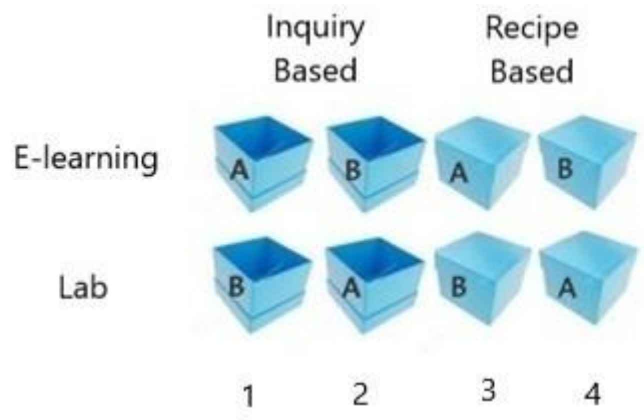


Fig 1. The structure of the experiment: the rows of the matrix correspond to the modality of carrying out the experiment, the columns the method used, A and B are referred to groups of the same couple of statistical samples of couples.

1.7 **Activity planning**.

The structure of the didactic experiment has eight sections which correspond to four laboratory experiences:

1) Gay-Lussac pressure and temperature law.
2) Determination of the linear thermal expansion coefficient.
3) Bose Einstein statistics.
4) Brownian motion.

each one carried on in two modalities Lab and E-Learning.

The first two experiments are carried on in the RB method, the other two, in the IB method.

For Lab modality, three gas thermometers, three dilatometers, three chessboards with pawns and three set of hexagons are provided.

Students know the “recipes” of the RB experiences and the operation of the instruments for the IB experiments.

The theoretical topics of the RB experiments were treated in the lessons, the topics of Statistical Mechanics were proposed to the students only from the macroscopic point of view, as it is traditional in a General Physics course.

For the E-learning modality, videos and simulations replaces the experimental apparatus.

In the RB method, experimental apparatus properties and measures necessary to carry out experiments can be obtained from video. To reproduce the real behaviour of the experiment, video editing is carried out within Moodle's Lesson activity. The Lesson module was designed to be adaptive and to use the student's choices to create a self-directed lesson: each choice the students make shows a different teacher response/comment and send the students to a different page in the lesson.

The structure of the teaching material in E-Learning modality and IB method is slightly more complex. As we will argue more in detail in the third chapter, in addition to reproducing some parts of the experiment through video editing into a Moodle lesson activity, it is necessary to provide students with a mathematical model of the system being studied. The model is provided in the form of interactive simulation, to allow students to develop conjectures and draw conclusions from the experimental results. Tutorials are developed on mathematical aspects, such as fitting techniques and use of functions and representation of data in Excel.

At the end of the educational paths, a summative test is scheduled on each of the four experiments carried out. Since all the work groups will be exposed to the four experiments, their differences in the evaluations will be correlated with the particular combination of Methods and Modalities to which the equivalent work groups have been exposed. The characteristics of the evaluation tool will be described in chapter 4.

Bibliography.

[1] Sternberg, R.J. (1987). Teorie dell'intelligenza. Milano: Bompiani.

[2] Sternberg, R.J. (2000). *Handbook of intelligence*. Cambridge University Press.

[3] Spearman, C.E. (1946). Theory of the general factor, *British Journal of Psychology,* 36**,** 117–131.

[4] Thurstone, L.L. (1938). *Primary mental abilities*. Chicago: University of Chicago Press.

[5] Elliott C.D., Murray G.J., Pearson L.S., Differential ability scales, San Antonio, Texas, 1990

[6] Timothy A. Brown, Confirmatory factor analysis for applied research, RH Hoyle Sage Publications, Inc, (2015).

[7] RJ Sternberg - Formal methods in developmental psychology, 1987 – Springer

[8] Carol S. Dweck, Ellen L. Leggett, A Social-Cognitive Approach to Motivation and Personality, Psychological Review 1988, Vol. 95, No. 2,256-273.

[9] Dweck, C.S., Chiu, C., & Hong, Y. (1995). Implicit theories and their role in judgments and reactions: A world from two perspectives. *Psychological Inquiry, 6*(4), 267-285.

[10] J Piaget, The psychology of intelligence, Routledge Classic 2001

[11] Henry C. Plotkin. Darwin machines and the nature of knowledge. Harvard University Press, 1994.

[12] Ernst von Glasersfeld. Radical Constructivism: a way of knowing and learning, Routledge. 6 ago 2013.

[13] Piaget J. (1975). L'équilibration des structures cognitives. Paris: PUF.

[14] Piaget, J. & Inhelder, B. (1969). *The Psychology of the Child.* NY: Basic Books.

[15] Piaget J, Duckworth E. Genetic Epistemology. American Behavioral Scientist. 1970.

[16] William Baker, Olen Dias, Bronislaw Czarnocha, Edme Soho. Creating action schema based on conceptual knowledge. DIDACTICA MATHEMATICAE Vol. 38, 2016, p. 5–31.

[17] David E. Rumelhart. Theoretical Issues in Reading Comprehension Perspectives from Cognitive Psychology, Linguistics, Artificial Intelligence and Education; Chapter 2: Schemata: The Building Blocks of Cognition. Edited by Rand J. Spiro, Bertram C. Bruce, William F. Brewer.

[18] Berger, Kathleen Stassen (2008). The developing person through the life span (7th ed.). Worth. p. 44, 45.

[19] Jack Block. "Assimilation, accommodation, and the dynamics of personality development". Child Development. **53** (2): 281–295 (1982).

[20] J Piaget, Understanding causality, Trans. D. & M. Miles (1974).

[21] Jean Piaget, L’ Epistemologia Genetica. From pag. 427 to pag.473. Edizioni Studium – Roma, 2016.

[22] Jean Piaget. L’ Epistemologia Genetica. Pag. 433. Edizioni Studium – Roma, 2016.

[23] Jean Piaget. L’ Epistemologia Genetica. Pag. 172. Edizioni Studium – Roma, 2016.

[24] Piaget, Jean, and Barbel Inhelder. *The psychology of the child*. Basic books, 2008.

[25] Nicholas Bourbaki - The Architecture of Mathematics - The American Mathematical Monthly, Vol. 57, No. 4. (Apr., 1950), pp. 221-232.

[26] Piaget J, Duckworth E. Genetic Epistemology. American Behavioral Scientist. 1970.

[27] Bruner, J. S. (1961). The act of discovery. Harvard educational review.

[28] Bruner, J. S. (2009). The process of education. Harvard University Press.

[29] Piaget, J, and Barbel I. The psychology of the child. Basic books, 2008.

[30] Boyer Commission on Educating Undergraduates in the Research University (1998). Reinventing undergraduate education: A blueprint for America's research universities. Stony Brook: State University of New York at Stony Brook.

[31] Ludwig Huber. Reflection. In Inquiry-Based Learning-Undergraduate Research. The German Multidisciplinary Experience. Harald A. Mieg Editor 2019. Springer Open.

[32] M.Abbrescia et al.; A study of upward going particles with the Extreme Energy Events telescopes; Nuclear Instruments and Methods in Physics Research A 816 (2016)

[33] Steinhauer, J. Observation of self-amplifying Hawking radiation in an analogue black-hole laser. *Nature Phys* **10,** 864–869 (2014).

[34] O. Petropoulou, K. Kasimatis, I. Dimopoulos, S. Retalis. LAe-R: A new learning analytics tool in Moodle for assessing students' performance. Bulletin of the IEEE Technical Committee on Learning Technology, Volume 16, Number 1, January 2014.

[35] Bruno D'Amore, Silvia Sbaragli. Analisi semantica e didattica dell'idea di "misconcezione". *La matematica e la sua didattica*. 2, 139-163. (2005).

Chapter 2: The statistical methodologies adopted.

2.1 Factor Analysis

As it is known, the correlation index $\rho(x; y)$ and the covariance index $cov(x; y)$ give us information on the concomitant trend of two random variables x and y,

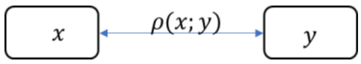


Fig 1: elementary graph representing the correlation between two random variables.

without telling us if x is the cause of y or vice versa (Fig 1).

An assessment test is built starting from the identification of a large number of random variables whose reciprocal relations generally show a complex nature. It is convenient to reduce the number of random variables for better understanding of this complexity and these relations. In order to reduce the number of variables, it is assumed that a certain number of them $x_1, x_2, x_3, \ldots$ has a relationship with a factor f responsible (cause) for their concomitant variation: f is called "common factor" of variables $x_1, x_2, x_3, \ldots$ and represents the trait that influences the behaviour of the variables (Fig 2).

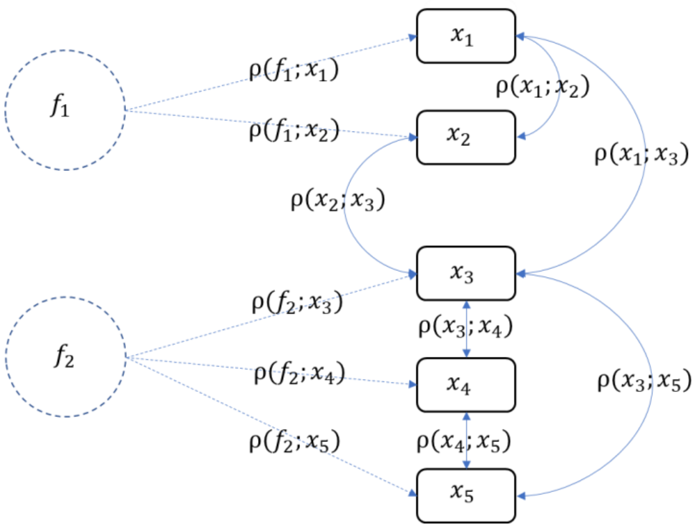


Fig 2: Example of factorial structure.

The features of the common factors are explained by the variables involved in the grouping, and the grouping of the variables is explained by the existence of the common factor. Therefore, the factor analysis has an inductive function in grouping the variables and in characterizing the factor, a deductive function in using the features of the factor to determine its action on the single variables.

The factor loadings $\rho(f;x)$ are equal to the value of the correlation $cor(f;x)$ or the value of covariance $cov(f;x)$ and measure the weight of the relationships between the factor and the random variables. The process of grouping random variables, can be done through the factor loading and it is useful to reduce the complexity of system if the number of factors will be less than the number of variables.

Factor analysis was first proposed by Spearman (1904) [1]. The author formulated a factorial model called simple factor model, according to which answers given to a set of skill tests can be traced back to a single general factor of intelligence, called g factor, and from "An infinity of specific capacities called s factors" (FIG 3). The general factor is superior to the specific factors, which can be considered a sub-differentiation of the g factor. The factor g determines the performance in all intellectual tasks, while each factor s determines the resolution of a single specific task.

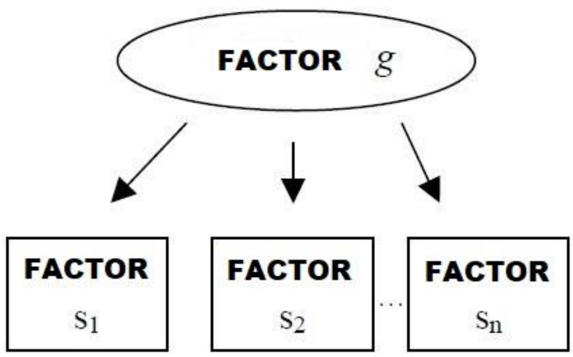


FIG 3: Graphic representation of Spearman's theory

Thurstone (1945) [2] proposed a modification of Spearman's theory known as multifactorial theory. He believed that there was not a single general factor but rather several common factors; the factors (traits) are independent and each one manifests itself on groups of measurable variables. His theory does not provide a hierarchical structure of the multiple factors identified but a disordered disposition, where each dimension is equally important and has the same weight (FIG 4). The author pointed out that an initial factor solution rarely lends itself to a simple interpretation. To facilitate the interpretation of the factors, Thurstone introduces the rotation procedure of the extracted factors, in order to obtain the "simple structure", that is, a structure in which an observer variable saturates (depend only on) on a single latent factor and shows negligible saturation on the other extracted factors. However, Thurstone's contribution to the technique of factor analysis does not end in "simple structure" concept. In fact, the author had the great merit of identifying the fundamental equation of the factor analysis expressed as

$$\boldsymbol{\Sigma}^2 = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Phi}$$

that represents the mail theorem of factor analysis.

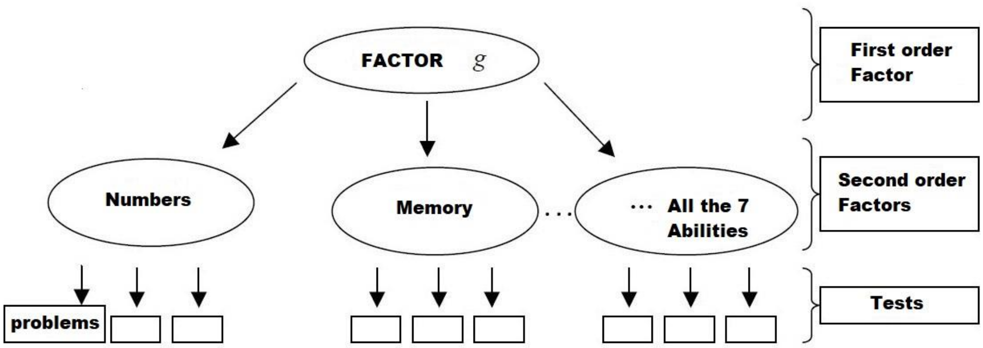


FIG 4: Graphic representation of Thurstone's theory.

Thurstone's theory is the theoretical framework of modern Factor Analysis based on multiple regression.

At the beginning, the theory was used for the exploration of the relations between many variables i.e., as an Exploratory Factor Analysis (EFA).

Progressively, with the introduction of the structural equations, factor analysis is increasingly used for the confirmation that hypothesized factors explain the measured variables, i.e., as Confirmatory Factor Analysis (CFA) [3].

EFA and CFA share the same mathematical methods with Principal Component Analysis (PCA). As we will see, the searching for common factors will coincide with the searching of components in spectral decomposition of the variance-covariance matrix.

2.2 Principal Component Analysis [4].

Throughout this paragraph the vectors and matrices will be indicated with bold letters unless the indices are explicit.

Let's consider the vector $\boldsymbol{x}$ of the p random variables (traits) x_i

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

which has the mean $\boldsymbol{\mu}$ on the statistical population

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_p \end{pmatrix}$$

and the vector$\boldsymbol{f}$ of m common factors f_i with $m \leq p$

$$\boldsymbol{f} = \begin{pmatrix} f_1 \\ \vdots \\ f_m \end{pmatrix}.$$

Suppose the mathematical model that describes the relationship between statistical variables and common factors is linear and takes the form:

$$\boldsymbol{x} = \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{f} + \boldsymbol{\varepsilon}$$

with the vector $\boldsymbol{\varepsilon}$ of specific factors ε_i

$$\boldsymbol{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{pmatrix}.$$

The mathematical model is a series of multiple regressions, which takes the explicit form

$$\begin{cases} x_1 = \mu_1 + l_{1,1}f_1 + \cdots + l_{1,m}f_m + \varepsilon_1 \\ \vdots \\ x_p = \mu_p + l_{p,1}f_1 + \cdots + l_{p,m}f_m + \varepsilon_p \end{cases}$$

The $l_{i,j}$, called factor loading coefficients, have index $\boldsymbol{i}$ linked to the random variable and index j linked to the factor. The problem is solved if $l_{i,j}$ and $\boldsymbol{\varepsilon}$ are determined from the experimental data $\boldsymbol{x}$.

The following relations are used to simplify the calculation [5]:

1) $\mu(\boldsymbol{\varepsilon}) = \mathbf{0}$
2) $\mu(\boldsymbol{f}) = \mathbf{0}$
3) $var(\boldsymbol{f}) = \mathbf{1}$
4) $var(\boldsymbol{\varepsilon}) = \boldsymbol{\Phi}$
5) $cov(f_i; f_j) = 0 \quad \forall i \neq j$
6) $cov(\varepsilon_i; \epsilon_j) = 0 \quad \forall i \neq j$
7) $cov(\varepsilon_i; f_j) = 0$

The direct consequences of these hypotheses are that

$$\mu(\boldsymbol{x}) = \boldsymbol{\mu}$$

$$cov(x_i; f_j) = l_{i,j}.$$

The calculation of variance and covariance of random variables establishes that

$$cov(x_i; x_j) = \sum_{k=1}^{m} l_{i,k} l_{j,k} + \Phi_i \delta_{i,j}$$

Putting together these last two expressions, we obtain the following expression for the variance-covariance matrix $\boldsymbol{\Sigma}^2$

$$\boldsymbol{\Sigma}^2 = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Phi}$$

being the specific variance $\boldsymbol{\Phi}$ a diagonal matrix with the elements $\boldsymbol{\Phi_i}$ on the main diagonal.

The quantity $\sum_{j=1}^{m} {l_{i,j}}^2$ is called communality of the random variable x_i. The larger the communality the better the model describes the real system for the $\boldsymbol{i-th}$ variable.

Now consider a statistical sample of size n extracted from the population considered above, let $\boldsymbol{x}_i$ the vector of the $i-th$ statistical unit of the sample

$$\boldsymbol{x}_i = \begin{pmatrix} x_{i,1} \\ \vdots \\ x_{i,p} \end{pmatrix}$$

which has mean $\boldsymbol{m}$

$$\boldsymbol{m} = \begin{pmatrix} m_1 \\ \vdots \\ m_p \end{pmatrix}.$$

$\boldsymbol{S^2}$ denotes the sample variance-covariance matrix, which is a $p \times p$ symmetric matrix, expressed by

$$\boldsymbol{S^2} = \frac{1}{n-1} \sum_{i=1}^{n} (\boldsymbol{x}_i - \boldsymbol{m})(\boldsymbol{x_i} - \boldsymbol{m})^T$$

$\boldsymbol{S^2}$ admits p eigenvalues as well as corresponding p orthonormal eigenvectors

$$\lambda_1, \dots, \lambda_p$$

$$\hat{\boldsymbol{e}}_1, \dots, \hat{\boldsymbol{e}}_p$$

which are expressed in such a way that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_p$.

The spectral decomposition allows us to express $S^{\boldsymbol{2}}$ as a function of eigenvalues and eigenvectors

$$\boldsymbol{S}^2 = \sum_{i=1}^{p} \lambda_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i^{\,T}$$

being $\boldsymbol{S}^2$ the estimator of $\boldsymbol{\Sigma}^2$.

The most important feature of the factorial analysis is that it gives the possibility to reduce the number of traits necessary to obtain a complete description of the system studied. Naively we could provide the criterion for reducing complexity by referring to the magnitude of the eigenvalue. Actually, the following theorem holds:

Let $\boldsymbol{x}$ be the vector of p random variables, let $\boldsymbol{S}^2$ be the invertible square symmetric matrix variance-covariance of a sample of $\boldsymbol{x}$, let $\widehat{\boldsymbol{a}}_1, \widehat{\boldsymbol{a}}_2, \dots, \widehat{\boldsymbol{a}}_p$ be its eigenvectors, then the random variables $y_1 = \widehat{\boldsymbol{a}}_1 \boldsymbol{x}$, $y_2 = \widehat{\boldsymbol{a}}_2 \boldsymbol{x}$, … , $y_p = \widehat{\boldsymbol{a}}_p \boldsymbol{x}$ have maximum variance.

Proof: Let $y = a_1 x_1 + \cdots + a_p x_p = \boldsymbol{a}^T \boldsymbol{x}$ be a linear combination of the component of $\boldsymbol{x}$, let $y_i = a_1 x_{i,1} + \cdots + a_p x_{i,p} = \boldsymbol{a}^T \boldsymbol{x}_i$ be the value of component y for $i-th$ statistical unit, then the mean of y is $\bar{y}_i = a_1\bar{x}_{i,1} + \cdots + a_p\bar{x}_{i,p} = \boldsymbol{a}^T\overline{\boldsymbol{x}}_i$.

The variance of component y is:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\,y_i - \bar{y}_i)^2;$$

recalling that

$$\boldsymbol{a}^T(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i) = (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)^T\boldsymbol{a},$$

it is obtained

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\boldsymbol{a}^T(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)^T\boldsymbol{a}$$

$$= \boldsymbol{a}^T\frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)(\boldsymbol{x}_i - \overline{\boldsymbol{x}}_i)^T\boldsymbol{a} = \boldsymbol{a}^T\boldsymbol{S}^2\boldsymbol{a}.$$

We must find the maximum of the function $\boldsymbol{a}^T\boldsymbol{S}^2\boldsymbol{a}$ with the constraint $\boldsymbol{a}^T\boldsymbol{a} = 1$. The problem, as usual, is solved with the Lagrange multiplier technique.

