JOURNAL OF BALTIC SCIENCE EDUCATION
ISSN 1648-3898 /Print/
ISSN 2538-7138 /Online/

ITEM DIMENSIONALITY EXPLORATION BY MEANS OF CONSTRUCT MAP AND CATEGORICAL PRINCIPAL COMPONENTS ANALYSIS

Gabriela Deliu, Cristina Miron, Cristian Opariuc-Dan

Abstract. *The aim of this research is to study the merits and complementarity of Construct Mapping and Categorical Principal Components Analysis as two approaches that explore the dimensionality of multiple-choice items in achievement tests. Data from the two forms of the Romanian National Assessment Tests on Science were used to explore the dimensionality of items and to identify potentially problematic items that affect the equivalence of the two parallel forms. The findings confirm that the two tests have at best partial equivalence, but while the two methods both agree on test unidimensionality, they flag in part different items as potentially problematic. The results enable researchers and practitioners to make coherent data-driven decision regarding the use of unidimensional vs multidimensional IRT models.*

Keywords: *categorical principal components analysis, construct map, item response theory, unidimensionality.*

Gabriela Deliu, Cristina Miron
University of Bucharest, Romania
Cristian Opariuc-Dan
"Ovidius" University, Romania

Introduction

An assessment should mean rigor and accuracy, regardless of it's field and purpose. In the domain of education, the applied tests aim at determining the students' acquired competences development level by participating in a particular course or for determining the level of cognitive development achieved at the end of a certain schooling period. Tests designed for assessing these competences are based on one of two great theories: Classical Test Theory (CTT) or Item Response Theory (IRT). Many of the educational assessment tests used at the present are elaborated in the paradigm of the CTT. This theory is used on a large scale owing to its simplicity of implementation (Chambers, 2014; Ding & Beichner, 2009; Engelhardt, 2009). Nevertheless, the CTT has a major drawback: it does not allow for the dissociation of the test characteristics from the characteristics of the test subjects, with the ability of the examinee defined only in terms of a particular test (Chambers, 2014). When the test is difficult, the examinee seems to have low ability, and when the test is easy, the examinee gives the impression of high ability, so the difficulty of the test/items hinges upon the measured ability of the examinees, while the ability of the examinees hinges upon the difficulty of the test/items (Xitao, 1998). The characteristics of the test and of the items change with a new group of examinees, and the characteristics of the examinees change in accordance with the test. It follows that it is extremely difficult both to compare examinees who take different tests and to compare the characteristics of the items obtained as a result of using different groups of examinees (Hambleton, Swaminathan, & Rogers, 1991). This major drawback of the CTT can be avoided if the results of the tests are analysed in the paradigm of the IRT. In this theory, the assessment of the examinees' ability does not depend upon the test used, and the characteristics of the items are not determined by the group that responded to the items (Hambleton & Jones, 1993). Even if IRT seems highly efficient and promises the identification of the examinees' real level of ability, it is also extremely restrictive from the point of view of assumptions (DeMars, 2010). In order to adequately assess the parameters of the items, one should verify in preliminary analyses the unidimensionality

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

(Wang & Bao, 2010) and the local independence of the items (Liu & Maydeu-Olivares, 2012), and afterwards check the adequacy of the measurement model. When these assumptions are not fulfilled, the application of a unidimensional IRT model leads to erroneous results for the performed analysis. (Hambleton, Swaminathan, & Rogers, 1991).

With all these restrictions, the rigor and accurasy of the IRT has led to its increasing use in the development and interpretation of educational assessment tests. The literature abounds in tests based on IRT (e.g., Ding, Chabay, Sherwood, & Beichner, 2006; Engelhardt & Beichner, 2004; Eshach, 2014; Önder, 2016).