In order to find the solution, the function $f(\boldsymbol{a}) = \boldsymbol{a}^T\boldsymbol{S}^2\boldsymbol{a} - \lambda(\boldsymbol{a}^T\boldsymbol{a} - 1)$ is defined, and then we calculate the vector derivative of $f(\boldsymbol{a})$ and impose that:

$$\frac{df(\boldsymbol{a})}{d\boldsymbol{a}} = 2\boldsymbol{S}^2\boldsymbol{a} - \boldsymbol{2}\lambda\boldsymbol{a} = \boldsymbol{0},$$

$$(\boldsymbol{S}^2 - \lambda\boldsymbol{I})\boldsymbol{a} = \boldsymbol{0};$$

this last equation is the eigenvalue-eigenvector problem for the variance-covariance matrix of $\boldsymbol{x}$: this demonstrates the theorem [6].

Let's return to the problem of how to reduce the complexity of the statistical system. The previous theorem entitled us to state with certainly this approximation:

$$\boldsymbol{\Sigma}^2 = \sum_{i=1}^{p} \lambda_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i{}^T = \sum_{i=1}^{m} \lambda_i \hat{\boldsymbol{e}}_i \hat{\boldsymbol{e}}_i{}^T + \boldsymbol{\Phi} =$$

$$= \left(\sqrt{\lambda_1}\hat{\boldsymbol{e}}_1 \quad \cdots \quad \sqrt{\lambda_m}\hat{\boldsymbol{e}}_m\right) \begin{pmatrix} \sqrt{\lambda_1}\hat{\boldsymbol{e}}_1{}^T \\ \vdots \\ \sqrt{\lambda_m}\hat{\boldsymbol{e}}_m{}^T \end{pmatrix} + \boldsymbol{\Phi} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Phi}.$$

The last part of the formula allows us to estimate the factor loading from the data obtained from the experiment:

$$l_{i,j} = \sqrt{\lambda_j}\hat{e}_{i,j}.$$

Recalling the multiple regression model

$$\boldsymbol{\Sigma}^2 = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Phi},$$

and equalling the corresponding elements of the diagonals of the three matrices we have that:

$$\sigma_i{}^2 = \sum_{j=1}^{m} \lambda_j \, \hat{e}_{i,j}{}^2 + \Phi_i$$

and

$$\Phi_i = \sigma_i{}^2 - \sum_{j=1}^{m} \lambda_j \, \hat{e}_{i,j}{}^2.$$

2.3 Example of Explorative Factor Analysis (EFA).

Suppose we have proposed to a class a test containing 9 different test questions:

x1 - Recognition of symmetries

x2 - Resolution of algebraic problems

x3 - Resolution of combinatorial problems

x4 - Understanding of the passage read

x5 - Synthetic transcription of a passage read

x6 - Oral description of events

x7 - Resolution of geometric problems

x8 - Invention and story telling

x9 - Resolution of riddles.

We collect the results of the 20 students who participated in the following table:

	x1	x2	x3	x4	x5	x6	x7	x8	x9
a	7	4	3	7	4	6	4	7	6
b	5	3	3	4	6	4	3	4	7
c	4	4	4	5	5	5	3	4	7
d	7	6	6	6	3	7	7	6	3
e	3	3	4	7	3	7	4	7	6
f	4	7	7	3	7	3	6	4	4
g	6	7	7	7	3	6	6	7	3
h	5	3	3	5	6	5	3	5	7
i	3	5	5	6	4	7	5	7	5
l	7	7	6	3	6	4	7	3	4
m	3	6	7	6	3	7	6	7	3
n	3	4	4	7	3	6	6	7	5
o	6	5	6	7	4	7	5	6	4
p	7	4	4	4	7	4	3	4	6
q	7	6	5	6	4	6	6	7	3
r	5	7	6	3	6	3	7	3	7
s	4	7	7	5	4	4	6	5	5
t	3	6	7	4	7	3	7	3	6
u	5	5	5	4	7	4	5	4	7
v	6	7	6	5	8	6	6	3	8

We want to identify the factors that influenced the outcome of the test, that is, which are the cause of the results achieved.

From the correlation matrix we obtain the scree plot (Fig 5) which represents the eigenvalues in descending order; the point where the curve changes concavity identifies the number of common factors to consider. In this case 3.

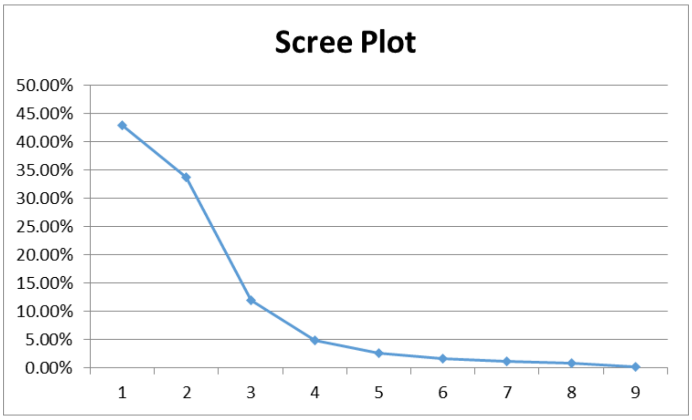


Fig 5. the point where the curve changes concavity identifies the number of common factors

At this point we consider the matrix L of the factor loadings, calculated by ordering the eigenvectors by descending order of the eigenvalues.

	1	2	3	4	5	6	7	8	9
x1	0.072328	0.051867	0.992748	0.032546	-0.03547	-0.04875	-0.01942	-0.02401	0.029704
x2	0.32423	0.905062	0.084839	0.146237	0.029714	-0.12978	0.008969	0.161102	-0.05826
x3	0.192632	0.919987	-0.18352	0.079847	0.233327	-0.08835	-0.04402	-0.08717	0.068401
x4	-0.94128	-0.03456	0.007024	0.246666	-0.00502	-0.14894	0.080466	-0.14327	-0.05195
x5	0.909586	-0.21055	0.096751	0.159581	0.177869	0.109263	0.221899	-0.0241	-0.00993
x6	-0.89036	0.025583	0.090742	0.349114	0.145029	0.212435	-0.0789	0.063938	0.00336
x7	0.129518	0.908524	-0.06695	0.14168	-0.32725	0.140744	0.057246	-0.05343	0.014871
x8	-0.96438	0.054982	-0.03281	-0.10383	-0.0202	-0.06034	0.187157	0.110941	0.060873
x9	0.507412	-0.70417	-0.17251	0.422219	-0.13105	-0.12659	-0.03211	0.047693	0.046254

We have coloured as usual the cells with factor loadings greater than 0.8, and we are ready for interpreting the result:

- the first common factor is connected with Understanding of the passage read, Synthetic transcription of a passage read, Oral description of events, Invention and story-telling.
- The second common factor is connected with Resolution of algebraic problems, Resolution of combinatorial problems, Resolution of geometric problems.
- The third common factor is connected with Recognition of symmetries.

Note that x9, associated to resolution of riddles is not represented in the reduction of complexity obtained with the three common factors.

Finally, we summarize the results in the following table:

Factor Matrix (unrotated)						
	1	2	3		Commonality	Specific
x1	0.072328	0.051867	0.992748		0.993469592	0.00653
x2	0.32423	0.905062	0.084839		0.931460916	0.068539
x3	0.192632	0.919987	-0.18352		0.917162195	0.082838
x4	-0.94128	-0.03456	0.007024		0.887247721	0.112752
x5	0.909586	-0.21055	0.096751		0.881039221	0.118961
x6	-0.89036	0.025583	0.090742		0.801632862	0.198367
x7	0.129518	0.908524	-0.06695		0.846672838	0.153327
x8	-0.96438	0.054982	-0.03281		0.934129887	0.06587
x9	0.507412	-0.70417	-0.17251		0.783088618	0.216911
	3.85783	3.038686	1.079388		7.975903848	1.024096

Let’s recall, that commonality represents the amount of variance of the random variable explained by the 3 common factors while it is indicated with specific what the factors do not explain.

2.4 The Item Response Theory (IRT)

The concept behind the IRT is that a subject's response to a test can be explained with a set of latent factors (traits) and parameters [7]. Conventionally the latent traits are called proficiencies and we will represent it with the Greek lowercase letter ϑ.

Each subject has a different amount of proficiency, so every individual subjected to the same test will have a different performance. However, the subject's response to a test cannot be explained only by the subject's abilities level possessed, but also on the parameters, i.e., the characteristics that are possessed by the test carried out. The test parameters are represented by Greek lowercase letters β, α and γ, corresponding respectively to difficulty, discrimination and guessing.

We are interested in these three parameters only for verifying the efficiency of the test as a measure instrument of the students' proficiency.

The IRT contains a series of models that differ in the number of proficiencies measured, in the number of parameters considered, and in the score of the test question carried out.

Based on the number of abilities, we distinguish between mono-dimensional or multi-dimensional model.

Based on the number of parameters taken into account, three different types of models are identified:

- the one-parameter Logistic Model (1PLM) (Rasch Model): it is assumed that only the difficulty parameter can influence the subject's response to the test questions.
- The two-parameter Logistic Model (2PLM): in addition to the difficulty, the discrimination is added.
- Three-parameter Logistic Model (3PLM): all the parameters of the test questions are considered simultaneously (difficulty, discrimination and guessing).

Based on the test question score, a distinction is made between:

- dichotomous model: all the test questions present alternative answers including the right one, and therefore the alternatives can be traced back to the dichotomous form.
- Polychotomous model: the test questions present alternative answers and do not admit a right answer.

Considering the dichotomous test questions, it is possible to identify two families of models differing in the probability distribution: Logistic Model (LM) and Normal Ogive Model (NOM). The aim is to calculate the probability of answering correctly to a test question as a function of the subject's proficiency level and test question parameters. In both families it is possible to identify different models based on the number of parameters considered. The LM is the more used in the construction of tests, because its function is simpler than those of the NOM.

The principle from which 1PLM is deduced is expressed by the relation on probability functions:

$$p(t;b) = p(kt;kb)$$

where t, b, $\frac{t}{b}$ are ability, difficulty, performance [8]. The relation expresses the invariance of the outcome with respect to the scale variation of the scores.

A solution of the mathematical problem is:

$$p(t;b) = \frac{\frac{t}{b}}{1 + \frac{t}{b}}$$

Note that

$$\frac{t}{b} = \frac{p(t;b)}{1 - p(t;b)}$$

and

$$\ln(t) - \ln(b) = \ln\left(\frac{p(t;b)}{1 - p(t;b)}\right)$$

Let

$$\theta_i = \ln\left(\frac{p_i}{1 - p_i}\right);\ p_i \in (0; 1)$$

and its inverse

$$p_i = \frac{e^{\theta_i}}{1 + e^{\theta_i}};\ \theta_i \in (-\infty; +\infty)$$

respectively, the proficiency of i-th subject and his probability of correctly answering one generic test question. It must be noted that the proficiency is estimated from the test outcome, being the relative frequency of i-th subject’s correct answer, a parameter having the properties of probability.

Let

$$\beta_j = \ln\left(\frac{1 - p_j}{p_j}\right);\ p_j \in (0; 1)$$

and its inverse

$$p_j = \frac{e^{-\beta_j}}{1 + e^{-\beta_j}};\ \beta_i \in (-\infty; +\infty)$$

respectively, the difficulty of the j-th test question and the probability of the correct answer to that test question. It must be noted that the difficulty is estimated from the test results, being the relative frequency of correct answers given to j-th test question, a parameter having the proprieties of a probability.

If the 1PM is consistent with empirical data, it is possible to estimate the proficiency of a subject without knowing the difficulty of the test questions, i.e., it is sufficient to know the profile of answers that the subject has provided to the test (sufficient statistic). It is possible to estimate the difficulty of a test question without knowing the proficiency of the subjects, i.e., it is sufficient to know the correct answers provided to the same test question by the sample (sufficient statistic). This property is called Invariance.

Linearity property of the scale adopted permits to express the performance of the i-th individual respect to the j-th test question by the location $\theta_i - \beta_j$:

$$\theta_i - \beta_j = \ln\left(\frac{p_{i;j}}{1 - p_{i;j}}\right)$$

and the probability of answering correctly as function of the location

$$p_{i;j} = \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}}$$

Another property of the model is the specific objectivity [9] i.e., with two test questions r and s and two subjects n and m we have:

$$(\theta_n - \beta_r) - (\theta_n - \beta_s) = (\theta_m - \beta_r) - (\vartheta_m - \beta_s) = \beta_s - \beta_r$$

$$(\theta_n - \beta_r) - (\theta_m - \beta_r) = (\theta_n - \beta_s) - (\vartheta_m - \beta_s) = \theta_n - \theta_m$$

that is, the relationship between the two parameters of test questions r and s is not influenced by the subjects' proficiency, the relationship between the two proficiency of subjects n and m is not influenced by the parameters of test questions.

The third property called local independence is imposed to the system: with two test questions r and s and two subjects n and m it is assumed that

$$p(u_{n;r}; u_{n;s}) = p(u_{n;r})p(u_{n;s})$$

being $u_{n;r}$ and $u_{n;s}$ pattern of responses (1 or 0) of test questions s and r given by subject n,

$$p(u_{n;r}; u_{m;r}) = p(u_{n;r})p(u_{m;r})$$

being $u_{n;r}$ and $u_{m;r}$ pattern of responses (1 or 0) of subjects n and m given to test question r.

Local independence differs from general independence. Persons with very high or very low θ values tend to respond correctly or incorrectly to many test questions, at the same time test questions with very large or very small β, will be correlated in the responses of persons. By eliminating these extreme values of proficiency and difficulty, general dependency should disappear.

A more realistic model that considers all the parameters of the IRT is the Birnbaum model [10] [11] in which the probability of answering correctly It is expressed by the equation:

$$p_{i;j} = \gamma_j + (1 - \gamma_j)e^{D\alpha_j(\theta_i - \beta_j)}$$

being $D = 1.7$ a number that makes the logistic model $L(x)$ practically equivalent to normal ogive $O(x)$ i.e., $|O(x) - L(x)|$ is minimized.

The 3PL model encapsulates the other 1PL and 2PL models, through appropriate choices of guessing and discrimination parameters.

Properties imposed, like local independence, are assumed into the models, while 2PL and 3PL no longer verify the property of specific objectivity [12].

2.5 Estimation of parameters.

In IRT the parameters are estimated starting from their probability distribution, which is supposed to be known. This hypothesis is brought to its extreme consequences which involve the elimination, from the experimental data set, of the data that do not correspond to the model adopted. Furthermore, the principle of maximum likelihood, which is adopted for determining the parameters, has local independence as its starting hypothesis.

Local independence can be violated for at least 2 reasons [13]:

- when the test is multidimensional, but proficiency is determined by a single parameter. In this case, it is necessary to divide the test into one-dimensional sub-tests, within which it is reasonable to consider valid the local independence, verifying that the non-diagonal elements of the correlation matrix of the subtests are close to zero. Then the proficiency is formalized by

  $$\theta_{i;s} = \theta_i + c_s \theta'_{i;s}$$

  being θ_i the common trait for all subsets, $\theta'_{i;s}$ the specific trait for each subset s, $c_s > 0$ a constant representing the magnitude of specific trait and θ_i and $\theta'_{i;s}$ supposed independent.

- When the response to a test question depends on the response to a previous test question. In this case the local independence is violated, as independence is a commutative relation. Algebraic formalization is obtained by adding to the location of test question k, dependent from test question j, a constant

  $$(2u_{i;j} - 1)d$$

  being $u_{i;j}$ the pattern of response to test question j given by subject i before giving the response to test question k , and being $d > 0$ a constant used to vary the dependence of test questions.

Let’s consider the simultaneous estimate of the subjects’ proficiency and parameters on a set of test questions. Assuming local independence, the likelihood function [14] which must be maximized, is formalized in the following way:

$$L(u|\theta;\beta;\alpha;\gamma) = \prod_{i=1}^{N}\prod_{j=1}^{n} {P_{i;j}}^{u_{i;j}}\, {Q_{i;j}}^{1-u_{i;j}}$$

being:

- u observed pattern of responses
- $P_{i;j}$ and $Q_{i;j}$ the probability of the correct and incorrect answer, respectively
- N the number of subjects
- n the number of test questions

Let's consider the problem in the simplest case of 1PLM. We note that:

$$\frac{\partial \ln(L)}{\partial \beta_j} = \sum_{i=1}^{N}(-u_{i;j} + \frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}})$$

$$\frac{\partial \ln(L)}{\partial \theta_i} = \sum_{j=1}^{n}(u_{i;j} - \frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}})$$

The expressions into brackets, the difference between patterns and the correspondent expected probability value, are called residuals. The problem is solved if the sum of the residuals of the first expression is null for every j, and the sum of the residuals of the second expression is null for every i. The solution is approximated by the Newton-Raphson’s iterative algorithm.

The Newton-Raphson algorithm allows to approximate the solution of an equation $f(x) = 0$ by means of the iterative formula

$$x_{k+1} \cong x_k - \frac{f(x_k)}{f'(x_k)}$$

starting from an initial value x_0. Noticing that

$$f(\theta_i) = \sum_{j=1}^{n} (u_{i;j} - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}})$$

and

$$\frac{\partial}{\partial \theta_i} f(\theta_i) = -\sum_{j=1}^{n} \frac{e^{\theta_i - \beta_j}}{\left(1 + e^{\theta_i - \beta_j}\right)^2} = -var(\theta_i)$$

we have the solution

$$\theta_i^{(k+1)} = \theta_i^{(k)} + \frac{f\left(\theta_i^{(k)}\right)}{var(\theta_i^{(k)})}$$

Similarly

$$\beta_j^{(k+1)} = \beta_j^{(k)} + \frac{f\left(\beta_j^{(k)}\right)}{var(\beta_j^{(k)})}$$

How well does the model describe the experimental data?

This problem is of great relevance and is currently much debated in teaching, since the synthetically described model is used to compare educational systems in Europe and other countries [15]. We will not deal with this topic and however we want to describe the elementary fit techniques, indicating some crucial points of the model.

The fit of model is a first check of the model's ability to adapt to real data. It is obtained by calculating the matrix:

$$\frac{(u_{i;j} - \frac{e^{\theta_i - \beta_j}}{1 + e^{\theta_i - \beta_j}})^2}{\frac{e^{\theta_i - \beta_j}}{\left(1 + e^{\theta_i - \beta_j}\right)^2}}$$

i.e., the square of residual divided the estimate of variance of expected values. It is intended that, if these values remain lower than 1.3 for all test questions and all individuals, we can believe that the model fits the experimental data.

The outfit value i.e., outlier-sensitive fit value, are the means of the corresponding standardized residuals:

$$\frac{1}{n}\sum_{j=1}^{n}\frac{(u_{i;j}-\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}})^2}{\frac{e^{\theta_i-\beta_j}}{\left(1+e^{\theta_i-\beta_j}\right)^2}}$$

$$\frac{1}{N}\sum_{i=1}^{N}\frac{(u_{i;j}-\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}})^2}{\frac{e^{\theta_i-\beta_j}}{\left(1+e^{\theta_i-\beta_j}\right)^2}}$$

The infit values i.e., information-weighted fit values are expressed by the formulas:

$$\frac{\sum_{j=1}^{n}(u_{i;j}-\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}})^2}{\sum_{j=1}^{n}\frac{e^{\theta_i-\beta_j}}{\left(1+e^{\theta_i-\beta_j}\right)^2}}$$

$$\frac{\sum_{i=1}^{N}(u_{i;j}-\frac{e^{\theta_i-\beta_j}}{1+e^{\theta_i-\beta_j}})^2}{\sum_{i=1}^{N}\frac{e^{\theta_i-\beta_j}}{\left(1+e^{\theta_i-\beta_j}\right)^2}}$$

Any infit or outfit values that exceed 1.3 are highlighted and may indicate a sub-optimal fit [16].

The other non-negligible details of the recursive procedure will be described by an example.

2.6 Example of IRT.

An example of IRT on 9 questions carried out by 9 subjects is given.

It is assumed, that all test questions belong to a subset identified through an appropriate factorial analysis, that proficiency calculated through the test results depends on the common trait of the individual and on the specific trait defined by the subset.

Before starting the analysis, any subject who answered all questions correctly or incorrectly and any test questions that all or no one answered correctly must be removed. After this removal, additional subjects/test questions may be recursively qualified for elimination.

The following table is given:

					ITEMS						
	i1	i2	i3	i4	i5	i6	i7	i8	i9	avg	ability
A	0	0	0	0	0	1	0	1	0	0.222222	-1.25276
B	1	1	1	1	1	1	0	1	1	0.888889	2.079442
C	0	1	1	0	0	1	0	1	1	0.555556	0.223144
D	1	1	1	0	1	1	1	1	1	0.888889	2.079442
E	1	0	1	1	1	0	1	1	1	0.777778	1.252763
F	0	1	1	1	0	1	1	0	1	0.666667	0.693147
G	0	1	1	1	0	1	1	1	1	0.777778	1.252763
H	0	1	1	0	0	1	0	0	1	0.444444	-0.22314
I	1	1	1	1	1	1	0	1	1	0.888889	2.079442
avg	0.444444	0.777778	0.888889	0.555556	0.444444	0.888889	0.444444	0.777778	0.888889		
difficulty	0.223144	-1.25276	-2.07944	-0.22314	0.223144	-2.07944	0.223144	-1.25276	-2.07944	-0.92195	
adj difficulty	1.145095	-0.33081	-1.15749	0.698808	1.145095	-1.15749	1.145095	-0.33081	-1.15749	0	

The data allow us to estimate, as was shown in previous paragraph, the ability and difficulty, which are the starting point of the iterative process.