In Romania, preoccupations for the IRT in the field of educational assessment in pre-university education remain largely theoretical. Standardized assessment tests based on the IRT are found on a single educational platform, BRIO (http://www.brio.ro), that has a private initiative backing, while national evaluation tests (http://www.subiecte.edu.ro), with a major impact on a pupil's educational path, are being developed by the National Center for Assessment and Examination in the CTT paradigm.

For instance, as of 2014, all the students belonging to the 12-13 age group of Romania have been administered, at the end of the 6th grade, a transdisciplinary evaluation test in mathematics and natural sciences. The two parallel forms of this test were built in CCT and aimed to assess the level of development for a set of abilities common to mathematics, physics and biology (National Center for Assessment and Examination, 2017). The results obtained allow for a (pre)orientation of students in higher stages of schooling either towards academic high schools (where students predominantly develop academic cognitive abilities) or towards technical high schools (where students predominantly develop practical abilities).

This study analyzes the Mathematics and Natural Sciences Test items administered to 6th grade students with the purpose of measuring item characteristics through the application of the IRT instruments.

Since the adequacy of the pattern model for measurement is conditional, in the IRT, on the assumption of the unidimensionality of the items, the analysis was focused on the answer to the following questions:

- What is the construct, respectively the scale of the construct used in the elaboration of the items in the Mathematical and Natural Sciences Tests?
- Are the items of the Mathematics and Natural Science Tests unidimensional?

The answer to these questions makes it possible to orientate the determination of the characteristics of Mathematics and Science Test items to unidimensional or multidimensional IRT models.

Research Methodology

General Background

In order to identify the constructs measured by the tests, the graphic instrument of construct map was used (Derbentseva, Safayeni, & Canãs, 2007; Novak & Canãs, 2007; Opariuc-Dan, 2014). Designing a test generally involves the following stages: defining the construct to be measured, identifying the set of behaviours associated to the said construct (National Research Council [NRC], 2000; Sabella & Redish, 2007), designing some items whose solving entails externalizing the behaviours associated to the measured construct (Brown & Wilson, 2011; Osterlind, 1998; Wilson, 2005). Nevertheless, the present analysis proceeded in reverse order: it started from the items of the Mathematics and Natural Science Tests, then identified the behaviours associated with the solving of these items, and, based on this, finally identified the construct measured by the test.

Aiming at identifying to what extent the constructs saturated the test items, the Categorical Principal Components Analysis (CATPCA) was used, a technique of processing nonparametric data, which does not differ significantly from the Exploratory Factor Analysis, whose application was not possible since the test items were processed as dichotomous items. These items do not respect (at least) the interval scale, do not correlate linearly, but logistically, and do not present a normal character of distribution (Opariuc-Dan, 2012). CATPCA, like the Exploratory Factor Analysis, aims at extracting a number of latent factors common to a set of variables, appearing initially under the form of independent factors, based on the analysis of correlations (Ding & Beichner, 2009; Jolliffe, 2002). In this case, the 15 items of each test represented 15 independent factors (independent variables). The responses to these items showed a proper variance, owing to some specific factors (such as the individualization of an item by means of its phrasing – a different phrasing would have led to slightly different results), and a common variance, owing to some common factors, whose nature is usually psychological (such as numerical, verbal, etc. abilities) (Sava, 2011). Using common variance, CATPCA reduced to a minimum the number of factors/eigenvectors responsible for the variance of the initial variables.

https://doi.org/10.33225/jbse/19.18.209

Therefore, the purpose of this analysis was to extract a number of eigenvectors, as small as possible, accounting for as much as possible the variance of the results. The nature of these factors was mentioned in the "Discussions" section. CATPCA was used in this paper to find whether the test items are unidimensional, in other words, whether the variance of the results was influenced by the existence of a single (dominant) factor, or there were several factors influencing the variability of the results.

Evaluation Instrument

The two tests, upon which the present analysis is based on, were administered in 2017, within the National Evaluation Examination for Sciences, 6th grade. Named Test 1 and Test 2, they were supposed to be equivalent and were administered simultaneously, almost half of the students taking Test 1, and the other half Test 2. Each test consisted of 15 items, built on the model of TIMSS and covering the disciplines: physics, mathematics and biology (5 items for each discipline). Structurally, the items are, at face value, interdisciplinary. A mere browsing of the tests showed that the items had, essentially, a monodisciplinary structure, with the interdisciplinary intention solved by integrating them awkwardly in a "story". This gave the tests at most a multidisciplinary character, but in no way an interdisciplinary one. The tests had a quasi-identical structure (the same number of items per discipline, the same position of items within the tests, the same wording of statements), the only items that had a certain variability from one test to another being the biology ones. Out of the 15 items, 5 were multiple-choice, and 10 were open-answer questions. Nevertheless, in the present analysis, in order to eliminate the factor of examiner subjectivity in allotting a response code to the open-answer items, all items were treated as dichotomous, scoring them 0 for an incorrect or incomplete answer and 1 for a correct/complete answer.

Database under Research

The primary data representing the object of processing in the present paper were collected from a large number of secondary schools, which resulted in a set of data made up of the pattern of responses given by 1104 students who took Test 1 in 2017 and 1053 students who took Test 2 in 2017 (out of 73711 students who took Test 1 and 69221 who took Test 2, nationwide). All the participants were 6th-grade students, aged 12-13, out of whom 1038 are girls (48.1%) and 1119 boys (51.9%). All students took the test in their own school class, at the same time, and had at their disposal a total of 60 minutes to solve it. The students received the test on paper support. The data was collected by consulting the assessment papers of the students taking the tests.

The research was done following the ethical prescriptions of research with human subjects, valid in Romania at the time of data collection.

Data Analysis

The used analysis consisted of the following steps:

In the first step of the analysis, in order to identify the construct and scale of the construction measured by each test, the items were grouped in clusters according to the stated tasks and according to the behaviors associated with solving them. This stage of the analysis has ended with the representation of the construct map of each test.

In the second stage of the analysis, the number of dimensions upon which the construction of the tests was based were evaluated. In order to achieve this, the analysis resorted to CATPCA. The IBM SPSS Statistics application with the Categories module was used, by accessing the Optimal Scaling option within the Dimension Reduction sub-menu of the Analyze menu (George & Mallery, 2009; Opariuc-Dan, 2012).

In the third stage, based on the results obtained in the first two stages, the items of the two tests were compared and a series of equivalence problems were discovered.

Research Results

The construct map elaborated during the first stage of the analysis, starting from the items of Test 1, has the structure shown in Figure 1.