The next step is to calculate the probability a subject has to answer correctly to each test question using the estimated proficiency and difficulty:

	i1	i2	i3	i4	i5	i6	i7	i8	i9		ability
A	0.08334	0.28456	0.47620	0.12438	0.08334	0.47620	0.08334	0.28456	0.47620		-1.50191
B	0.71796	0.91761	0.96220	0.79909	0.71796	0.96220	0.71796	0.91761	0.96220		2.3957
C	0.28456	0.63505	0.79909	0.38328	0.28456	0.79909	0.28456	0.63505	0.79909		0.276516
D	0.71796	0.91761	0.96220	0.79909	0.71796	0.96220	0.71796	0.91761	0.96220		2.3957
E	0.52689	0.82971	0.91761	0.63505	0.52689	0.91761	0.52689	0.82971	0.91761		1.502621
F	0.38890	0.73574	0.86420	0.49858	0.38890	0.86420	0.38890	0.73574	0.86420		0.851978
G	0.52689	0.82971	0.91761	0.63505	0.52689	0.91761	0.52689	0.82971	0.91761		1.502621
H	0.20290	0.52689	0.71796	0.28456	0.20290	0.71796	0.20290	0.52689	0.71796		-0.27937
I	0.71796	0.91761	0.96220	0.79909	0.71796	0.96220	0.71796	0.91761	0.96220		2.3957
difficulty	1.051363	-0.03875	-0.73257	0.722783	1.051363	-0.73257	1.051363	-0.03875	-0.73257	0.177963	
adj difficulty	0.8734	-0.21671	-0.91053	0.54482	0.8734	-0.91053	0.8734	-0.21671	-0.91053	0	

Then column of ability and rows of difficulty are calculated with the recursive formulas shown in previous paragraph using the matrix of residuals:

					RESIDUALS					
	i1	i2	i3	i4	i5	i6	i7	i8	i9	sum
A	-0.08334	-0.28456	-0.4762	-0.12438	-0.08334	0.5238	-0.08334	0.71544	-0.4762	-0.37211
B	0.282044	0.082394	0.037799	0.200907	0.282044	0.037799	-0.71796	0.082394	0.037799	0.325225
C	-0.28456	0.364947	0.200907	-0.38328	-0.28456	0.200907	-0.28456	0.364947	0.200907	0.095659
D	0.282044	0.082394	0.037799	-0.79909	0.282044	0.037799	0.282044	0.082394	0.037799	0.325225
E	0.473109	-0.82971	0.082394	0.364947	0.473109	-0.91761	0.473109	0.17029	0.082394	0.372037
F	-0.3889	0.264257	0.135798	0.501415	-0.3889	0.135798	0.611102	-0.73574	0.135798	0.27063
G	-0.52689	0.17029	0.082394	0.364947	-0.52689	0.082394	0.473109	0.17029	0.082394	0.372037
H	-0.2029	0.473109	0.282044	-0.28456	-0.2029	0.282044	-0.2029	-0.52689	0.282044	-0.10092
I	0.282044	0.082394	0.037799	0.200907	0.282044	0.037799	-0.71796	0.082394	0.037799	0.325225
	-0.16735	0.405515	0.420736	0.041813	-0.16735	0.420736	-0.16735	0.405515	0.420736	0.82518

and the matrix of variances calculated with the Bernoulli's formula for variance:

					VARIANCE OF EXPECTED VALUES					
	i1	i2	i3	i4	i5	i6	i7	i8	i9	var ability
A	0.076391	0.203586	0.249434	0.108911	0.076391	0.249434	0.076391	0.203586	0.249434	1.493557
B	0.202495	0.075605	0.036371	0.160544	0.202495	0.036371	0.202495	0.075605	0.036371	1.028351
C	0.203586	0.231761	0.160544	0.236376	0.203586	0.160544	0.203586	0.231761	0.160544	1.792285
D	0.202495	0.075605	0.036371	0.160544	0.202495	0.036371	0.202495	0.075605	0.036371	1.028351
E	0.249277	0.141291	0.075605	0.231761	0.249277	0.075605	0.249277	0.141291	0.075605	1.48899
F	0.237656	0.194425	0.117357	0.249998	0.237656	0.117357	0.237656	0.194425	0.117357	1.703888
G	0.249277	0.141291	0.075605	0.231761	0.249277	0.075605	0.249277	0.141291	0.075605	1.48899
H	0.161734	0.249277	0.202495	0.203586	0.161734	0.202495	0.161734	0.249277	0.202495	1.794828
I	0.202495	0.075605	0.036371	0.160544	0.202495	0.036371	0.202495	0.075605	0.036371	1.028351
var difficulty	1.785407	1.388447	0.990152	1.744023	1.785407	0.990152	1.785407	1.388447	0.990152	

The goal of the procedure is to reduce to zero the sum of square of subjects' residuals i.e., the red-written cell in residuals sheet.

The result after 10 iterations is:

	i1	i2	i3	i4	i5	i6	i7	i8	i9	ability	s.e.
A	0.029569	0.216881	0.474466	0.054976	0.029569	0.474466	0.029569	0.216881	0.474466	-1.79979	0.903224222
B	0.750499	0.964714	0.988904	0.851695	0.750499	0.988904	0.750499	0.964714	0.988904	2.794026	1.125758091
C	0.205025	0.700963	0.88428	0.329931	0.205025	0.88428	0.205025	0.700963	0.88428	0.336885	0.834417483
D	0.750499	0.964714	0.988904	0.851695	0.750499	0.988904	0.750499	0.964714	0.988904	2.794026	1.125758091
E	0.528207	0.910521	0.970737	0.681274	0.528207	0.970737	0.528207	0.910521	0.970737	1.805556	0.907996648
F	0.343235	0.826089	0.939338	0.499443	0.343235	0.939338	0.343235	0.826089	0.939338	1.043471	0.849856242
G	0.528207	0.910521	0.970737	0.681274	0.528207	0.970737	0.528207	0.910521	0.970737	1.805556	0.907996648
H	0.114441	0.540141	0.79292	0.197899	0.114441	0.79292	0.114441	0.540141	0.79292	-0.35444	0.829842985
I	0.750499	0.964714	0.988904	0.851695	0.750499	0.988904	0.750499	0.964714	0.988904	2.794026	1.125758091
diff	1.692557	-0.51534	-1.69756	1.04568	1.692557	-1.69756	1.692557	-0.51534	-1.69756	iter	10
s.e.	0.795907	0.982197	1.228509	0.817867	0.795907	1.228509	0.795907	0.982197	1.228509	prec	4.05196E-06

where standard error (s.e.) is calculated as the reciprocal of the total standard deviation question or subject.

All fit statistics are represented in the following matrix: the values that exceed 1.3 are highlighted and are considered values that don't fit the data very well.

	i1	i2	i3	i4	i5	i6	i7	i8	i9	out fit	in fit
A	0.03047	0.276945	0.902825	0.058174	0.03047	1.107634	0.03047	3.610821	0.902825	0.772293	1.135923
B	0.332446	0.036577	0.01122	0.174129	0.332446	0.01122	3.008006	0.036577	0.01122	0.439316	0.903106
C	0.257901	0.426609	0.130864	0.492384	0.257901	0.130864	0.257901	0.426609	0.130864	0.279099	0.316085
D	0.332446	0.036577	0.01122	5.742878	0.332446	0.01122	0.332446	0.036577	0.01122	0.760781	1.159603
E	0.893197	10.17584	0.030145	0.467839	0.893197	33.17262	0.893197	0.098272	0.030145	5.183828	2.102739
F	0.522614	0.210524	0.064579	1.002231	0.522614	0.064579	1.913458	4.750055	0.064579	1.012804	1.185384
G	1.119573	0.098272	0.030145	0.467839	1.119573	0.030145	0.893197	0.098272	0.030145	0.431907	0.742641
H	0.12923	0.851369	0.261161	0.246726	0.12923	0.261161	0.12923	1.174579	0.261161	0.38265	0.489156
I	0.332446	0.036577	0.01122	0.174129	0.332446	0.01122	3.008006	0.036577	0.01122	0.439316	0.903106
out fit	0.438925	1.349921	0.161487	0.980703	0.438925	3.86674	1.162879	1.140926	0.161487	iter	10
in fit	0.54615	1.175951	0.433381	0.919176	0.54615	1.93136	1.343758	1.636749	0.433381	prec	4.05E-06

Let's put our attention on critical data of the fit of model. The strategy adopted is to consider these data as missing data, i.e., data not available for the calculation of the model. This choice can lead to the elimination of other test questions and subjects as indicated above. After this deletion of data, the analysis must be repeated [17].

The ability misfits reported by the infit, and outfit values can be caused by the subject's guessing or deviation from the average behaviour of the other subjects. There are then two possibilities: the subject is eliminated, the test questions involved are considered missing data. In the latter case, it can be assumed that the subject did not understand the test questions. After these deletions, analysis must be repeated.

Reasons for test question misfit are the question is confusing, not well expressed, test question is not testing the intended ability. Again, the approach to dealing with a test question misfit is to remove the test question from the data set and repeat the analysis.

2.7 The analysis of variance.

Analysis of Variance (ANOVA) [18] is the generalization of the T test in the case when there are more than two samples (hereinafter referred to as samples) to compare, when the samples extracted from the same population are differentially influenced.

It is a fundamental statistical method in searching the causal explanation.

The analysis of variance is used to test the differences between the sample means, and this is caried out just considering the corresponding variances. The purpose of this test is to establish whether more than two sample means can derive from populations with the same a priori average.

The ANOVA must necessarily be used when the means are more than two, or when it is necessary to divide the grouping variable into variables, to eliminate the sources of variation beyond to those produced by the factor, whose effect is to be evaluated.

It will be called "sample" the group extracted from a homogeneous statistical population, "treatment" the sample subjected to a treatment that modifies its average characteristics. It will be used the index j for the sample, the index i for the statistical unit, n_j for the total number of statistical units in each sample, k for the total number of samples, n for the total number of statistical units selected, $\bar{x}_j$ for the sample mean, $\bar{x}$ for the total mean.

The quantities:

$$SS_j = \sum_{i=1}^{n_j} (x_{i;j} - \bar{x}_j)^2$$

$$SS_T = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{i;j} - \bar{x})^2$$

$$SS_W = \sum_{j=1}^{k} SS_j = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i;j} - \bar{x}_j)^2$$

$$SS_B = \sum_{j=1}^{k} n_j\,(\bar{x}_j - \bar{x})^2$$

are respectively: the sum of square of j-th sample, the sum of squares of the total sample, the sum of squares within the sample, the sum of the squares between sample means, being $\bar{x}_j$ the mean of j-th sample, $\bar{x}$ the mean of total sample.

The following equality holds:

$$SS_T = SS_W + SS_B$$

Proof:

$$SS_T = \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i;j} - \bar{x})^2 = \sum_{j=1}^{k}\sum_{i=1}^{n_j}[(x_{i;j} - \bar{x}_j) + (\bar{x}_j - \bar{x})]^2 =$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i;j} - \bar{x}_j)^2 + \sum_{j=1}^{k}\sum_{i=1}^{n_j}(\bar{x}_j - \bar{x})^2$$

$$+ 2\sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i;j} - \bar{x}_j)(\bar{x}_j - \bar{x}) =$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i;j} - \bar{x}_j)^2 + \sum_{j}^{k} n_j\,(\bar{x}_j - \bar{x})^2$$

$$+ 2\sum_{j=1}^{k}(\bar{x}_j - \bar{x})\sum_{i=1}^{n_j}\left(x_{i;j} - \bar{x}_j\right) =$$

$$= \sum_{j=1}^{k}\sum_{i=1}^{n_j}(x_{i;j} - \bar{x}_j)^2 + \sum_{j}^{k} n_j\,(\bar{x}_j - \bar{x})^2 + 2\sum_{j=1}^{k}(\bar{x}_j - \bar{x}) \cdot 0$$

$$= SS_W + SS_B$$

Being

$$df_T = n - 1$$

and

$$df_W = \sum_{j=1}^{k} n_j = n - k$$

respectively the degrees of freedom of SS_T and SS_W, it follows that

$$df_B = k - 1$$

are the degrees of freedom of SS_B.

Dividing each mean square by its degrees of freedom we obtain: the variance of total sample

$$MS_T = \frac{1}{n-1} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{i;j} - \bar{x}\right)^2$$

the variance within samples

$$MS_W = \frac{1}{n-k} \sum_{j=1}^{k} \sum_{i=1}^{n_j} \left(x_{i;j} - \bar{x}_j\right)^2$$

the variance between samples

$$MS_B = \frac{1}{k-1} \sum_{j=1}^{k} n_j \left(\bar{x}_j - \bar{x}\right)^2$$

The statistical variable

$$F = \frac{MS_B}{MS_W}$$

is a quotient between two different estimates of the population variance.

The continuous probability distribution $F(df_B; df_W)$ of F is known as Snedecor-Fisher's, depends on the degrees of freedom of the variances, and has mean value $df_W/(df_W-2)$, which, in standard conditions i.e., normal distributed population and $df_W >> 1$, does not depart from 1 being, MS_B and MS_W estimates of population variance [19].

Now it is defined a structural model for treatments

$$x'_{i;j} = x_{i,j} + a_j$$

$$x_{i;j} = \bar{x}_j + e_{i;j}$$

being $e_{i;j}$ gaussian-distributed with mean 0, a_j the parametrization of treatment effects.

It is simply to verify that:

$$\bar{x}'_j = \bar{x}_j + a_j$$

and

$$\bar{x}' = \bar{x} + \bar{a}$$

These relationships are true:

- $MS'_W = MS_W$
- $MS'_B = MS_B + \frac{1}{k-1}\sum_j n_j \, (a_j - \bar{a})^2$

Proof:

$$MS'_W = \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x'_{i;j} - \bar{x}'_j\right)^2$$

$$= \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{i;j} + a_j - \bar{x}_j - a_j\right)^2 =$$

$$= \frac{1}{n-k}\sum_{j=1}^{k}\sum_{i=1}^{n_j}\left(x_{i;j} - \bar{x}_j\right)^2 = MS_W$$

$$MS'_B = \frac{1}{k-1}\sum_{j=1}^{k} n_j \, (\bar{x}'_j - \bar{x}')^2 = \frac{1}{k-1}\sum_{j=1}^{k} n_j \, (\bar{x}_j + a_j - \bar{x} - \bar{a})^2 =$$

$$= \frac{1}{k-1}\sum_{j=1}^{k} n_j \left[\left(\bar{x}_j - \bar{x}\right)^2 + 2\left(\bar{x}_j - \bar{x}\right)\left(a_j - \bar{a}\right) + \left(a_j - \bar{a}\right)^2\right]$$

$$= MS_B + \frac{1}{k-1}\sum_{j=1}^{k} n_j\left(a_j - \bar{a}\right)^2$$

being x and a uncorrelated.

If the quantity

$$\frac{1}{k-1}\sum_{j=1}^{k} n_j\left(a_j - \bar{a}\right)^2$$

is significantly greater than 0 i.e., if F is significantly greater than 1, treatments can no longer be considered samples of the same population. However, it is necessary to remember that the outcome of a statistical test depends on whether or not:

- the hypotheses, from which the test has origin, occur
- the same result can be achieved (reproducibility), in proportion to the degree of confidence expected.

2.8 Example of ANOVA.

The following Gaussian statistical population of 84 statistical units is considered:

population					k	
17	18	15	17	16	15	19
18	16	18	17	17	16	18
17	16	17	17	15	17	16
15	16	16	19	18	15	17
17	17	16	16	17	15	19
19	17	16	16	17	17	19
16	15	18	18	17	17	17
15	17	18	16	18	18	17
18	17	17	16	17	17	15
14	13	19	17	19	17	16
16	18	18	18	17	16	14
18	16	17	16	16	15	16
p.mean	p.var					
16.7381	1.574263					

3 samples of 5 statistical units each are extracted from this population:

sample	g1	g2	g3
	17	17	15
	16	17	17
	14	13	19
	17	17	16
	16	16	15
s.mean	16	16	16.4
s.var	1.5	3	2.8
t.s.mean	16.13333		
t.s.var	2.12381		
SST	SSW	SSB	SSW+SSB
29.73333	29.2	0.533333	29.73333

The samples are tested on the random variable F, from which it is concluded that $F < F_{crit}$ i.e., samples come from the same population, with a 95% confidence level (hypothesis H0), as it is shown by the following table.

DESCRIPTION					Alpha	0.05		
Group	*Count*	*Sum*	*Mean*	*Variance*	*SS*	*Std Err*	*Lower*	*Upper*
g1	5	80	16	1.5	6	0.697615	14.48003	17.51997
g2	5	80	16	3	12	0.697615	14.48003	17.51997
g3	5	82	16.4	2.8	11.2	0.697615	14.88003	17.91997
ANOVA								
Sources	*SS*	*df*	*MS*	*F*	*P value*	*F crit*	*RMSSE*	*Omega Sq*
Between Groups	0.533333	2	0.266667	0.109589	0.897089	3.885294	0.148047	-0.13472
Within Groups	29.2	12	2.433333					
Total	29.73333	14	2.12381					

The samples are subjected to a differentiated treatment a_j:

treatment	g1+3	g2-3	g3-2
	20	14	13
	19	14	15
	17	10	17
	20	14	14
	19	13	13
s.mean	19	13	14.4
s.variance	1.5	3	2.8
t.s.mean	15.46667		
t.s.variance	9.12381		

The treatments are tested on the random variable F, from which it is concluded that $F > F_{crit}$ i.e., treatments do not come from the same population, with a 95% confidence level (hypothesis H1), as it is shown in the following table.

ANOVA: Single Factor								
DESCRIPTION					Alpha	0.05		
Group	*Count*	*Sum*	*Mean*	*Variance*	*SS*	*Std Err*	*Lower*	*Upper*
g1+3	5	95	19	1.5	6	0.697615	17.48003	20.51997
g2-3	5	65	13	3	12	0.697615	11.48003	14.51997
g3-2	5	72	14.4	2.8	11.2	0.697615	12.88003	15.91997
ANOVA								
Sources	*SS*	*df*	*MS*	*F*	*P value*	*F crit*	*RMSSE*	*Omega Sq*
Between Groups	98.53333	2	49.26667	20.24658	0.000143	3.885294	2.012291	0.71959
Within Groups	29.2	12	2.433333					
Total	127.7333	14	9.12381					

Bibliography.

[1] Spearman, C.E. (1946). Theory of the general factor, British Journal of Psychology, 36**,** 117–131.

[2] Thurstone, L.L. (1938). Primary mental abilities. Chicago: University of Chicago Press.

[3] Bryant, F. B., & Yarnold, P. R. (1995). Principal-components analysis and exploratory and confirmatory factor analysis, psycnet.apa.org.

[4] Thurstone, L. L. (1931). Multiple factor analysis. Psychological *Review, 38*(5), 406–427

[5] Guttman, L. Some necessary conditions for common-factor analysis. Psychometrika **19,** 149–161 (1954).

[6] Paul Kline, An Easy Guide to Factor Analysis, eBook Published: 25 February 2014, Taylor & Francis Group.

[7] Susan E. Embretson, Steven P. Reise, Item Response Theory, Psychology Press, 5 set 2013.

[8] Rasch, G. (1960). *Probabilistic Model for some Intelligence and Attainment Test*. Copenhagen: Danish Institute for Educational Research.

[9] Maja Planinic, William J. Boone, Ana Susac, Lana Ivanjek, Rasch analysis in physics education research: Why measurement matters, PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH 15, 020111 (2019).

[10] Birnbaum, A. (1968). *Some Latent Trait Models and Their Use to Inferring an Examinee's Ability*. In Lord, F.M. Novick, M.R. (1968).

[11] Brandon LeBeau, Aaron McVay, Validity of the three parameters test question response theory model for field test data, Iowa Testing Programs, Research Series (2017)

[12] Baker, F. B. (1987). Methodology review: test question parameter estimation under the one, two and three-parameter logistic models. *Applied Psychological Measurement*, *11* (2), 111–141.

[13] I Marais, D Andrich - Formalizing dimension and response violations of local independence in the unidimensional Rasch model – J Appl Meas, 2008.

[14] Rossi, Richard J. Mathematical Statistics: An Introduction to Likelihood Based Inference, New York: John Wiley & Sons, 2018.

[15] PISA 2015 TECHNICAL REPORT, Scaling procedures and construct validation of context questionnaire data, OECD 2017.

[16] Smith RM, Suh KK. Rasch fit statistics as a test of the invariance of test question parameter estimates. Journal of Applied Measurement. 2003; 4(2):153-163.

[17] Christine Hohensinn, Klaus D. Kubinger, On the impact of missing values on test question fit and the model validness of the Rasch model, Psychological Test and Assessment Modeling, Volume 53, 2011 (3), 380-393.

[18] Snedecor, G. W. (1934). *Iowa State College Division of Industrial Science monographs: Vol. 1. Calculation and interpretation of analysis of variance and covariance.* Collegiate Press.

[19] Patnaik, P. B. "The Non-Central χ^2- and F-Distribution and Their Applications." *Biometrika*, vol. 36, no. 1/2, 1949, pp. 202–232. *JSTOR*, www.jstor.org/stable/2332542.

Chapter 3: The didactical tools.

3.1 E-learning structure.

Can a laboratory experience be carried out through the support of Information and Communication Technology (ICT) teaching techniques? Answering this question positively means specifying in which conditions and to what extent the same experiments carried out in Lab and E-learning modality achieves the same learning objectives.

Since the IB and RB experiments pursue different objectives, carrying out them in E-learning and Laboratory environments requires a different approach, and this is particularly true for the measurement devices used in the RB experiments. It is evident that one can never learn to use a measurement apparatus by only reding an online tutorial, because the interaction between the student and the device cannot be fully replaced; to state it simply: a tutorial cannot teach anyone to swim even though it can provide useful advice.

Knowledge of the physical principles of operation and specific physical characteristics of the instruments are necessary prerequisites for RB experiments, so that the screenplay of video of an experiment should allow the students to extract not only the experimental data necessary to verify the physical law or principle, but also the qualitative and quantitative details of the measuring instruments necessary for processing the experimental data. Indeed, the determination and processing of the experimental data depend on the physical characteristics of the measuring instrument, on its sensitivity, on its operating modes.

In RB experiments the recipe becomes the screenplay, the sequence of actions and measurements, recorded and edited in the video, starting from the assembly and determination of the characteristics of the measuring instrument, up to the determination of the experimental data to be processed.

The alternation of videos and questions within the adaptive activity Moodle Lesson allows the extraction of experimental data, the control by the students of the correctness of the acquired data and is a navigation tool within the sequences of the experiment. Other tools necessary for the data processing are provided.

The IB laboratory activity incorporates the design process of an experiment, builds knowledge, makes independent thinking and creativity emerge in problem-solving, exposing the students to an authentic research process, so that there is not a screenplay. A video of an experiment is not suitable for representing the complexity of the physical system in all its possible evolutions. The unidirectional nature of the cinematographic representation does not allow a complex and rich interaction with the reality of the studied system. The IB laboratory activity is bidirectional since the students must have the possibility to modify the characteristics and constraints which the physical system is subject to, to determine relationships or verify conjectures [1][2].

However, the filmed representation of the parts of the experiment performs some special functions:

- it defines the physical system through one of its evolutions,
- it suggests the physical quantities to be observed,
- it poses the theoretical problems to be solved,
- it suggests methods and strategies to face the theoretical difficulties,

We used it as the content of the Moodle Lesson activity.

To be more specific, we assigned a part of the scaffolding function to the filmed representation of the experiment, while the questions in the Moodle Lesson activity contain essentially the simulation of the physical system on which one can verify the system behaviour and confirm the working hypotheses. This choice made the didactic support sufficiently flexible and adaptive to correspond to the needs imposed by the teaching methodology adopted. We observe that the complete integration of the HTML pages within Moodle allows to add the Java-Script code and therefore to insert the mathematical model of the physical system studied wherever it is necessary for the teaching process.

Other tools necessary for data processing and the production of the scientific report on the experiments are provided in simple sequence, recalling that the simulation of the system is the most important tool of the whole sequence in the IB method.

3.2 A real simulator of thermal interactions.

The simulator is a chessboard with pawns (Fig 1) [3]. A random number generator identifies the position of pawns and used two consecutive times permits the exchange of a pawn position.

Fig 1. Thermodynamic system simulator: the chessboard represents a set of harmonic oscillators; the pieces represent the quanta of energy.

The macroscopic state is the number of chess-board squares with 0, 1, 2, … pawns, the microscopic state is the configuration of the pawns on the chessboard.

Let's suppose a transition between macroscopic states is described by the following one:

$$(n_0 = 0, n_1 = 64, n_2 = 0, \dots) \rightarrow (n_0 = 1, n_1 = 62, n_2 = 1, \dots)$$

While the initial macroscopic state corresponds to one microscopic state, the next state is realized in 64 · 63 = 4032 ways. Since the number of squares with one pawn is greater than the number of squares without or with two pawns, with great probability a second transition could be:

$$(n_0 = 1, n_1 = 62, n_2 = 1, \dots) \rightarrow (n_0 = 2, n_1 = 60, n_2 = 2, \dots)$$

and the final state is realized in 64 · 63 · 62 · 61:4 = 3812256 ways.

Let W be the number of ways a generic macroscopic state can be realized; its value can be evaluated considering that it is given by the ratio of the number of all chessboard squares with the product of permutation of squares with the same number of pawns, that is:

$$W = \frac{64!}{n_0!\,n_1!\,n_2! \cdot \ldots}.$$

Since this number, after about sixty exchanges, becomes of the order of $\mathbf{10^{33} \div 10^{34}}$, greater than the Avogadro's number, it is useful to consider the natural logarithm of W, which in addition to being an extensive parameter, it is also an increasing function that tends to infinity very slowly. The statistical entropy S is defined:

$$S = ln \frac{64!}{n_0!\,n_1!\,n_2! \cdot \ldots}$$

so that:

$$W = e^S.$$

The exchanges of pawns are random, there are no rules that prevent their realization. It is assumed that every microscopic state is as probable as any other and, therefore, a macroscopic state will be the more likely as more microscopic states will contribute to achieving it. A further consequence is that our system will tend to spend more time in those macroscopic states realized with more microscopic states, and the return to the initial macroscopic state, realized with only one microscopic state, will be highly unlikely.

By dividing the students into groups, we will have generated a set of copies of the system (a chessboard by each group), and each copy will evolve independently of the others; at the end of the experience the groups will produce comparable macroscopic states of the system and experimental data as that represented in the following table:

n0	n1	n2	n3	n4	n5	n6	n. of states	exchanges	entropy	squares	pieces
0	64	0	0	0	0	0	1	0	0.000	64	64
15	34	15	0	0	0	0	2.51E+26	20	60.789	64	64
22	23	16	3	0	0	0	3.48E+31	40	72.627	64	64
23	23	14	3	1	0	0	8.35E+33	60	78.107	64	64
27	19	12	4	1	1	0	8.33E+33	80	78.106	64	64
28	18	12	4	0	2	0	2.83E+33	100	77.025	64	64
29	15	14	3	3	0	0	3.50E+33	120	77.237	64	64
32	16	6	6	2	2	0	1.11E+34	140	78.394	64	64
35	12	6	7	2	1	1	3.53E+33	160	77.247	64	64
36	9	8	7	2	2	0	1.16E+33	180	76.131	64	64
32	17	7	2	4	1	1	5.60E+33	200	77.709	64	64
34	15	4	6	3	1	1	3.17E+33	220	77.139	64	64
32	17	6	6	0	1	2	1.31E+33	240	76.254	64	64

The graph of the statistical entropy fully illustrates the trend of the system: for the first sixty exchanges the entropy increases, quickly saturating around the values $76 \div 79$, for subsequent exchanges, the entropy fluctuates below the maximum value [4]:

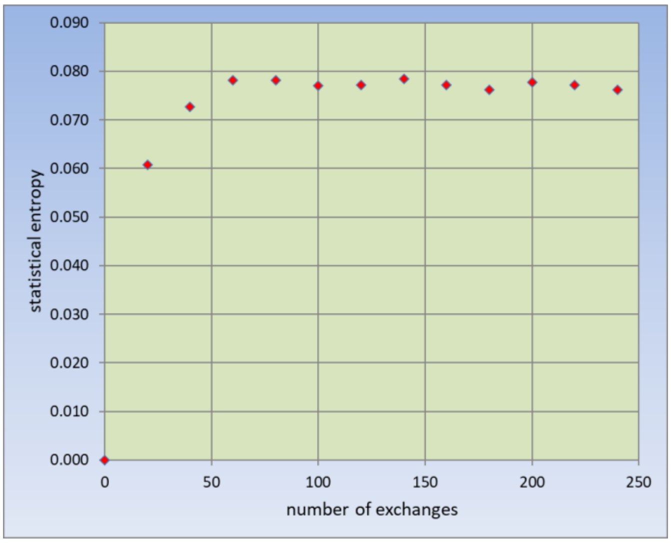


Fig 2. Temporal behaviour of the system: note the fluctuation of the statistical entropy value

The state of the system

$$n_0 = 30;\ \ n_1 = 16;\ \ n_2 = 10;\ n_3 = 5;\ n_4 = 2;\ n_5 = 1;\ n_6 = 0$$

which corresponds to a statistical entropy S≅79.25, is a relative maximum; to verify it, the occupation numbers can be changed of a unit, in all possible ways, respecting the pieces number conservation constraints, all together 24 cases, showing that the statistical entropy of the varied states is always smaller than that of the proposed macroscopic state.

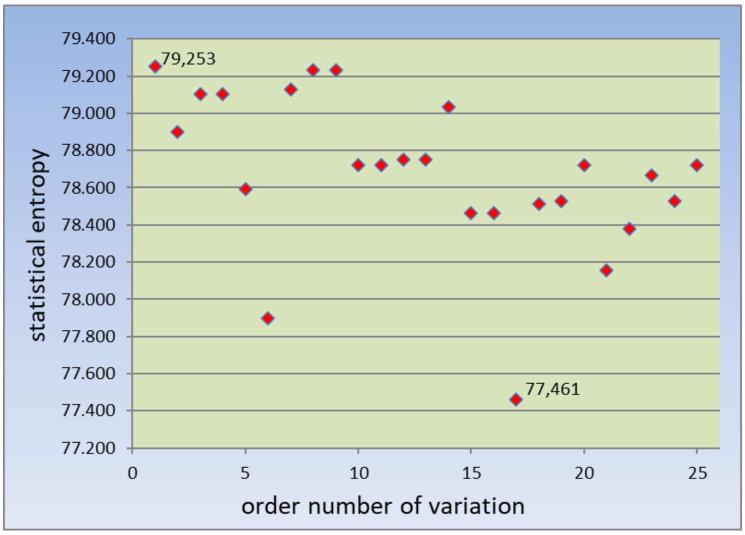


Fig 3. Entropy values around the relative maximum point.

The macroscopic state of maximum statistical entropy is the equilibrium state.

The simulation operated by the students can reach the state of equilibrium, but it is evident from the graph of the statistical entropy that the fluctuation is unavoidable: continuing the exchanges, the system continues to pass, randomly, between different macroscopic states characterized by an average statistical entropy.

By performing the exponential regression from the data on the macroscopic state of relative maximum entropy we obtain the analytical form of the distribution of pieces:

$$n_i = 33.33e^{-0,68i}$$

which corresponds to the Boltzmann distribution of Statistical Mechanics [3].

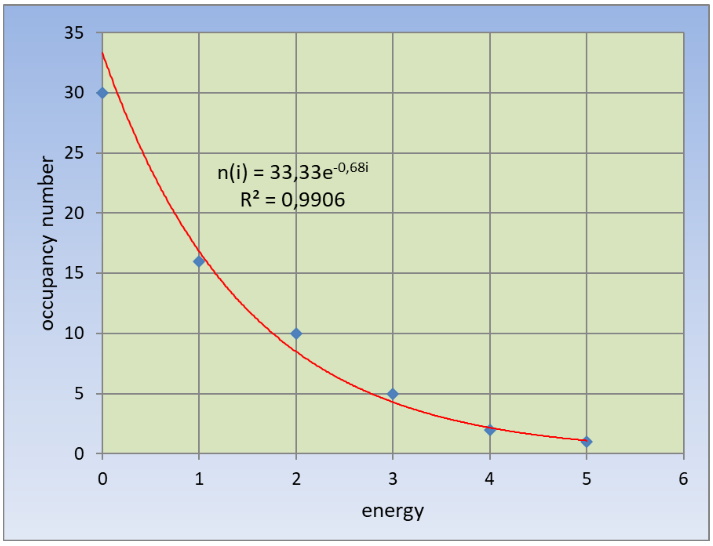


Fig 4. Boltzmann distribution: the fitting was performed on the occupation number of the various energies i.e., the number of squares occupied by the same number of pieces.

From the analytical distribution follows the probability that a square has n pieces:

$$p_i = \frac{n_i}{n} = 0.52e^{-0,68i}$$

The value β = 0.68, extracted from the previous report, is the parameter of the distribution depending on the average number of pieces per pawn; this aspect will become clear, through the further activities in which the students will participate.

3.3 Generalization of the real Simulator of Thermodynamic System.

As we described in the previous chapters, the construction of the model proceeds through the progressive generalization and abstraction of what the students have already learned.

A simple algorithm can carry out an interactive simulation of the system. The software

- permits to choose the number of pawns,
- permits to choose the number of squares of chessboard
- communicates the instantaneous macroscopic state,
- communicates the relative maximum of entropy reached,
- provides students with other parameters that the teacher considers necessary.

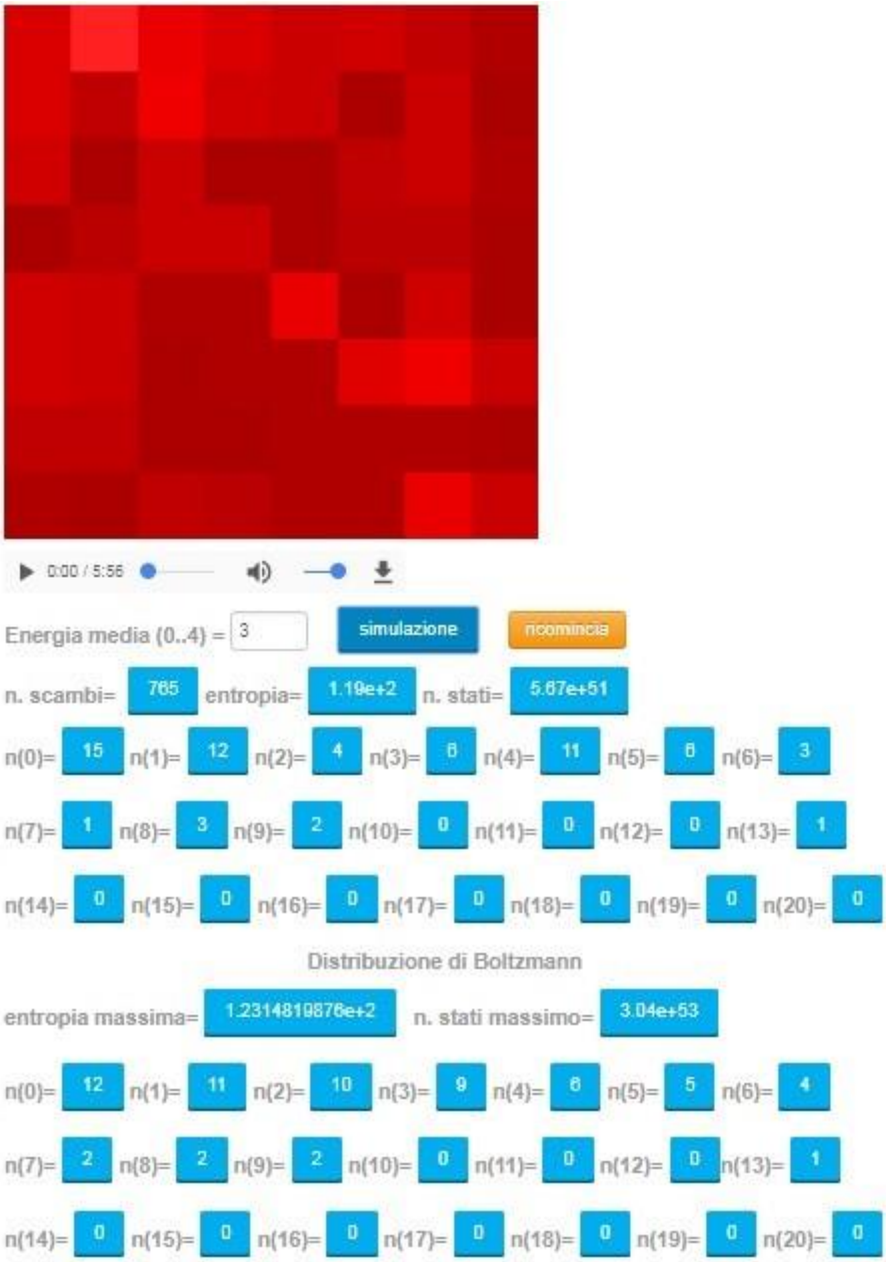


Fig 5. Simulator of the evolution of a thermodynamic system: the algorithm shows the current state and the maximum entropy state reached.

The use of the algorithm allows to develop other didactical activities:

- to determine the function β = β (j), where j is the average number of pawns per square.
- To determine the function S = S (j), where S is the maximum statistical entropy and j is the average number of pawns per square.
- To determine the relationship between the number of added pawns and the statistical entropy.

The first activity builds the concept of temperature starting from its statistical meaning. Indeed, the data extracted from the simulator $\beta\ (1)\ =$

$0.684, \beta\,(2) = 0.371, \beta\,(3) = 0.248, \beta\,(4) = 0.182, \ldots$, lead to the equation:

$$\frac{1}{\beta} \cong kJ$$

being k≅1.4.

The second activity leads the student to determine the function $S(j)$:

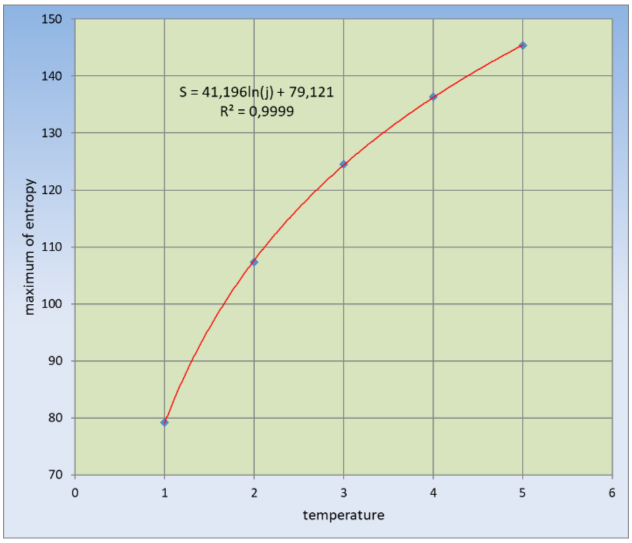


Fig 6. Behaviour of entropy: the fitting was performed on the maximum of statistical entropy at the values of 1, 2, 3, 4, 5 energies i.e., the mean number of pieces.

The third activity leads the student to determine the Boltzmann relation between the statistical entropy and the physical entropy. Indeed, the passage between two successive values of the maximum statistical entropy is interpreted as having supplied the chessboard, in addition to the previous ones, with a number of pieces, so that we can write the relation $dN = ndj$, where N is the total number of pawns and n the number of squares of chessboard; by performing a logarithmic regression of the function $S = S(j)$ and differentiating the student gets the relation:

$$dS \cong \frac{dN}{kJ}$$

from which it is deduced that, by calling

$$S = klnW$$

the thermodynamic entropy, we reach the equation:

$$dS \cong \frac{dN}{J}$$

which recalls the well-known relationship of Classical Thermodynamics [5]

$$dS = \frac{\delta Q}{T}.$$

3.4 The system reproduces the results of Classical Statistical Mechanics.

The conceptual schemes assimilated with the use of simulations become prerequisites for understanding the main laws and relations of Classical Statistical Mechanics.

From the point of view of the results produced by the simulation, it is irrelevant to consider an oscillator as an element of the Gibbs statistical set or rather the set of oscillators as the statistical population on which to perform probabilistic measurements. However, from the didactic point of view the two things are not equivalent and the second point of view clearly needs less mathematical prerequisites than the first. Adopting the second approach, it is easier to understand the imposition of the constraints of conservation of the number of oscillators and of the total energy, constraints automatically verified by the simulator. For this purpose and for completeness we have rather suggested to consider each chessboard as an element of a statistical system, so that the practical experience of comparing chessboards induces understanding of the features that remain common, as maximum entropy or total energy, and those that change, as temporal evolutions or realized microscopic states corresponding to the same macroscopic states.

In conclusion, the analogy realizes on the chessboard a Markovian type of process that, starting from a not equilibrium state, makes the simulator converge to the results of Bose-Einstein Statistics and therefore, approximately to the Maxwell-Boltzmann statistics, if the mean number of pieces per square j is greater or equal 1, as students have been able to prove by the experiment. We resume this fact saying that the given system solves the Boltzmann problem:

$$\begin{cases} \sum_i \delta n_i = 0 \\ \sum_i \varepsilon_i \delta n_i = 0 \\ \delta \ln(W) = \sum_i \delta \ln (n_i!) = 0 \end{cases}$$

being n_i the number of harmonic oscillators with energy ε_i and lnW the statistical entropy.