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

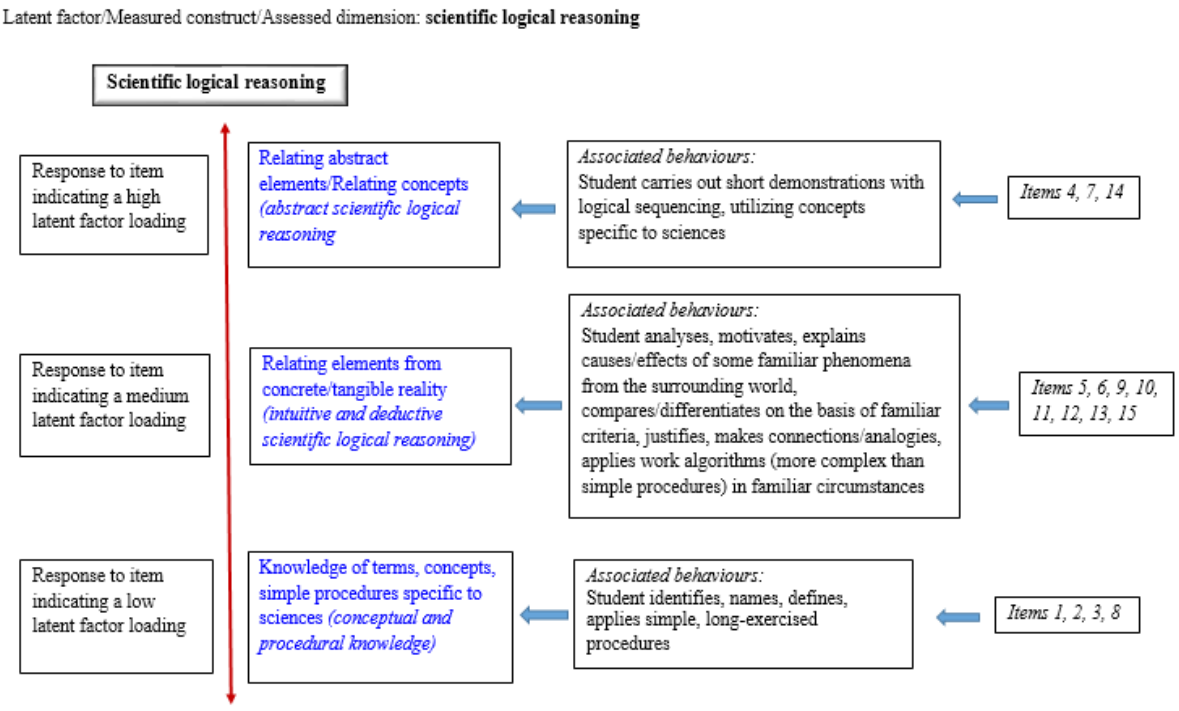


Figure 1. Map of the "scientific logical reasoning" construct – Test 1 of 2017.

Test 2 (2017) had the same structure as Test 1, which is only natural considering the necessity of equivalence for the two tests. Under these circumstances, the construct map for this test has a similar structure to the construct map for Test 1.

In the second stage of the analysis, 1081 students' results obtained from Test 1 (2017) were processed in CATPCA, 23 cases being excluded from the analysis. The convergence of the model with two components defined a priori was obtained after only one iteration, resulting in that the first component accounted for about 22.40% of the variance of all items (Eigenvalue=3.361; Cronbach Alpha=0.753). The second component had a low consistency (Cronbach Alpha =.257), accounting for only 8.82% of the common variance of the items (Eigenvalue=1.323). Therefore, the considered bidimensional or axial-dimensional model accounted for about 31.22% of the common variance of the items (Eigenvalue=4.685; Cronbach Alpha =.843), the difference being accounted for by potential other components or representing the proper variance of the items (residual variance).

Item 5 had extremely low values of the coordinates in both dimensions ($Centroid_{(1)}$=0.006; $Centroid_{(2)}$=0.014; $m_{(Centroid)=}$0.010; $Vectorial_{(Total)=}$0.019). Considering this, it was decided to remove it from the analysis. Resuming the procedure without *Item 5*, a slight increase in the proper value of the correlation matrix (Eigenvalue=4.728) was found, which indicated that a bidimensional model could account for 33.75% of the variance of the 14 items. Dimension 1 remained the most important, saturating the items in a ratio of 23.95% (Eigenvalue=3.354; Cronbach Alpha =.756). A second dimension was identified, which saturated the items in a ratio of 9.80% (Eigenvalue=1.373; Cronbach Alpha =.293), with *Item 5* affecting the second rather than the first component.

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

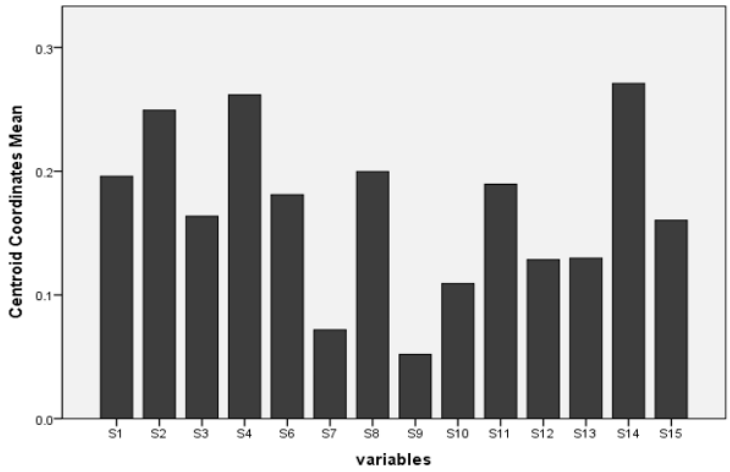


Figure 2. Variance of each item, accounted for by the bifactorial model – Test 1 of 2017.

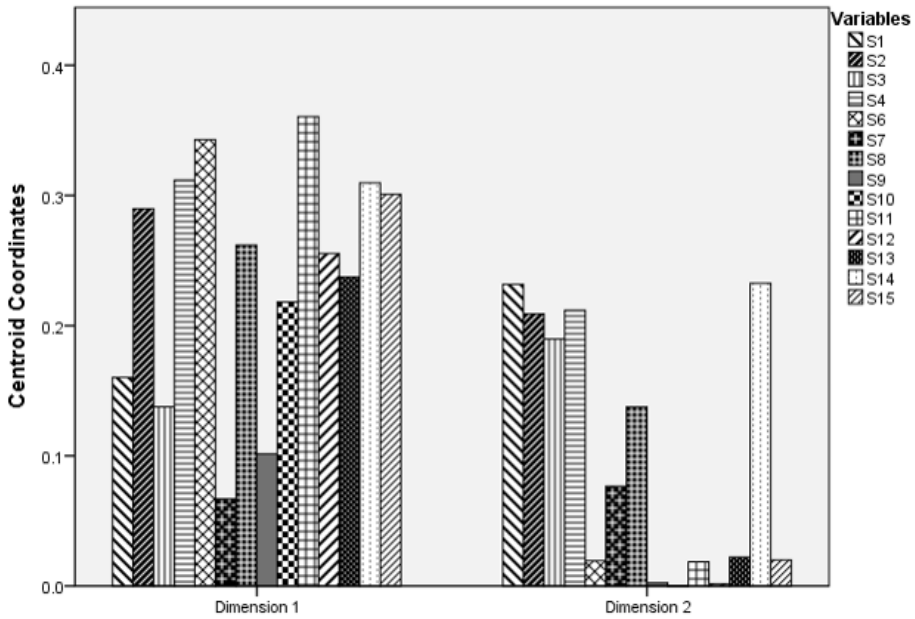


Figure 3. Variance of each item, accounted for by the first and the second dimension – Test 1 of 2017.

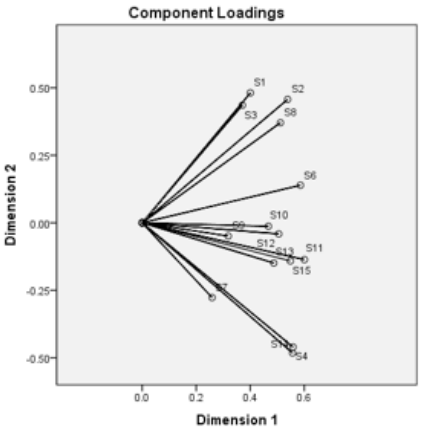
The table of factor saturation for Test 1:

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Table 1. Item loading in components – Test 1 of 2017.

	Dimensions	
	1	2
S1	.400	.481
S2	.538	.457
S3	.371	.436
S4	.558	-.460
S6	.586	.139
S7	.259	-.277
S8	.512	.371
S9	.319	-.049
S10	.467	-.013
S11	.600	-.136
S12	.505	-.041
S13	.487	-.149
S14	.556	-.482
S15	.548	-.141



The graphical representation of the answers of the students who took Test 1 in relation to the two presumed existing dimensions is shown in Figure 4, Figure 5, Figure 6 and Figure 7.