The consistence of the system is proved by its ergodicity; indeed, since the probability p of the macroscopic state is proportional to the number of microscopic states forming the same macroscopic state, we have:

$$dp = \frac{e^S}{e^{S_{max}} - 1} dS$$

being S and S_{max} respectively the entropy and the maximum entropy of the system; calculating the average of the entropy $\mu(S)$, we get

$$\mu(S) = \int_0^{S_{max}} \frac{Se^s}{e^{S_{max}} - 1} dS = \frac{S_{max} e^{S_{max}} - e^{S_{max}} + 1}{e^{S_{max}} - 1} \cong S_{max} - 1$$

We estimate the values of $\mu(S)$ and S_{max} from the experiment; data from one of the 12 groups that performed the experiment (the table shown above), we get $\mu(S) = 77{,}3 \pm 0{,}2$ and $S_{max} - 1 = 77{,}4$ and similar results for the other groups (the error is calculated, as usual, using estimate of standard deviation on a sample of 10 values). The thesis is proved recalling that the estimation of $\mu(S)$ is a temporal mean.

3.5 The real Brownian Motion simulator.

We consider now a second example of analogous system simulating a Brownian motion [6]. We will see that the system allows the student to determine, in a practical way, all the characteristics of Brownian motion, including the solution of the diffusion equation in presence or absence of external forces.

An analogy is established between the central point of a hexagon and the instantaneous position of a Brownian Particle on a plan. By randomly adding hexagons, as in the tessellation of a plane, starting from a first black hexagon, a random walk is realized, that simulates the spreading of a Brownian particle on a plan. Adding a yellow hexagon after f steps, a red one after $4f$ steps and a green one after $9f$, the data from the random walk are collected, recording the positions $(x; y)$ of the three coloured hexagons.

Fig 7. The Brownian motion simulator: each hexagon represents a position occupied by the particle.

A statistical sample of these data allows students to determine, with a precision depending on the sample size:

- the law $\mu_t(r) = k\sqrt{t}$ of the average distance r from the black hexagon at time t,
- the law $\sigma_t(r) = h\sqrt{t}$ for standard deviation of the average distance r from the black hexagon at time t.

From a sample of 150 random walks, students get $m(r_{t=1}) = (12.4 \pm 0.5)cm$, $m(r_{t=4}) = (22.2 \pm 0.5)\ cm$ for the sample mean of r and $s(r_{t=1}) = 6.3\ cm$, $s(r_{t=4}) = 13.4\ cm$ for the sample standard deviation. The error is determined by the sampling standard deviation divided the square root of the number of statistical units.

Assuming that the frequency of the steps is constant, it is possible to establish the relationship between the sampling mean of r and time t, the sampling standard deviation of r and time t.

3.6 Generalization of the real system through an algorithm.

The practical activity develops the basic knowledge of the mathematical model of the system. The construction of the mathematical model of the Brownian motion proceeds through the progressive generalization and abstraction of what the students have already learned.

A simple algorithm can carry out an interactive simulation of the system. The software permits

- to choose frequency and number of positions detected,
- to add external interactions
- to provide the students with the measure of positions.

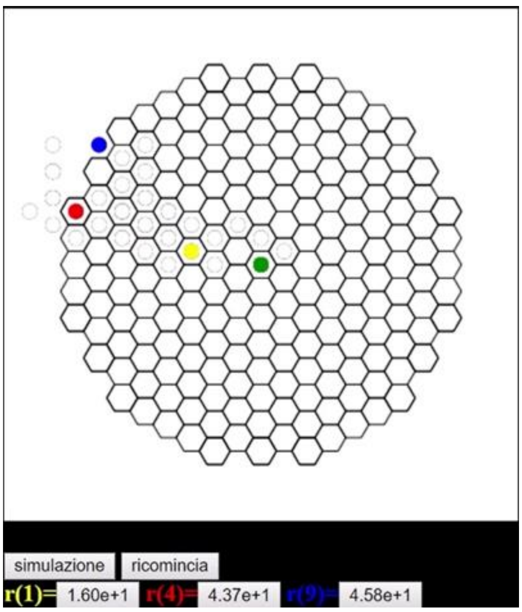


Fig 8. Algorithm that simulates the Brownian motion: the motion of the particle is reproduced, and the positions are given after n, 4n, 9n displacements from the initial position

This tool solicits students to other and more complex guided research, such as those leading to the concept of probability distribution of the random variable r, or those which allow them to determine the external forces acting on the Brownian particle by studying its drift.

Using the algorithm, students collect data of $r_{t=1}$, $r_{t=4}$, $r_{t=9}$ from a number of random walks; with the data collected the students calculate the sampling mean and the sampling standard deviation by obtaining: $m(r_{t=1}) = 12.21$, $m(r_{t=4}) = 23.21$, $m(r_{t=9}) = 36.31$, $s(r_{t=1}) = 5.87$, $s(r_{t=4}) = 12.48$, $s(r_{t=9}) = 19.76$ suggesting a property of two dimension Brownian motion, that we will show theoretically as follow:

$$\mu_t(r) \cong 2\sigma_t(r).$$

The distribution of the random variable r_t has been determined.

- the students calculate the absolute frequencies of the 7 classes by which the data are divided, from the minimum value to the maximum value of r_t,
- they represent the data in the form of a histogram

The relative frequencies of the 7 classes are positive numbers, their sum is 1, the classes are disjoint, the relative frequency of the union of two classes is the sum of the relative frequency of each, therefore a Probability Space has been established for which the relative frequencies of the classes represent the probability that an experiment yields a value of r_t falling into those classes

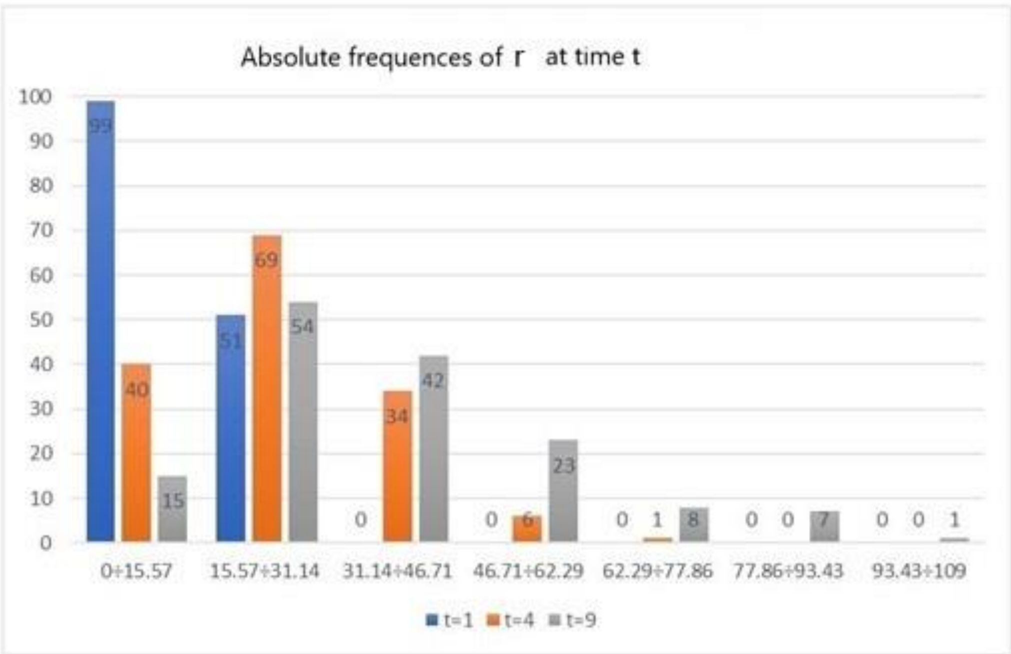


Fig 9. Probability distribution for different time classes and times.

Finally, students use the same HTML-JavaScript algorithm in which the 6 verses defined by the hexagon are no longer equiprobable, thus simulating a constant force field acting on the Brownian particle. The algorithm provides the coordinates of the arrival points after a fixed number of steps. The random variables are the coordinates (x_f; y_f), and the task consists in establishing if there is a more probable direction of Brownian particle, and what it is.

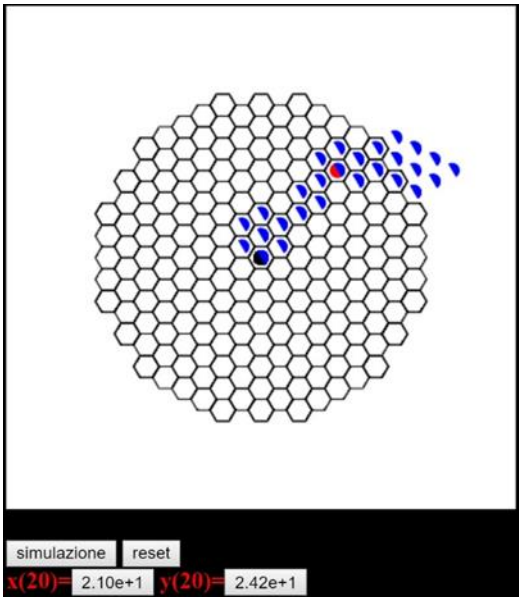


Fig 10. The algorithm that simulates the Brownian motion with drift: the motion of the particle is reproduced, the positions x(t) and y(t) are given after at a fixed time.

With the help of the calculation sheet and the sample extracted from the simulation, students realize the graph of the dispersion of r_{t_f}; the points show an inhomogeneity, they are not positioned around the centre of the coordinates. It is suggested to calculate the average of coordinates obtaining $m(x_{t=4}) = 16.38$, $m(y_{t=4}) = 9.42$. These values are well away from the origin of the axes since we expect a Brownian Motion with drift.

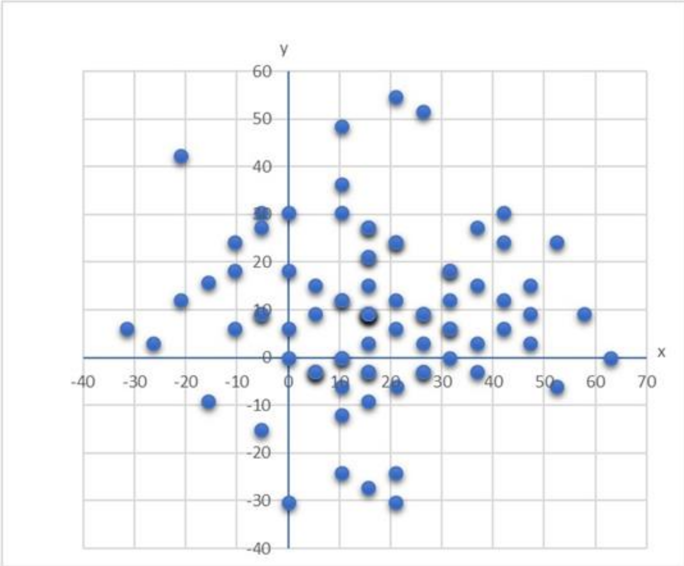


Fig 11. Graphical representation of the final positions of a sample of Brownian particles subjected to an external force.

From the equation $\theta = tg^{-1}(9.42/16.38)$ follows $\vartheta = 29.94°$ in agreement with the choice made by teacher to double the probability of a particular side of the hexagon, corresponding to a direction that forms an angle of $30°$ with the x axis.

3.7 The system reproduces the results of Brownian Motion theory.

The system, together with the application of the elementary Sampling Theory, solves the equation for the probability density of the stochastic variable r at the time t:

$$\begin{cases} \frac{\partial f(r;t)}{\partial t} = -\frac{1}{4}\frac{r}{t}\frac{\partial f(r;t)}{\partial r} + \frac{D}{2}\frac{\partial^2 f(r;t)}{\partial r^2} \\ f(r;0) = \delta(r) \end{cases}$$

being $f(r;t)$ the above-mentioned probability density.

We use a heuristic method to prove it. We start by noting that x and y at the time t are Gaussian independent stochastic variables with mean 0 and variance usually written ${\sigma^2}_t = 2Dt$ as it will be shown below; the stochastic variable $\frac{r^2}{{\sigma_t}^2}$ is a χ^2 with 2 degrees of freedom, and for the probability p at constant time t it can be written

$$\Delta p(a) = p(a \leq r \leq a + \Delta a) = p\left(\frac{a^2}{{\sigma_t}^2} \leq \frac{r^2}{{\sigma_t}^2} \leq \frac{a^2}{{\sigma_t}^2} + \Delta\left(\frac{a^2}{{\sigma_t}^2}\right)\right)$$

$$= \int_{\frac{a^2}{{\sigma_t}^2}}^{\frac{a^2}{{\sigma_t}^2}+\Delta\left(\frac{a^2}{{\sigma_t}^2}\right)} g\left(\frac{r^2}{{\sigma_t}^2}\right) d\left(\frac{r^2}{{\sigma_t}^2}\right)$$

being $g\left(\frac{r^2}{{\sigma_t}^2}\right) = \frac{1}{2} e^{-\frac{r^2}{2{\sigma_t}^2}}$ the probability density of the χ^2 variable. It follows that

$$\frac{\Delta p(a)}{\Delta a} = \frac{1}{\Delta a}\int_{\frac{a^2}{{\sigma_t}^2}}^{\frac{a^2}{{\sigma_t}^2}+\Delta\left(\frac{a^2}{{\sigma_t}^2}\right)} g\left(\frac{r^2}{{\sigma_t}^2}\right) d\left(\frac{r^2}{{\sigma_t}^2}\right) =$$

$$= \frac{2a}{{\sigma_t}^2}\frac{1}{\frac{2a}{{\sigma_t}^2}\Delta a}\int_{\frac{a^2}{{\sigma_t}^2}}^{\frac{a^2}{{\sigma_t}^2}+\Delta\left(\frac{a^2}{{\sigma_t}^2}\right)} g\left(\frac{r^2}{{\sigma_t}^2}\right) d\left(\frac{r^2}{{\sigma_t}^2}\right)$$

performing the limit as Δa approaches 0, noticing that $\Delta \frac{a^2}{{\sigma_t}^2}$ approaches 0, and $\frac{2a}{{\sigma_t}^2} \Delta a$ approaches $\Delta\left(\frac{a^2}{{\sigma_t}^2}\right)$, we obtain

$$f(a;t) = \frac{dp(a)}{da} = \frac{2a}{{\sigma_t}^2} g\left(\frac{a^2}{{\sigma_t}^2}\right) = \frac{a}{{\sigma_t}^2} e^{-\frac{a^2}{2{\sigma_t}^2}}$$

and

$$f(r;t) = \frac{r}{{\sigma_t}^2} e^{-\frac{r^2}{2{\sigma_t}^2}}.$$

This is a solution of the partial differential equation given at the beginning of this calculation, as can be proved by replacing $f(r;t)$ in the equation and considering the dependence from t of the variance.

It must be noticed that in three dimensions, with the same method, one gets the solution

$$f(r;t) = \sqrt{\frac{2}{\pi}} \frac{r^2}{{\sigma_t}^3} e^{-\frac{r^2}{2{\sigma_t}^2}}$$

and the equation associated to that solution is

$$\begin{cases} \dfrac{\partial f(r;t)}{\partial t} = D \dfrac{\partial^2 f(r;t)}{\partial r^2} \\ f(r;0) = \delta(r) \end{cases}$$

as it is expected.

This shows the theoretical consistency of the results obtained by the students; we verify the relation according to which the average of the distances from the start point is about twice its standard deviation. With the probability density of the variable r, calculating the mean and the variance of r_t:

$$\mu(r) = \int_0^{+\infty} \frac{r^2}{{\sigma_t}^2} e^{-\frac{r^2}{2{\sigma_t}^2}} dr = \frac{\sigma_t}{2}\sqrt{2\pi}$$

and

$$\mu(r^2) = \int_0^{+\infty} \frac{r^3}{{\sigma_t}^2} e^{-\frac{r^2}{2{\sigma_t}^2}} dr = 2{\sigma_t}^2$$

as it is expected, being x and y independent with same variance.

Therefore

$$\mu(r) = \frac{\sigma_t}{2}\sqrt{2\pi} \cong 1.25\sigma_t$$

$$\sqrt{var(r)} = \sqrt{\mu(r^2) - \mu^2(r)} = \sqrt{2{\sigma_t}^2 - \frac{\pi}{2}{\sigma_t}^2} \cong 0.66\sigma_t$$

Recalling the dependence of σ_t from the root of t, it results that the expressions obtained are in complete agreement with what was established by the students through the Analogous System. We notice that the same evaluation in the case of three-dimensional system isn't longer valid, being:

$$\mu(r) = \sqrt{\frac{8}{\pi}}\sigma_t \cong 1.60\sigma_t$$

$$\sqrt{var(r)} = \sqrt{\mu(r^2) - \mu^2(r)} = \sqrt{3{\sigma_t}^2 - \frac{8}{\pi}{\sigma_t}^2} \cong 0.67\sigma_t$$

To demonstrate that ${\sigma^2}_t = 2Dt$, we consider the random variable $x = (2k - n)l$, being k the number of steps on the right, n-k the number of steps on the left, n the total number of steps, l the length of a step. The random variable k verifies the Bernoulli's probability distribution and:

$$\mu(x) = 0$$

$$var(x) = {\sigma^2}_t = nl^2$$

But $n = \nu t$, being ν the frequency of the random walk, so that ${\sigma^2}_t = \nu t l^2 = 2Dt$.

3.8 Computer tools.

The construction of the mathematical model of a physical system, starting from real experiments, requires experiment repetitions with different constraints and initial conditions. This operation is not always easy if performed on the real analogous system, while it becomes easy if there is a computer simulation of the system. In practice, the construction of the model requires the generalization to other situations of an acquired scheme, up to identify all its essential characteristics, but this can be done only at the price of many repeated tests and experiences. In these terms the speed of the numerical simulation is fundamental.

Students understands what is happening on the screen of their electronic device, because they assimilate it to a previous scheme, but now, paradoxically, they have the possibility to widen the perception of the physical system, as they can act on the characteristics of the situation, and by changing it, they can confirm the expected result or discover a new possibility. The freedom acquired in the search for new relationships between known objects lies in the possession of previous knowledge.

The software, which generalizes the two analogous experiences used in the experimentation, is designed primarily to reproduce a virtual environment in which students can operate in the same way as in a real environment. The software must allow the teacher to set the characteristics of the physical system that are necessary for the student to proceed along the cognitive path. Therefore, the entire project is realized if the software has two levels:

- the teacher level (TL), in which the teacher constructs the physical system to be studied.
- The student level (ST) in which the student uses the system produced by the teacher to inquire into its physical properties.

Other element that determines the design of the software is portability, that means that software created by the teacher in the TL must be executable on any operative system. This goal is achieved by adopting the solution of

HTML5 mark-up language for the front-end and procedural languages JavaScript and PHP for the back end. The PHP language is a server-side programming language whose script, if executed, produces dynamic and interactive pages whose content may vary depending on the user's choices. The pages produced by PHP are HTML files that contain JavaScript, which is a client-side programming language.

In the simplest version, the TL could be realized with a main HTML file containing a form section through which the teacher sends to a main PHP file on an Apache HTTP Server the data about the characteristics of the HTML file that must be produced and sent to the SL.

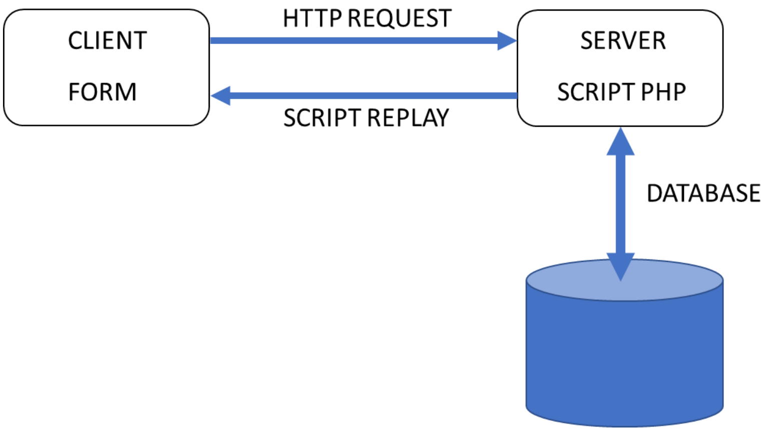


Fig 12. Project structure: the form sent from the client side to the server generates an HTML-JavaScript file to the end user.

The thermodynamic system simulator has characteristics suitable to be generalized. For each oscillator we can associate three variables of type integer and one variable of type char. The variables represent: identity, energy value, position in space and physical state (solid liquid gas).

The thermal interactions are simulated by the random exchange of energy quanta, as we have shown previously. Adiabatic interactions are simulated from the relationship between variation of volume and variation of temperature:

$$\frac{j_f}{j_i} = \left(\frac{N_i}{N_f}\right)^{\alpha}$$

being N the volume, α a positive constant (α=0 correspond to free expansion of gas).

3.9 The details of algorithms

The development of the algorithms was favoured by the choice of HTML and JavaScript. These Object-Oriented Programming languages (OOP) are sufficiently expressive and contain many object classes; it follows that the number of instructions used in the algorithms is small and the files produced are legible and easily adaptable to specific programming needs. We recall that a class of objects is an Abstract Description of a data Type (ADT) and that an object is an instance of the class.

JavaScript manages attributes and methods of these objects making the realization of the algorithm simple and immediate.

For example, in HTML the Canvas class is instantiated in the *myCanvas* object via the tag:

<canvas id = "*myCanvas*" width = "*myWidth*" height = "*myHeight*"> </ Canvas>

which defines the object; then the object is managed by JavaScript, assigning it to a variable via the statement:

var c = document.getElementById ("*myCanvas*");

and loading its methods via the statement:

var ctx = c.getContext ("2d") ;.

This simple sequence allows you to manage all the graphics needed to simulate the physical systems used; for example

```
function point (x, y, r, color)
{
  ctx.beginPath ();
  ctx.arc (x, y, r, 0.2 * Math.PI);
  ctx.fillStyle = color;
  ctx.fill ();
```

}

produces point with center (x; -y), radius r, color color, using the attribute fillStyle and the methods beginPath(), arc(), fill().

Not all objects must be instantiated, some of them are created when an HTML file is started, this is the case of the objects Math and Window.

Attributes and methods of Math are numerical constant and mathematical functions. The mathematical model of analogous system bases their temporal evolution on the random exchange of pawns or on random displacements of hexagons, in both cases the Math.floor(k*Math.random())+1 statement is used. For example, if k = 6 the function randomly generates integers between 1 and 6, including 1 and 6.

Window object is created every time a browser opens a window. This object has many attributes and methods and we used them to manage the graphic representation of the temporal evolution of AS. In particular:

- setInterval(*function, milliseconds*) calls a function or evaluates an expression at time intervals expressed in millisecond
- setTimeout(*function, milliseconds*) calls a function or evaluate an expression after a time interval expressed in milliseconds
- clearInterval(*myVar*) clears timer set with setInterval()
- clearTimeout(*myVar*) clears timer set with setTimeout()

Animation of analogous system is achieved through the simple structure:

var *myVar*=setInterval(*myFunction*,*myMilliseconds*);

function *myFunction()*

{

. . .

If . . .

{

clearInterval(*myVar*);

}

```
}
```

The procedure is performed at *myMilliseconds* time intervals until the condition expressed in the if statement is satisfied, in this case the clearInterval(*myVar)* method stops the execution of the procedure.

As we have mentioned, the nucleus of the two analogous systems is based on the random number generator Math.random(), we now illustrate how the evolution of the systems is generated.

For the simulator of thermodynamic systems two random numbers are generated which identify the emitter and the receptor of the quantum of energy; if the emitter is not in the fundamental state, it decreases its energy by one, while the receptor increases its energy by one; the code of exchange(n):

```
function exchange(n)
{
  k=Math.floor(n*Math.random());
  if(energy[k]>0)
  {
     h=Math.floor(n*Math.random());
     energy[k]=energy[k]-1;
     energy[h]=energy[h]+1;
  }
}
```

where n is the number of oscillators involved in the system.

This simple algorithm is the engine of all the thermal transformations that the system can simulate.

As regards the Brownian motion simulator, the mathematical model adopted is that of the random walk, therefore it consists of a sequence of isometries starting from an initial hexagon. Once the orientation of the initial hex is established, the 6 transformations are determined, in our case the function move(l,b):

```
function move(l,b)
{
  if (b==1)
  {
      y=y+l*Math.sqrt(3);
  }
  if (b==2)
  {
      x=x+l+l/2;
      y=y+0.5*l*Math.sqrt(3);
   }
   if (b==3)
   {
       x=x+l+l/2;
       y=y-0.5*l*Math.sqrt(3);
     }
     if (b==4)
     {
       y=y-l*Math.sqrt(3);
     }
     if (b==5)
     {
       x=x-l-l/2;
       y=y-0.5*l*Math.sqrt(3);
```

```
    }
    if (b==6)
    {
        x=x-l-l/2;
        y=y+0.5*l*Math.sqrt(3);
    }
}
```

where l is the length of side of hex and b is a random number from 1 to 6.

Let’s note a random walk with drift can be easily simulates by suitably modifying the random number generator.

Bibliography

[1] Ludwig Huber. Reflection. In Inquiry-Based Learning-Undergraduate Research. The German Multidisciplinary Experience. Harald A. Mieg Editor 2019. Springer Open.

[2] Boyer Commission on Educating Undergraduates in the Research University (1998). Reinventing undergraduate education: A blueprint for America's research universities. Stony Brook: State University of New York at Stony Brook

[3] I. Borsini. An analogous experience to make the knowledge of Statistical

Mechanics meaningful. GIORNALE DI FISICA VOL. LX, N.1, gennaio, marzo 2019.

[4] Don S. Lemons, A student's Guide to Entropy, Cambridge University Press, (2013).

[5] Ludwig Boltzmann, Lectures on Gas Theory, Dover Books on Physics.

[6] I. Borsini. Understanding Brownian motion through the Analogous Experience. GIORNALE DI FISICA VOL. LX, N. 3 Luglio-Settembre 2019.

Chapter 4: The Implementation and the results of the didactic experimentation.

4.1 The placement test (PT).

The tests measuring multiple attitudes aim to detect one or more specific skills which can be considered the necessary requisites for carrying out work tasks. These tools are designed for specific purposes and validated on the samples by statistical techniques based on IRT and CFA.

The test we used is inspired by the Differential Aptitude Test (DAT) [1], which measures seven different aptitudes. It can be applied independently to specific professional profiles that must be selected. The seven attitudes are: verbal reasoning, numerical reasoning, abstract reasoning, speed and precision, mechanical reasoning, spatial relationships and use of language.

The attitudes that can be considered relevant and useful for the final purpose of the test were identified. The choice was made based on the scientific literature, and on the analyses of psycho-aptitude tests already standardized and widely used for the selection or orientation in education and vocational [2] [3] [4].

Our Multiple Psycho-attitudinal Battery (MPB) is made up of 28 dichotomous multiple choices items with 5 alternatives. The test measures three dimensions that are Verbal Reasoning (VR), Abstract Reasoning (AR) and Logical-Arithmetic Reasoning (LAR).

The VR attitude is measured by three sub-tests, which are: synonyms (items 1-4), contrary (items 5-8) and comprehension of sentences (item 9-12).

The AR attitude is measured by two sub-tests, which are: abstract series (items 13-16) and abstract analogies (items 17-20).

The LAR attitude is measured by two sub-tests, which are: numerical series (items 21-24) and arithmetic problems (items 25-28).

The tests measure individual abilities on each of the three dimensions, in particular:

1) the VR test provides a measure of the ability to understand and evaluate the information contained in verbal statements of events or facts, to grasp the relationships between words,
2) the AR test measures ability to understand and use for the deductions the relationships between series of abstract figures and propositions,
3) the LAR measures the ability to master the concepts of numbers and understand the relationships between the numbers (mathematical reasoning, not the ability of calculation).

The test was carried out collectively, with a duration of 30 minutes.
The students' answers were dichotomized in the following way:

- for VR items: correct answer=2.78, wrong answer or missing values=0,
- for LAR and AR: correct answer=4.17, wrong answer or missing values=0.

The maximum score of each dimension is 33.36.

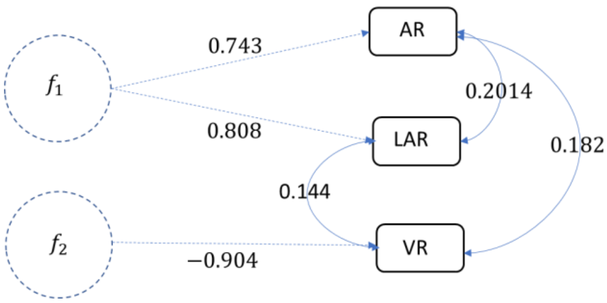
The test was carried out by the students of the first year of Chemical and Pharmaceutical Technologies degree course at the University of Camerino. A total of 65 students took the test, based on which 12 equivalent work groups, each of 5 students, have been selected. Subsequently 6 pairs of A-B work groups were formed, as it will be described.
In addition, an analysis was made on all test results aimed at investigating the level of reliability in terms of internal consistency, factor structure and psychometric characteristics of test and items. All the analyses were performed using Excel Real Statistics Resource Pack [5].

4.2 The validation of the Placement Test.

The first analysis was aimed at confirming the three-dimension structure (VR, AR, and LAR) of the whole test, the three sub-dimensions of the VR test and the two sub-dimensions of AR and LAR tests.

The CFA was performed on the sample’s data collected for the whole test and for each dimension. The results of the survey, collected in the graph and in the communality table, shows that:

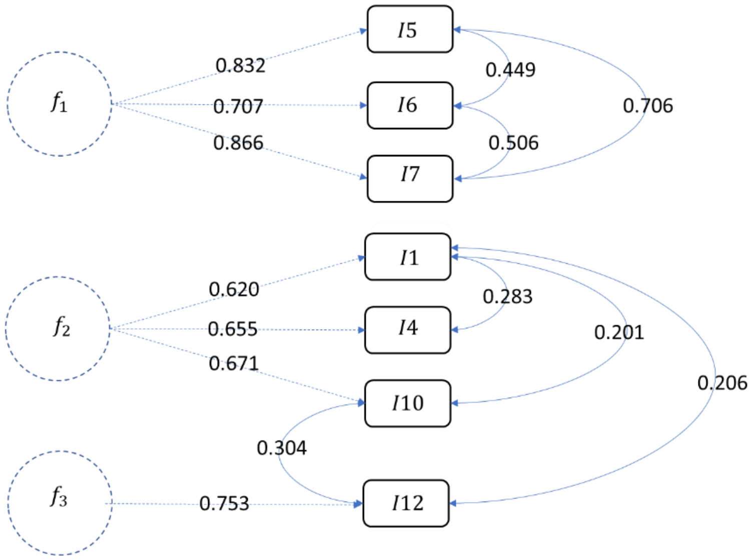
- the AR and LAR dimensions are linked to a single factor, while the VR dimension forms the remaining factor. It means that AR and LAR dimensions are strictly connected (Fig. 1).



Factor Matrix (unrotated)					
	f1	f2		Commun	Specific
verbal	0.101359	-0.90374		0.827014	0.172986
abstract	0.743274	0.452085		0.756837	0.243163
math	0.807921	-0.30253		0.744261	0.255739
	1.215466	1.112647	tot	2.328112	0.671888

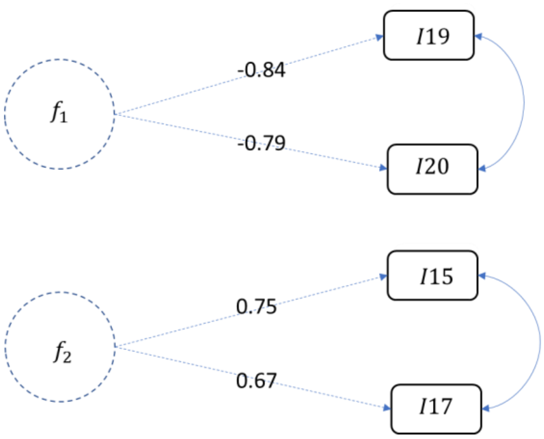
Fig 1. Factor analysis shows that the AR and LAR dimensions are linked to a single factor.

Then, CFA analysis was performed within each dimension, to confirm the existence of the three sub-dimensions for VR, the two sub-dimensions for AR and LAR.



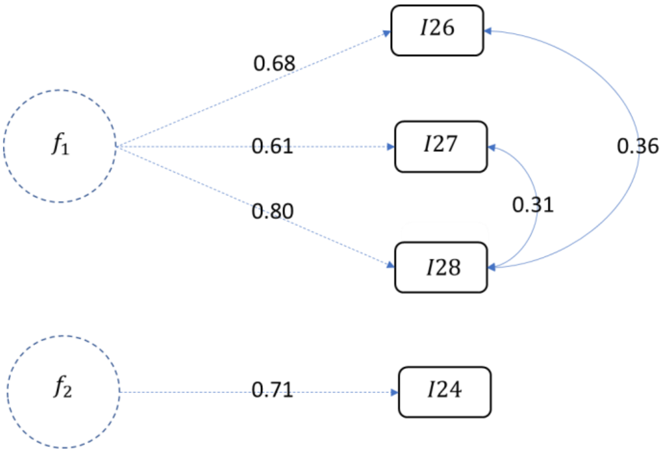
Factor Matrix (unrotated)						
	f1	f2	f3		Commun	Specific
D. 1 /2,78	0.156281	0.620775	-0.49929		0.659077	0.340923
D. 2 /2,78	-0.33319	0.447529	-0.23812		0.368001	0.631999
D. 3 /2,78	0.245997	0.228536	0.130304		0.129722	0.870278
D. 4 /2,78	-0.04882	0.65454	-0.08224		0.43757	0.56243
D. 5 /2,78	0.831878	-0.10829	-0.146		0.725064	0.274936
D. 6 /2,78	0.706942	0.22699	-0.19303		0.588551	0.411449
D. 7 /2,78	0.866339	-0.03881	0.051118		0.754663	0.245337
D. 8 /2,78	0.431068	0.185184	0.538608		0.510212	0.489788
D. 9 /2,78	-0.07814	0.135766	0.51206		0.286743	0.713257
D. 10 /2,7	-0.1813	0.671224	0.306908		0.577606	0.422394
D. 11 /2,7	0.160575	0.170617	-0.19557		0.093143	0.906857
D. 12 /2,7	0.094492	0.1709	0.752811		0.60486	0.39514
	2.400178	1.692637	1.642396	tot	5.735211	6.264789

Fig 2. The factor analysis of the VR dimension. To read the data: the three sub-dimensions are made up of four consecutive items.



Factor Matrix (unrotated)					
	1	2		Commun	Specific
D. 13 /4,17	0.290969	-0.2365		0.140596	0.859404
D. 14 /4,17	-0.23041	0.389519		0.204816	0.795184
D. 15 /4,17	0.204017	0.748453		0.601805	0.398195
D. 16 /4,17	0.364508	-0.46727		0.351211	0.648789
D. 17 /4,17	0.164866	0.666663		0.47162	0.52838
D. 18 /4,17	-1.1E-16	-5.9E-18		1.3E-32	1
D. 19 /4,17	-0.83706	0.000641		0.700677	0.299323
D. 20 /4,17	-0.78923	-0.08467		0.630045	0.369955
	1.662977	1.437794		3.10077	4.89923

Fig 3. The factor analysis of AR. To read the data: the two sub-dimensions are made up of four consecutive items.



Factor Matrix (unrotated)					
	1	2		Commun	Specific
D. 21 /4,17	5.34E-16	-3.3E-07		1.1E-13	1
D. 22 /4,17	9E-16	3.31E-07		1.1E-13	1
D. 23 /4,17	0.271101	0.244141		0.133101	0.866899
D. 24 /4,17	0.023326	0.709898		0.504499	0.495501
D. 25 /4,17	0.066085	0.386062		0.153411	0.846589
D. 26 /4,17	0.67969	-0.46531		0.678488	0.321512
D. 27 /4,17	0.611785	0.33099		0.483835	0.516165
D. 28 /4,17	0.798975	0.006897		0.638409	0.361591
	1.553028	1.038714		2.591742	5.408258

Fig 4. The factor analysis of LAR. To read the data: the two sub-dimensions are made up of four consecutive items.

Recalling that the factor loading, i.e., the correlation between the factor and the statistical variable, determines the weight of the connection between the factor and the variable, that the sub-dimensions are grouped into successive sets of four test questions, it can be said that the CFA substantially confirm the hypotheses on sub-dimensions that led to the construction of the test (Fig 2, 3, 4).

The psychometric characteristics of the test and of its items were determined with IRT analysis of the items.

Fig 6. The distribution of scores: the shape of the graph confirms the regularity of the test carried out by the students.

Item Analysis VR						
Discrim cutoff	0,27					
	D. 1 /2,78	D. 2 /2,78	D. 3 /2,78	D. 4 /2,78	D. 5 /2,78	D. 6 /2,78
Difficulty	0.62	0.48	0.53	0.23	0.91	0.91
Discrimination	0.63	0.30	0.54	0.44	0.18	0.28
Correlation	0.47	0.25	0.45	0.46	0.32	0.44
	D. 7 /2,78	D. 8 /2,78	D. 9 /2,78	D. 10 /2,78	D. 11 /2,78	D. 12 /2,78
Difficulty	0.92	0.82	0.97	0.77	0.85	0.44
Discrimination	0.23	0.45	0.06	0.52	0.18	0.53
Correlation	0.40	0.43	0.15	0.47	0.28	0.36

Item Analysis AR								
Discrim cutoff	0,27							
	D. 13 /4,17	D. 14 /4,17	D. 15 /4,17	D. 16 /4,17	D. 17 /4,17	D. 18 /4,17	D. 19 /4,17	D. 20 /4,17
Difficulty	0.91	0.53	0.92	0.68	0.97	1.00	0.86	0.86
Discrimination	0.24	0.79	0.18	0.39	0.05	0.00	0.32	0.32
Correlation	0.30	0.63	0.28	0.33	0.14	0.00	0.45	0.45

Item Analysis LAR								
Discrim cutoff	0,27							
	D. 21 /4,17	D. 22 /4,17	D. 23 /4,17	D. 24 /4,17	D. 25 /4,17	D. 26 /4,17	D. 27 /4,17	D. 28 /4,17
Difficulty	1.00	1.00	0.91	0.92	0.70	0.92	0.91	0.85
Discrimination	0.00	0.00	0.29	0.14	0.67	0.19	0.24	0.47
Correlation	0.00	0.00	0.47	0.25	0.54	0.43	0.47	0.64

Fig 7. The IRT parameters of the three dimensions of the assessment test.

The analysis shows that the distribution of the scores is regular with positive kurtosis (Fig 6).

To calibrate the test, questions with a high difficulty index (close to 1) and at the same time a low discrimination index (close to 0) should be replaced, indeed these questions are difficult and, at the same time, not able to select the students, reducing the sensitivity of the test. Subsequently, items 21 and 22 were eliminated (Fig 7).

4.3 The equivalent work groups creation.

The 12 work groups were formed by imposing that they had the same mean and the same variance. The equivalence conditions were verified by an ANOVA test on the sample of the 12 groups.

Recalling that the application condition of the ANOVA test is that the samples do not have significantly different population's variance (homoscedasticity), we first verify this condition as usual through Bartlett's test [6][7].

Bartlett's test is based on the variable:

$$B = \frac{df_W \ln(S^2) - \sum_{j=1}^{k} df_j \ln\left(S_j^{\,2}\right)}{1 + \frac{1}{3(k-1)}\left(\sum_{j=1}^{k} \frac{1}{df_j} - \frac{1}{df_W}\right)} \sim \chi^2(k-1)$$

being $S^2 = MS_W$ the variance within, $S_j^{\,2}$ the variance of the sample unit j, k the numbers of groups, df the degrees of freedom.

Bartlett test														df	11
														b-num	0.972199
work group	a1	b1	a2	b2	a3	b3	a4	b4	a5	b5	a6	b6		b-den	2.532179
c1	97.22	95.83	94.44	94.44	91.67	91.67	91.67	90.28	90.28	88.89	88.89	88.89		b	0.383938
c2	84.72	86.11	86.11	86.11	86.11	86.11	87.50	87.50	87.50	87.50	87.50	87.50			
c3	81.94	81.94	83.33	83.33	83.33	84.72	84.72	84.72	84.72	84.72	84.72	84.72		alfa	0.05
c4	81.94	80.56	80.56	80.56	79.17	79.17	79.17	77.78	77.78	77.78	77.78	76.39		p-value	1
c5	76.39	76.39	76.39	70.83	73.61	72.22	72.22	72.22	75.00	75.00	75.00	75.00		b-crit	19.67514
mean	84.44	84.17	84.17	83.05	82.78	82.78	83.06	82.50	83.06	82.78	82.78	82.50	error W		
df	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	4.00	48	sig.	no
1/df	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.02		
variance	60.18	54.56	45.88	73.68	46.89	54.61	57.31	54.60	41.86	37.23	37.23	41.09	50.42		
ln(var)	4.10	4.00	3.83	4.30	3.85	4.00	4.05	4.00	3.73	3.62	3.62	3.72	3.92		

Fig 8. Bartlett test: the table shows the sample's homoscedasticity.

Bartlett's test shows complete homoscedasticity of the sample, as evident from the b and b_{crit}b values (Fig 8).

Then we verified the equivalence of the groups through the ANOVA test at the 95% of confidence level and the null hypothesis that the means of the groups are equal (Fig 9). As $F \ll F_{crit}$, the test confirms the null hypothesis and at the same time provides the organizational structure of the work groups.