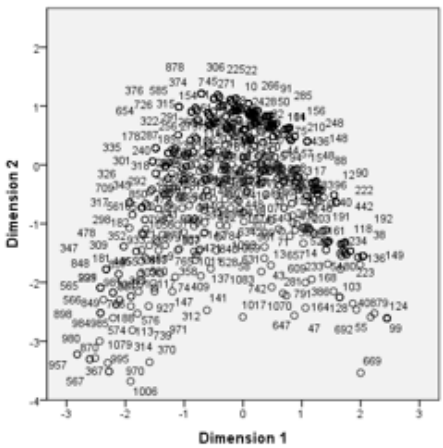


Figure 4. Point cloud of scores on the two dimensions – Test 1 of 2017.

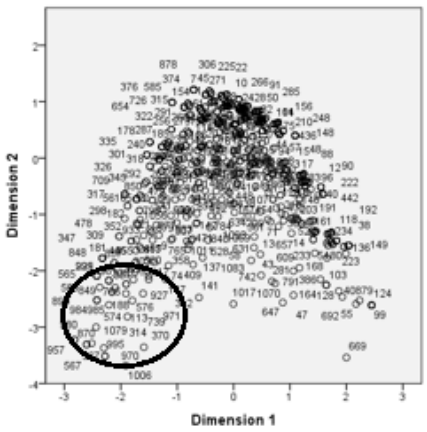


Figure 5. Point cloud of scores for the students with very low performance – Test 1 of 2017.

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

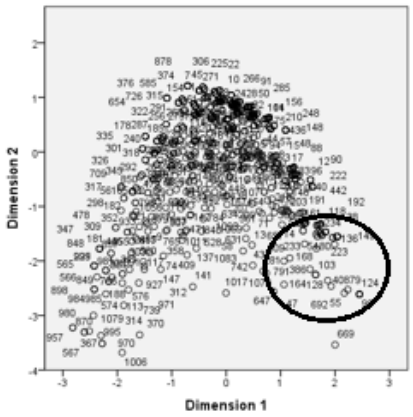


Figure 6. Point cloud of scores for the students with very high performance – Test 1 of 2017.

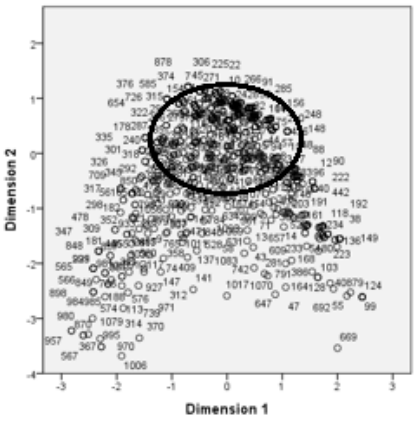


Figure 7. Point cloud of scores for the students with high performance in the area of conceptual and procedural knowledge and in the area of intuitive scientific logical reasoning – Test 1 of 2017.

For the second stage of the analysis for Test 2 (2017) we performed a CATPCA analysis , based on a number of 1035 cases, 18 not being included in the analysis because of the absence of some data. The matrix convergence was achieved after a single iteration, as there was no significant increase of variance at the level of the last iteration. The proper value was 4.660, indicating the fact that a bidimensional model could account for 31.06% of the common variance of the items. The most important dimension was the first, which saturated the items in a ratio of 22.20% (Eigenvalue=3.330). A second dimension was also identified, saturating the items in a percentage of 8.86 (Eigenvalue=1.330). In the case of the first dimension, the consistency of the items was good (Cronbach Alpha=.750), but the consistency of the second dimension was quite low (Cronbach Alpha =.266). Overall, considering both dimensions, the consistency of the items was 0.842. The same as in the case of Test 1, *Item 7* had low centroid coordinates for both dimensions ($Centroid_{(1)}$=0.062 and $Centroid_{(2)}$=0.078), while all the other items had relatively high centroid coordinate values in the first dimension. *Items 6* and *11* were the most characteristic for Dimension 1 (Figures 8 and 9). Their centroid coordinates had the values $Centroid_{(1)}$=0.351 for *Item 6* and $Centroid_{(1)}$=0.349 for *Item 11*, but they had very low values for Dimension 2: $Centroid_{(2)}$=0.003 for *Item 6* and $Centroid_{(2)}$=0,065 for *Item 11*.