ANOVA: Single Factor								
DESCRIPTION					Alpha	0.05		
Group	*Count*	*Sum*	*Mean*	*Variance*	*SS*	*Std Err*	*Lower*	*Upper*
a1	5	422.21	84.442	60.17732	240.7093	3.175677	78.05687	90.82713
b1	5	420.83	84.166	54.56313	218.2525	3.175677	77.78087	90.55113
a2	5	420.83	84.166	45.87563	183.5025	3.175677	77.78087	90.55113
b2	5	415.27	83.054	73.67563	294.7025	3.175677	76.66887	89.43913
a3	5	413.89	82.778	46.88612	187.5445	3.175677	76.39287	89.16313
b3	5	413.89	82.778	54.60757	218.4303	3.175677	76.39287	89.16313
a4	5	415.28	83.056	57.30973	229.2389	3.175677	76.67087	89.44113
b4	5	412.5	82.5	54.6034	218.4136	3.175677	76.11487	88.88513
a5	5	415.28	83.056	41.85988	167.4395	3.175677	76.67087	89.44113
b5	5	413.89	82.778	37.22562	148.9025	3.175677	76.39287	89.16313
a6	5	413.89	82.778	37.22562	148.9025	3.175677	76.39287	89.16313
b6	5	412.5	82.5	41.08565	164.3426	3.175677	76.11487	88.88513
ANOVA								
Sources	*SS*	*df*	*MS*	*F*	*P value*	*F crit*	*RMSSE*	*Omega Sq*
Between Groups	25.76954	11	2.342685	0.046459	0.999998	1.99458	0.096394	-0.21185
Within Groups	2420.381	48	50.42461					
Total	2446.151	59	41.46018					

FIG 9. The ANOVA test shows the equivalence of the 12 working groups.

4.4 The Multiple Knowledge and Abilities Test (MKAT).

The following describes the design and the implementation of the tool for assessing the educational impact of the knowledge acquired in the four didactic paths. We will see that the evaluation will be done in terms of proficiency and awareness of the acquired knowledge, as it is necessary for the completeness of the measure.
The summative assessment tool of the educational processes is a test with a different structure from the previous one, made up of multiple choices questions. The objective of MKAT is no longer to evaluate the individual aptitudes suitable for carrying out a specific task but is to ascertain the individual possession of knowledge and abilities built by work groups through educational contact with IB or RB teaching methods, in Lab or E-Learning modalities, through which laboratory experiment is carried on.
The working hypothesis is that the level of individual acquisition of knowledge and skills, being a function of the combination of methods and modalities, can be evaluated from the equivalent work groups' performances through an appropriate score derived from individual one's. Furthermore, the division into 6 pairs of equivalent groups A, B, creates another statistical sample. Each statistical unit, i.e., each pair of equivalent groups A, B, is subjected to the stimulus by all four experiments conducted in both modalities. This allows the complete statistical analysis on the test results obtained.
The individual's knowledge and skills that the test must measure were selected:

O.1 knowledge of experimental apparatuses and of their operating principles.

O.2 knowledge of physical principles and/or relations between physical quantities to be verified or discovered.

O.3 ability to use knowledge acquired through the experiment.

With the selected objectives, a test was built, structured into 4 sections, one for each experiment. Every Section is made up of 4 multiple choice items (5 alternatives of which only one is correct), one short answer question, one numerical question: 6 questions each experiment, 24 totally. Each question is coupled with another, that aims to determine the student's perception about the correctness of the answer given (FIG 10).

What do not these two representations of a thermodynamic system have in common?

- o Macroscopic state
- o Microscopic state
- o Entropy
- o Temperature
- o Internal energy

Of the answer given I am:

- o Sure
- o Quite sure
- o Uncertain
- o Very uncertain

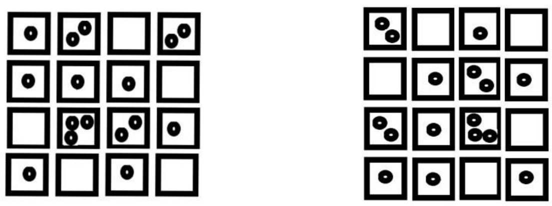


Fig 10. Example of item: The figure represents two microscopic states of the same macroscopic state. The question posed to the student is accompanied by another one which has the purpose of ascertaining the level of awareness of the answer given.

Following the approach according to which the perception arises from the encounter between external stimuli and the expectations, values, knowledge, and interests of the subject, who becomes an active builder of his own perceptions [8] [9], following the theoretical approach of the Implicit Theory of Knowledge, that was summarily described in chapter 1, we will statistically discriminate the impact of the teaching technique adopted in matching perception and results achieved.

Each item of the test is coupled with an item in which it is asked to self-evaluate whether the answer given previously is correct, choosing in a categorical set of four values. We considered two random variables $\beta = \{1; -1\}$ which corresponds to the correct or incorrect answer, $\gamma = \left\{1; \frac{1}{3}: -\frac{1}{3}; -1\right\}$ which corresponds to one of the four categories: sure, quite sure, uncertain, very uncertain. The random variable $\alpha = \beta\gamma$ measures no more the perception but the awareness, in fact it corresponds to a score greater than zero in cases in which the two random variables have the same sign, and to a score less than zero in the cases in which the two random variables have opposite signs.

The working hypothesis is that, on a scale between -1 and 1, the values of α greater than zero represent the correspondence between perception and reality, while those less than zero represent the conflicting perception with reality. [10]

Since not all members of the group underwent the test, group's dichotomous total score of each test section was normalized using the number of participants of that group. The normalized score of the test Sections is considered as a measure of impact of the combination of methods and modalities.

4.5 The analysis of the scores.

A1	1	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	1	1	0	0	0	1
A1	1	0	0	0	1	0	0	0	0	0	0	1	1	1	0	1	0	0	1	1	0	0	1	0
A1	1	0	0	0	0	0	0	0	1	0	1	1	0	1	1	0	1	0	1	0	1	0	0	0
A1	1	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	0	1	0	0	0	0
A2	1	0	0	1	1	0	1	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0
A2	1	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	1	0	1	0	0	1	0	1
A3	1	1	0	0	1	0	1	0	0	0	0	1	1	0	0	0	0	0	1	0	0	1	1	0
A3	1	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0
A4	1	0	0	1	0	0	0	0	1	1	0	1	1	1	0	0	0	0	1	1	0	0	0	0
A4	0	1	0	0	1	0	1	0	1	0	1	1	0	0	0	0	0	0	0	1	0	0	0	0
A4	1	0	0	1	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	0
A5	1	0	0	0	1	0	1	0	0	0	0	1	1	0	0	0	1	0	1	0	0	0	0	0
A5	0	1	0	1	0	0	0	0	0	0	1	1	1	1	0	0	1	0	1	1	0	1	1	0
A6	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0
A6	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	1	0	1	0	0	1	0	0
A6	1	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	1	0	1	0	0	1	0	0
A6	1	0	0	0	1	0	0	0	0	1	1	0	0	1	0	0	1	0	1	1	0	0	0	0
B1	0	0	0	0	1	0	0	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	0	0
B1	0	0	0	0	0	0	0	0	1	1	1	0	0	1	1	1	1	1	1	0	0	0	1	1
B1	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	1	0	0	1	0	1	0	0
B1	1	0	0	0	0	0	0	0	1	1	1	0	0	1	0	0	0	0	1	0	0	1	1	1
B2	1	1	0	1	1	0	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1
B2	0	0	0	0	1	0	0	0	1	1	1	1	1	0	0	0	0	0	1	0	0	0	0	0
B2	1	0	0	1	1	1	0	0	0	1	0	1	0	0	0	0	0	0	0	1	0	1	0	0
B2	1	0	1	0	1	1	1	0	1	0	1	0	1	1	0	0	0	0	1	0	0	0	0	0
B2	0	0	0	0	1	0	0	0	1	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0
B3	1	0	0	0	0	0	0	0	1	1	1	1	1	1	0	0	0	0	1	0	0	0	0	0
B4	0	1	0	0	0	0	0	0	1	1	1	1	0	0	0	0	0	0	1	0	0	0	0	0
B4	1	0	0	1	1	0	0	0	1	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0
B5	1	0	0	0	0	0	0	0	1	0	1	1	0	1	0	0	0	0	0	0	0	0	0	0
B6	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0
B6	0	0	0	0	1	0	0	0	1	1	1	1	0	0	0	0	0	0	0	1	0	0	0	0

Fig 11. MKAT score code of colours: E-Learning, Laboratory, IB, RB, Boltzmann Statistics, Brownian Motion, Gay-Lussac's Law, Thermal Expansion.

The data collected are sufficient to show that statistical methodology adopted has had an efficiency in discriminating the impact of methods and modalities (Fig 11).

We collected MKAT data by sorting it by rows and columns; the modalities are ordered by rows and the methods by column, according to the scheme identified in chapter 1 (Fig 12).

		R-B	I-B
AB1	LABORATORY	2.50	1.25
AB2	LABORATORY	1.14	2.71
AB3	LABORATORY	0.67	1.67
AB4	LABORATORY	1.20	2.20
AB5	LABORATORY	1.67	1.67
AB6	LABORATORY	1.83	1.67
AB1	E-LEARNING	2.25	2.00
AB2	E-LEARNING	1.43	2.71
AB3	E-LEARNING	2.33	2.67
AB4	E-LEARNING	0.80	2.80
AB5	E-LEARNING	2.00	2.33
AB6	E-LEARNING	1.33	1.67

Fig 12. The scores are sorted by rows with respect to the methodologies, by columns with respect to the methods.

Subjecting the data to the analysis, it has been established that the combination (Lab, RB) has the lower overage score and the higher variance than the other combinations. Furthermore, (E-learning-, IB) has the highest average score and the lowest variance, thus being the most efficient combination of Modalities and Methods from the point of view of the learning achieved by students.

Descriptive Statistics			
COUNT	balanced		
	R-B	I-B	
LAB	6	6	12
E-LEARNING	6	6	12
	12	12	24
MEAN			
	R-B	I-B	
LAB	1.50	1.86	1.68
E-LEARNING	1.69	2.36	2.03
	1.60	2.11	1.85
VARIANCE			
	R-B	I-B	
LAB	0.41	0.27	0.34
E-LEARNING	0.36	0.20	0.38
	0.36	0.28	0.38

Fig 13. It is shown that the combination (E-LEARNING; I-B) is associated with a higher score and a smaller variance, i.e., a higher didactical efficiency compared to the other combinations.

The didactical impacts of the different modalities are compared on each statistical unit A, B. The scores of the couple's E-learning experiments are added together, the same operation is performed for the scores of the Lab experiments, finally the result of each pair is exemplified in a graph. The graph shows the equivalence of the modalities scores. This hypothesis will be subsequently verified with a statistical test

Fig 14. Comparison between laboratory and online experiments: the sequence graphically shows the equivalence of the scores into the couples A, B.

The didactic impacts of the different methods are compared between each statistical unit A, B. The scores of the couple's I-B experiments are added together, the same operation is performed for the scores of the R-B experiments, finally the result of each pair is exemplified in a graph. The graph shows that IB scores are higher than RB scores. This hypothesis will be subsequently verified with a statistical test.

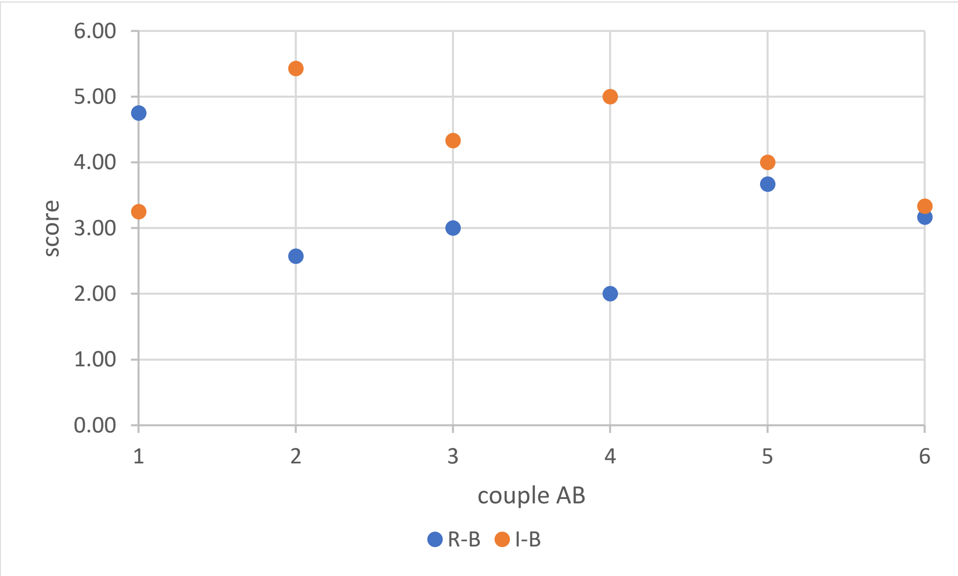


Fig 15. Comparison between I-B and R-B experiments: the sequence shows that I-B experiments are associated with higher scores than recipe-based experiments into each couple A, B.

A more in-depth analysis of results is obtained by means of test on Gosset's random variable T. For this purpose, we reorder the data in the following four tables:

		R-B	I-B				R-B	I-B
AB1	LAB	2.50	1.25		AB1	E-LEARNING	2.25	2.00
AB2	LAB	1.14	2.71		AB2	E-LEARNING	1.43	2.71
AB3	LAB	0.67	1.67		AB3	E-LEARNING	2.33	2.67
AB4	LAB	1.20	2.20		AB4	E-LEARNING	0.80	2.80
AB5	LAB	1.67	1.67		AB5	E-LEARNING	2.00	2.33
AB6	LAB	1.83	1.67		AB6	E-LEARNING	1.33	1.67
	mean	1.50	1.86			mean	1.69	2.36
		LAB	E-LEARNING				LAB	E-LEARNING
AB1	I-B	1.25	2.00		AB1	R-B	2.50	2.25
AB2	I-B	2.71	2.71		AB2	R-B	1.14	1.43
AB3	I-B	1.67	2.67		AB3	R-B	0.67	2.33
AB4	I-B	2.20	2.80		AB4	R-B	1.20	0.80
AB5	I-B	1.67	2.33		AB5	R-B	1.67	2.00
AB6	I-B	1.67	1.67		AB6	R-B	1.83	1.33
	mean	1.86	2.36			mean	1.50	1.69

Fig 16. Each table corresponds to a particular combination of methods and modalities, on which Gosset's test can be performed.

The scores are subjected to one-tailed test with the hypothesis H0: the overages are the same.

Comparing RB and IB methods in Lab modalities, the statistical tests confirm that the methods are equivalents, while IB scores is significantly larger than RB score in E-learning modality. This means that inquiry type teaching method is more efficient than the recipe teaching method when the didactics are carried on through the e-learning platform (fig 17).

T Test: Two Independent Samples on LAB									
SUMMARY			Hyp Mean	0					
Groups	*Count*	*Mean*	*Variance*	*Cohen d*					
R-B	6	1.501587	0.410174						
I-B	6	1.860714	0.265927						
Pooled			0.33805	0.617671					
T TEST: Equal Variances				Alpha	0.05				
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*
One Tail	0.335684	1.069838	10	0.154918	1.812461			no	0.320469
Two Tail	0.335684	1.069838	10	0.309835	2.228139	-1.10708	0.388823	no	0.320469
T TEST: Unequal Variances				Alpha	0.05				
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*
One Tail	0.335684	1.069838	9.56463	0.155473	1.820869			no	0.326919
Two Tail	0.335684	1.069838	9.56463	0.310947	2.241963	-1.11172	0.393463	no	0.326919

T Test: Two Independent Samples on E-LEARNING									
SUMMARY			Hyp Mean	0					
Groups	*Count*	*Mean*	*Variance*	*Cohen d*					
R-B	6	1.690873	0.362246						
I-B	6	2.363492	0.204822						
Pooled			0.283534	1.263184					
T TEST: Equal Variances				Alpha	0.05				
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*
One Tail	0.307427	2.187898	10	0.026763	1.812461			yes	0.568969
Two Tail	0.307427	2.187898	10	0.053526	2.228139	-1.35761	0.012371	no	0.568969
T TEST: Unequal Variances				Alpha	0.05				
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*
One Tail	0.307427	2.187898	9.284474	0.027774	1.826739			yes	0.583256
Two Tail	0.307427	2.187898	9.284474	0.055547	2.251636	-1.36483	0.019595	no	0.583256

Fig 17. I-B and R-B experiments carried out in the Lab (first table), or E-leaning (second table) modalities are compared.

Comparing Lab and E-learning modalities in R-B or I-B methods, the statistical tests confirm the equivalence of these combination i.e., the two methods can equivalently be carried on in laboratory or through an e-learning platform (Fig 18).

T Test: Two Independent Samples on R-B										
SUMMARY			Hyp Mean	0						
Groups	*Count*	*Mean*	*Variance*	*Cohen d*						
LAB	6	1.501587	0.410174							
E-LEARNING	6	1.690873	0.362246							
Pooled			0.38621	0.304583						
T TEST: Equal Variances				Alpha	0.05					
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*	
One Tail	0.358799	0.527554	10	0.304657	1.812461			no	0.164553	
Two Tail	0.358799	0.527554	10	0.609314	2.228139	-0.98874	0.610168	no	0.164553	
T TEST: Unequal Variances				Alpha	0.05					
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*	
One Tail	0.358799	0.527554	9.961647	0.304679	1.813169			no	0.164861	
Two Tail	0.358799	0.527554	9.961647	0.609357	2.229302	-0.98916	0.610585	no	0.164861	

T Test: Two Independent Samples on I-B									
SUMMARY			Hyp Mean	0					
Groups	*Count*	*Mean*	*Variance*	*Cohen d*					
LAB	6	1.860714	0.265927						
E-LEARNING	6	2.363492	0.204822						
Pooled			0.235375	1.036326					
T TEST: Equal Variances				Alpha	0.05				
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*
One Tail	0.280104	1.794969	10	0.051446	1.812461			no	0.493639
Two Tail	0.280104	1.794969	10	0.102892	2.228139	-1.12689	0.121333	no	0.493639
T TEST: Unequal Variances				Alpha	0.05				
	std err	*t-stat*	*df*	*p-value*	*t-crit*	*lower*	*upper*	*sig*	*effect r*
One Tail	0.280104	1.794969	9.834304	0.051699	1.815565			no	0.496762
Two Tail	0.280104	1.794969	9.834304	0.103399	2.233238	-1.12832	0.122761	no	0.496762

FIG 18. Laboratory and e-Learning modalities of carrying out experiments, carried out with the didactic method R-B (first table) or I-B (second table), are compared.

In summary: IB and RB methods have a stronger impact on learning in E-learning modality, while the other combination of Methods and Modalities are equivalent. The best learning results are achieved when the mathematical model of the physical system is provided to students through computer. It is reasonable to think that the computer permits to use models of real physical systems efficiently, allowing the student to determine model properties through conjecture and corroboration-refutation [11], while this road seams more laborious to be followed with the real system.

4.6 The analysis of Awareness.

A1	-0	0.3	1	0.3	-0	0.3	0.3	0.3	0.3	0.3	0.3	-0	-0	-0	-0	1	0.3	1	-0	-0		0.3	1	-0
A1	0.3	0.3		0.3	0.3	0.3	0.3	0.3	-0	0.3	0.3	-0	1	0.3	-1	0.3	-1	-0	1	0.3	-0	-0	0.3	0.3
A1	0.3	-0	1	-0	-0	0.3	0.3	0.3	-0	0.3	-0	-0	-0	0.3	-0		-0		-0	-0	-1	0.3		0.3
A1	0.3	-0	0.3	0.3	0.3					-1	0.3		-1	0.3			0.3	-0		0.3				
A2																								
A2	1	0.3		-1	-1	1	-1	1	0.3		1		-1	-1	-0		0.3		-0	-0		-1		-1
A3	0.3	0.3	-1	0.3	0.3		1		-1	-0	-1	1	1	-1	-1	0.3	-1	-1	1	-0		1	1	-0
A3	1	-0	-0	0.3	-1	0.3			0.3	0.3	-0	0.3	-0	-0	-0	-0	0.3		-0	0.3		0.3	-0	0.3
A4	0.3	-0	-0	0.3	-0		-0	-0	1	1	-0	0.3	0.3	1	-1		-1		1	0.3		-0	0.3	-0
A4																								
A4	-0	0.3		-0	0.3								-0	0.3	-0				-0	-0				
A5	-0	0.3	-0	0.3	-0	1	-0	1	1	0.3	1	-1	-1	1	1	1	-1	1	-1	1	1	1	1	1
A5	-1	0.3	0.3	1	-1	-1	-1	0.3	-0	-0	0.3	-0	1	1	-1	1	1	1	1	1	-0	0.3	0.3	-1
A6	0.3	1		1	1	0.3	0.3		0.3	0.3	-0	-0	0.3	0.3	-0		-0	0.3	-0	-0	1	1		
A6	0.3	-1	0.3	-0	-1	1	1	1	0.3	-0	-0	-0	-1	-0	1	1	-0	1	0.3	-0	0.3	0.3		1
A6	-0	0.3	1	-1	-1	1	1	1	1	1	0.3	0.3	-0	0.3	-0	1	0.3	0.3	1	1	1	-1	1	1
A6	0.3	-0	0.3	0.3	0.3					-0	0.3		-0	0.3	-0		0.3		1	-0				
B1	0.3	1		1	0.3				0.3			1	-0	0.3	0.3		-0	-0	-0	0.3		-0		-0
B1	-0	0.3	1	1	0.3	-0	1	1	1	1	0.3	1	-0	1	0.3	1	0.3	0.3	0.3	-1	1	0.3	0.3	0.3
B1	0.3	-0		0.3	0.3		0.3	0.3	-1	-0	0.3	0.3	-1	-1	-0		0.3	-0	0.3	0.3		-0		0.3
B1	0.3	-0		0.3	-1	1	-0	0.3	1	-0	-0		-1	1	-1				1	-1		-0	0.3	0.3
B2	1	0.3	1	0.3	1	1	0.3	0.3	-0	0.3	1	0.3	-1	0.3	-0		-1		1	-0		-0		1
B2	0.3	-0		0.3	0.3	0.3			-0	0.3	0.3	-0	-0	-0	-0		-0		0.3	0.3		0.3		1
B2	1	-0	1	0.3	1	-0	-1	-1	0.3	-0	0.3	-0	-1	-0	-0	1	-0			0.3	0.3	-0		0.3
B2	1	-0	1	0.3	1	-0	-1	-1	0.3	-0	0.3	-0	-1	-0	-0	1	-0			0.3	0.3	-0		0.3
B2	-0	0.3	-1	1	0.3	-1	0.3	-0	0.3	1	0.3	-0	0.3	-0	-0		-0		0.3	-0				0.3
B2	0.3	-0		0.3	0.3		0.3	0.3	-1	-0	0.3	0.3	-1	-0	-1		1	-0	0.3	0.3		0.3		0.3
B3	0.3	0.3	1	0.3	0.3	0.3	0.3	0.3	0.3	-0	-0	-0	0.3	-0	-0		0.3	-0	0.3	0.3	0.3	0.3		0.3
B4	-0	-0		0.3	-0	-0	-0	0.3	0.3	1	1	0.3	-1	-0	-1		-0		-0	-0		0.3		0.3
B4	-0	0.3	-0	0.3	0.3	1	-0	0.3	0.3	-0	0.3	0.3	-0	0.3	-0	-1	0.3	1	-0	-0	1	0.3	0.3	0.3
B5	0.3	-0		0.3	-1		1	1	-0		-0	-0	-1	-0	-0		0.3	1	0.3	-0		0.3		0.3
B6	-0	-0		0.3	0.3	1			-0			1	0.3	0.3	-0		-0		0.3	0.3		0.3	-1	
B6	0.3	-0		0.3	0.3	1			-0	1	0.3	0.3	-0	-1	-1		0.3		-0	0.3		0.3		0.3

Fig 19. Awareness score of MKAT, code of colours: E-Learning, Laboratory, IB, RB, Boltzmann Statistics, Brownian Motion, Gay-Lussac's Law, Thermal Expansion.

We subjected data to analysis of Awareness of knowledge acquired, as mentioned in the previous paragraph. We estimated mean and variation of stochastic variable α at 95% of confidence level, assuming that the behaviour of its sample mean is Gaussian, for each of the four experiments performed with different methods and modalities.

0.10	0.16	0.02	0.21
0.26	0.15	-0.19	0.18

-0.03	0.24	0.02	0.30	-0.14	0.18	0.05	0.37
0.14	0.38	0.03	0.28	-0.33	-0.05	0.07	0.28

Fig 20. Code of colours: E-Learning, Laboratory, IB, RB, Boltzmann Statistics, Brownian Motion, Gay-Lussac's Law, Thermal Expansion. The table shows the value and the confidence interval of the random variable Awareness.

As it is evident from the tables, the higher level of Awareness is achieved by the experiments in Modality LAB (yellow one's). This result is not surprising, we expect that the real manipulation of tools and devices contributes to the creation of knowledge and to raise awareness of the abilities reached. It becomes surprising considering that at the same time, modality E-learning (red one's) achieved the best MKAT score [12].

The experiment with the lowest level of Awareness is Gay-Lussac's one. The experimental apparatus is the most complex between those used in this research, moreover, carrying out the experiment involves complex manipulations of apparatus. Even for this experiment, while the scores obtained in MKAT do not distinguish between the Lab and E-Learning modalities, the level of awareness in E-learning modality is dramatically lower. We could conclude that the two modalities produce the same level of learning, but not the same level of awareness: E-learning modality does not favour the level of awareness of the acquired knowledge.

The critical data are subjected to factor explorative analysis.

Let's summarize the abilities and knowledge that each of the items wanted to determine:

I13 - Knowledge of pressure definition

I14 - Knowledge of cause of pressure in gas

I15 - Detailed knowledge of the mathematical relationship between p and T

I16 - Ability of preparing gas initial state

I17 - Knowledge of instrument operation principle

I18 – Ability of evaluating 0 K temperature.

Starting with Lab Modalities, PCA identifies two factors, the first connects items 13, 16, 18, which ascertain awareness about the knowledge necessary to carry out the experience with the R-B method i.e., knowledge of pressure definition, ability of preparing the gas initial state, ability of evaluating 0 K temperature; the second factor coincides with Item 17 i.e., Awareness about knowledge of instrument operation principle.

Fig 21. Scree plot of the awareness variable

Descriptive statistics						
	I13	I14	I15	I16	I17	I18
Mean	-0.15556	0.155556	-0.33333	0.666667	-0.14286	0.333333
Std dev	0.733189	0.640767	0.640513	0.503953	0.662994	0.745356
Skewness	0.506129	-0.49835	1.214305	-1.32288	-0.1229	-0.68999
Kurtosis	-0.91591	-0.33393	1.329545	0.875	-1.19495	-0.8

Factor Matrix Lab R-B					
	1	2		Commun	Specific
I13	0.812005	0.546498		0.958012	0.041988
I14	-0.49494	0.428061		0.428199	0.571801
I15	-0.75054	-0.64436		0.978511	0.021489
I16	-0.95871	0.196033		0.957555	0.042445
I17	-0.43499	0.812297		0.849041	0.150959
I18	-0.95075	0.183262		0.937508	0.062492
	3.479886	1.62894		5.108825	0.891175

Fig 22. Factor analysis of awareness variable

The statistics identify in item 16 the highest level of awareness. Item asks student to list the operations to be performed for preparing initial state of gas; this measure confirms the previously formulated hypothesis that operations carried out on the real instrument contribute to level of Awareness.

The item 14 concerns the microscopic gas model. The specific knowledge of the statistical properties of the system does not directly pertain to the execution of the experience but only to the possible conjectures that the student can formulate about how pressure is generated.

Switching to E-learning mode, PCA is performed by eliminating item 16 or item 18. The items elimination is due to the small number of responses which does not allow the algorithm to run.

It must be observed that system is not stable with respect to this elimination.

Descriptive statistics E.learning, R-B					
	I13	I14	I15	I17	I18
Mean	-0.54167	-0.08333	-0.41667	-0.02222	0.142857
Std dev	0.5288	0.590041	0.412759	0.495482	0.634126
Skewness	0.661965	0.432714	0.060192	0.130266	0.763587
Kurtosis	-1.0062	-0.15879	0.054841	0.182475	-1.68698

Factor Matrix E-learning, R-B without I16					
	1	2		Commun	Specific
I13	0.644023	-0.32957		0.523381	0.476619
I14	0.75207	0.40027		0.725825	0.274175
I15	0.933668	-0.04323		0.873604	0.126396
I17	-0.81325	0.232852		0.715591	0.284409
I18	0.151775	0.928669		0.885462	0.114538
	2.536517	1.187347		3.723864	1.276136

Descriptive statistics E-learning R-B					
	I13	I14	I15	I16	I17
Mean	-0.54167	-0.08333	-0.41667	0.5	-0.02222
Std dev	0.5288	0.590041	0.412759	1	0.495482
Skewness	0.661965	0.432714	0.060192	-2	0.130266
Kurtosis	-1.0062	-0.15879	0.054841	4	0.182475

Factor Matrix E-learning, R-B without I18					
	1	2		Commun	Specific
I13	0.982634	-0.18555		1	4.11E-15
I14	0.967946	0.251158		1	4.88E-15
I15	0.718827	0.695189		1	1.78E-15
I16	-0.41582	0.909446		1	0
I17	0.982634	-0.18555		1	3.55E-15
	3.55768	1.44232		5	1.43E-14

Fig 23. Descriptive statistics and factor matrix of the factorial analysis on data that differ in the elimination of item 16 or item 18. the main different factor is highlighted in yellow in the two cases.

as the first factor without item 16 is quite the complement of first factor of PCA without item 18.

Items 16 and 18 concern "Ability of preparing gas initial state" "Ability of evaluating 0 K temperature" respectively. Once included, they become the second factor of the PCA.

Like any other mathematical technique that tries to reduce the complexity of a system, stability is fundamental to establish the value of the solutions found.

Assuming that these results are reproducible, more in-depth research should identify technical, psychological, and behavioural reasons that contribute to these outcomes.

Bibliography.

[1] Bennett, G. K., Seashore, H. G., & Wesman, A. G. (1947). *Differential aptitude tests.* Psychological Corporation.

[2] Birnbaum, A. (1968). *Some Latent Trait Models and Their Use to Inferring an Examinee's Ability*. In Lord, F.M. e Novick, M.R.

[3] Boncori, L. (2002). Teoria e tecniche dei test. Torino: Bollati Boringhieri.

[4] Dunnette, M.D. (1990). La psicologia nella selezione del personal*e*. Milano: Franco Angeli.

[5] Charles Zaiontz, Real Statistics resource pack, Available: www. real-statistics. com/free-download/real-statistics-resource-pack/. (September 2016).

[6] GV Glass – Testing homogeneities of variances - American Educational Research Journal, 1966.

[7] M. W. J. Layard (1973) Robust Large-Sample Tests for Homogeneity of Variances, Journal of the American Statistical Association, 68:341, 195-198.

[8] Bruner, J. (1992). Another Look at New Look 1. *American Psychologist, 47*(6), 780–783.

[9] Bruner J. S., Goodman C. C. (1947). Value and need as organizing factors in perception. *The Journal of Abnormal and Social Psychology*.

[10] Morin A., Self-Awareness Part 1: Definition, Measures, Effects, Functions, and Antecedents, Social and Personality Psychology Compass 5/10 (2011).

[11] Karl R. Popper, CONJECTURES AND REFUTATIONS: The Growth of Scientific Knowledge, BASIC BOOKS, Publishers NEW YORK LONDON, (1962).

[12] D Richardson, Student perceptions and learning outcomes of computer-assisted versus traditional instruction in physiology, Advances in physiology education, 1997.

Conclusions and future work.

In psychology, the diagnostic tools are validated by their use and the experimental evidence, while this scientific necessity is not always properly considered in teaching. In teaching, the focus is placed on ascertaining the student's knowledge and skills, sometimes leaving in background the problem of verifying that, the instrument adopted for those assessments really measures what one wishes to measure. Recently the scientific literature has begun to take this topic head on [1][2]

In this thesis we adopted the opposite point of view, focusing the attention on the necessity to validate and standardize, not only the assessment tools, but also the adopted teaching methodologies. In this context the assessment of the student's knowledge and skills was aimed to determine the impact of the different didactic paths, and it was considered not a goal.

The objectives are typical of big educational institutions. To be achieved, they need structured organizations capable of carrying out the didactic methods and large statistical samples to verify their effectiveness. This work has had the sole ambition to show the feasibility the utility of the procedure on small samples, and to demonstrate the feasibility, we identified the statistical techniques suitable for the purpose in Factor Analysis, Item Response Theory and Analysis of Variance.

Factor analysis in all its forms (linear, non-linear, explorative, confirmative) is certainly the most powerful statistical technique to interpret the complex reality of educational processes. Moreover, the scale invariance of the raw test scores is a starting condition, even if insufficient, to be able to compare the results of identical experiments carried out in different educational environments. We discussed this inadequacy in chapter 2 of this thesis, associating it with the adoption of a probability distribution, from which a whole series of other requirements follow, such as that of having to get rid of the non-conforming experimental data. Finally, the comparison techniques between statistical samples, such as ANOVA, massively used in medicine, ethology, and biology, should be the fundamental technique for comparing samples also in the field of education and teaching. In this thesis we gave some examples of its application.

The European tradition in pedagogy, dating back to the thirties of the last century, supports the validity of the subject's experimental activity in the construction of reality. [3][4][5]. Sometimes this evidence does not seem to have compelling consequences on the organization of educational agencies. As far as we are concerned, we again observed and documented (chapter 3) various examples of this spontaneous construction in the didactic paths adopted. Indeed, it is simple to record the behaviours of physical systems (normally a solution of differential equations) when one can directly experience them. We called this knowledge empirical explanation (legal explanation in the Piaget's definition) and we do not consider this knowledge the most significant step of the learning process.

We argued that individual's need and desire to find a causal explanation is the most important stimulus to proceed through the learning process, and it is the specific characteristic making the student's activity like the scientist's one. This phase, which with Piaget we called "abstraction process", is entirely influenced by culture (by the set of the knowledge and skills acquired and possessed in the field of the personal experiences).

This aptitude to seek the causal explanation offers evident adaptive advantages in terms of the ability to predict, to recognize the situations and finally it forms the basis of the ability to consciously manage one's social life. In the actual information and communication society, the attitude of searching for the causal explanation has an evident social value and should be considered as one of the objectives of every educational agency.

The operationalization of a scientific content is a specific phase of science teaching. It consists in identifying the operations to be carried out, that is, the relevant set of didactic interventions aimed at teaching. The didactic action always consists in the substitution or metaphorization of a content and in this sense an analogous system realizes the substitution, this time in a material way, that is, through a device that simulate another physical reality, which cannot be experienced directly.

An advantage of the analogous systems we adopted is that they reproduce, on a macroscopical scale, what is hypothesized to happen on a microscopic or mesoscopic scale, making visible the temporal evolution of the systems and transforming the mathematical model into a student's practical-manipulative activity, without any approximation. In this sense, the macroscopic system becomes the tool for determining the evolution of the associated microscopic systems, that is, given the initial conditions, it provides the solutions of the common mathematical model.

While the analogous systems are an important cognitive tool in physics research, especially if it is not possible to access one of the systems linked by analogy [6], they are less frequently used for didactic purposes. In our concrete experience, for the results collected by the students and for the investigation carried out on the static experiment we are confident, and we have partly verified on samples, that the didactical use of analogous systems:

- promotes the awareness of the knowledge achieved,
- exposes the students to the chance of obtaining original experimental results,
- replaces a part of theoretical knowledge that allows for the synthesis of the causal explanation [7].

Another example on this last point is given. Using a chessboard and pawns we realized a Markovian type of process that, starting from a not equilibrium state, makes the system converge to the results of Bose-Einstein Statistics. It means that, in every case it is necessary to reproduce these statistics, the chessboard can provide the result. The chessboard is used like a calculator of the particular algorithm we are interested in, as in the case of black body radiation; associating the number of standing electromagnetic waves in a cavity per unit of volume and frequency to the number of squares on the board, the average energy of the standing waves at a given temperature to the average number of pawns per square, it is immediate to establish, from the pawn distribution on the board at equilibrium, the energy distribution in the cavity [8].

The idea of being able to create a particular analogous system, suitable for the operationalization of a particular phase of a didactic process, has led us to think about the possibility of producing an algorithm capable of facilitating these productions. Statistical mechanics is a very large field on which to test this possibility to the extent that this part of physics can simulate real thermodynamic systems.

We identified the computer structure that would allow the production and distribution of the experiment to the students, the algorithm that simulates the thermal and adiabatic interactions. Starting from an initial state of one or more sets of oscillators, exchanging quanta of energy each other, the system evolves, probing the states accessible, until the system reaches and fluctuates around the states of maximum entropy.

We characterized the states of matter in the following way: for the solid and liquid by the positions and any movements of the oscillators in a minimum volume, for the gas by the position inside a maximum volume; the changes of state are determined by conditions on the position and on the energy of the single oscillator. In addition to the algorithms used in our didactic experimentation, others have been created to prove the functioning and correspondence of the simulation results with the properties of thermodynamic equilibria.

An “author system” can also be realized for Brownian motion, but with more limited possibilities than in the previous case. In addition to relating the frequency and length of the random walk step with the dynamic and thermodynamic characteristics of the medium in which the particle floats, it is possible to simulate the motions in the presence of fields of external forces by appropriately varying the probabilities of the random walk steps.

The results of MKAT showed that the level of learning obtained from the didactic path that brings together inquiry experiments and their execution by means of the e-learning platform is significantly higher than the other three possible combinations. Indeed, (IB E-learning) combination has the highest score average and the lowest score variance respect the other paths.

Based on the results documented in chapter 4, the "recipe" and "inquiry" experiments, if carried out in the laboratory, achieve the same learning levels, while the carrying out through the e-learning platform significantly improve the outcome of the "inquiry" experiment. This result shows that there is an environment (in this thesis we have used the term modality) in which it is more efficient to carry out one type of experiment rather than another. We have formulated some hypotheses that can justify this outcome. Certainly, when an "inquiry" type experiment is carried out, the efficiency in terms of clarity of the results and in terms of the time taken to obtain them is one of the hypotheses we have indicated. But there is another element in the efficiency of the combination we must consider, that is the commonality of representative languages used between the type of experiment and its algorithmic representation, which must be investigated, not only from the point of view of learning, but also from that of the awareness of acquired learning, as it has been clarified in the last paragraph of chapter 4 and as we recall in the following.

The data extracted from MKAT allowed us to estimate the value of the awareness by interval. The Lab-IB experiment "Boltzmann statistics" is the experiment that obtained the best awareness values. The score mean is larger, and its standard deviation is smaller than any other combination of methods and modalities. All the experiments carried out in the laboratory obtained higher awareness values than the same experiments carried out in E-learning modality.

We were presented with an apparent paradox: while the knowledge acquired by the students is higher for experiments in the E-learning modality, the corresponding level of awareness of the acquired knowledge is higher for the experiments in Lab modality.

We have verified that this paradox present itself in other circumstances and for other disciplines [9]. For the theoretical framework we are referring to, these score behaviours seem completely natural, that is, it seems natural that a surgeon who performs a surgery for the first time after having trained himself through a simulator, should feel a certain insecurity. But learning obtained through a didactic path does not involve, if not towards itself, the responsibilities that a surgeon normally must take on for the others. It must be concluded that the manipulation of a real system acts at a deeper and more elementary level because it is more certain, less subject to relationships with other knowledge that could endanger those acquired through practical experience. This contributing to greater the awareness, as an expected coincidence between what one has rationally thought and what it is found experimentally. We described in chapter 1 this process of simplification, acting during the assimilation to an action scheme.

The EFA on the awareness data would have been necessary, to understand on which specific objectives the level of awareness falls and maybe to confirm our hypotheses, but the incompleteness of the data collected and, above all, the instability of the system with respect to the necessary deletion of data did not allow the analysis, as we documented in chapter 4.

Bibliography

[1] Cole Walsh, Katherine N. Quinn, C. Wieman, N. G. Holmes. Quantifying critical thinking: Development and validation of the physics lab inventory of critical thinking. PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH 15, 010135 (2019).

[2] Maja Planinic, William J. Boone, Ana Susac, Lana Ivanjek. Rasch analysis in physics education research: Why measurement matters. PHYSICAL REVIEW PHYSICS EDUCATION RESEARCH 15, 020111 (2019).

[3] Maria Montessori. The Montessori Method with an introduction of J. McV. Hunt and J. Valsiner. Transaction Publisher (2014).

[4] Bruner, J. S. (1961). The act of discovery. Harvard educational review.

[5] Ernst von Glasersfeld. Radical Constructivism: a way of knowing and learning, Routledge. 6 ago 2013.

[6] Hong Guo, Hang Liu, Xiao-Mei Kuang, and Bin Wang. Acoustic black hole in Schwarzschild spacetime: Quasinormal modes, analogous Hawking radiation, and shadows. Phys. Rev. D **102**, 124019 – 7 December 2020.

[7] Marco Cattaneo, Gabriele De Chiara, Sabrina Maniscalco, Roberta Zambrini, Gian Luca Giorgi. Collision Models Can Efficiently Simulate Any Multipartite Markovian Quantum Dynamics. Phys. Rev. Lett. **126**, 130403 – Published 2 April 2021.

[8] I. Borsini. An analogous experiment on the energy quantization and black body radiation. Giornale di Fisica, VOL. LXII, N. 2 Aprile-Giugno 2021

[9] D Richardson, Student perceptions and learning outcomes of computer-assisted versus traditional instruction in physiology, Advances in physiology education, 1997.