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

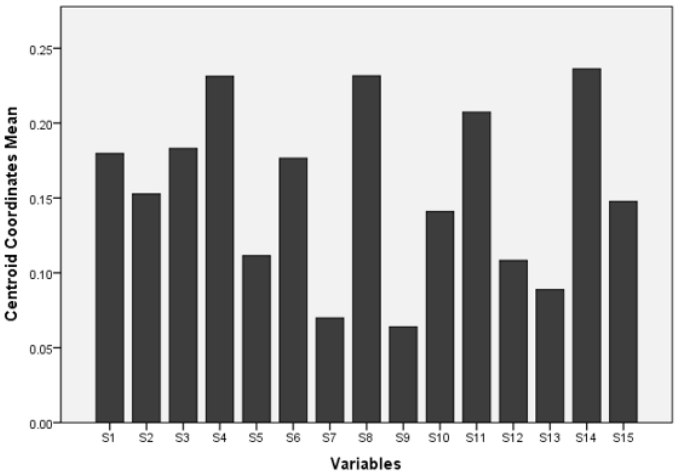


Figure 8. Variance of each item, accounted for by the bifactorial model – Test 2 of 2017.

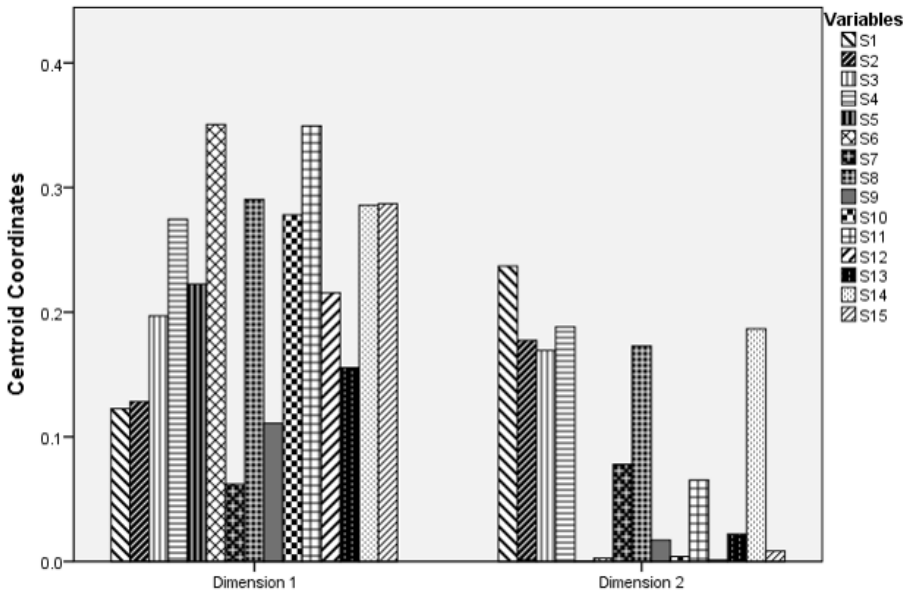


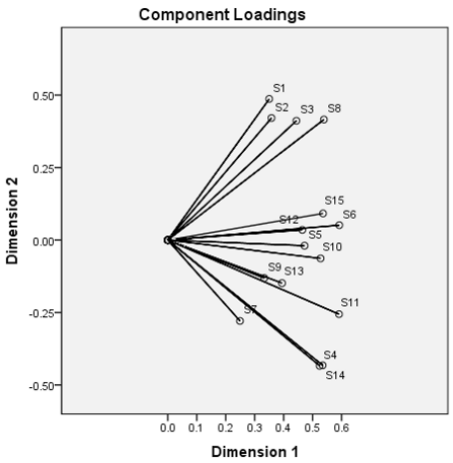
Figure 9. Variance of each item, accounted for by the first and the second dimension – Test 2 of 2017.

The table of factor saturation for Test 2:

https://doi.org/10.33225/jbse/19.18.209

Table 2. Item loading in components – Test 2 of 2017.

	Dimensions	
	1	2
S1	.350	.487
S2	.358	.421
S3	.444	.411
S4	.524	-.434
S5	.472	-.019
S6	.592	.051
S7	.249	-.279
S8	.539	.416
S9	.333	-.131
S10	.527	-.063
S11	.591	-.256
S12	.464	.035
S13	.394	-.149
S14	.535	-.432
S15	.536	.092



The third stage of the analysis is based on the interpretation of the results shown in Figure 10, Figure 11, Figure 12 and Figure 13.

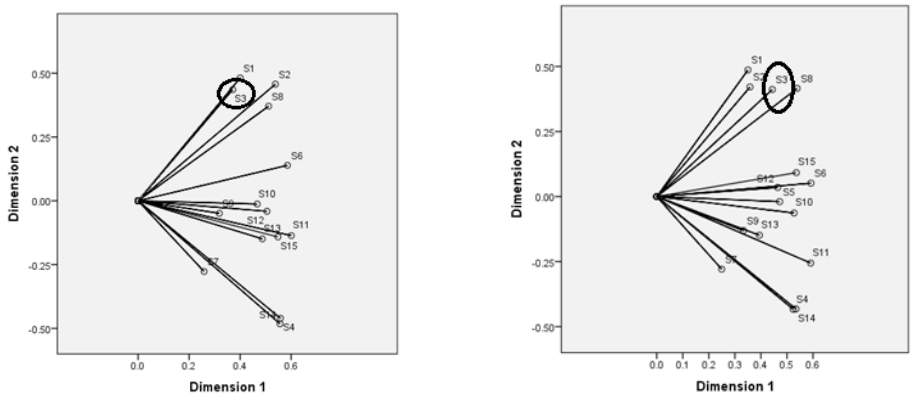


Figure 10. Position of the vector associated with Item 3 – Test 1 of 2017 (left), Test 2 of 2017 (right).

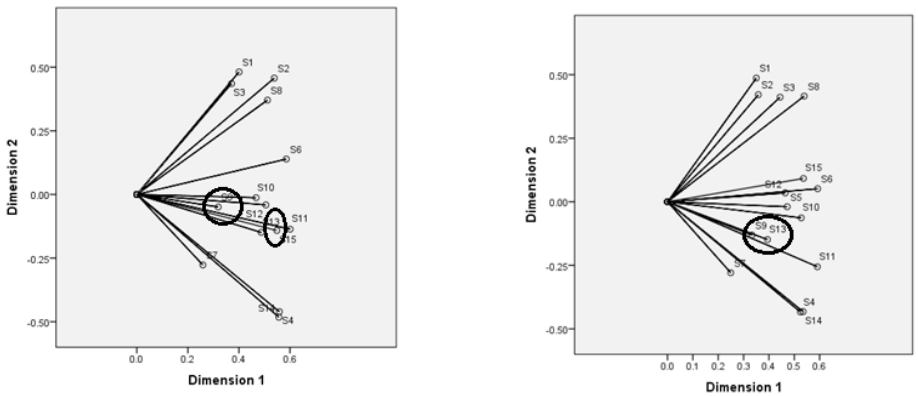


Figure 11. Position of the vectors associated with Items 9 and 13 – Test 1 of 2017 (left), Test 2 of 2017 (right).

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

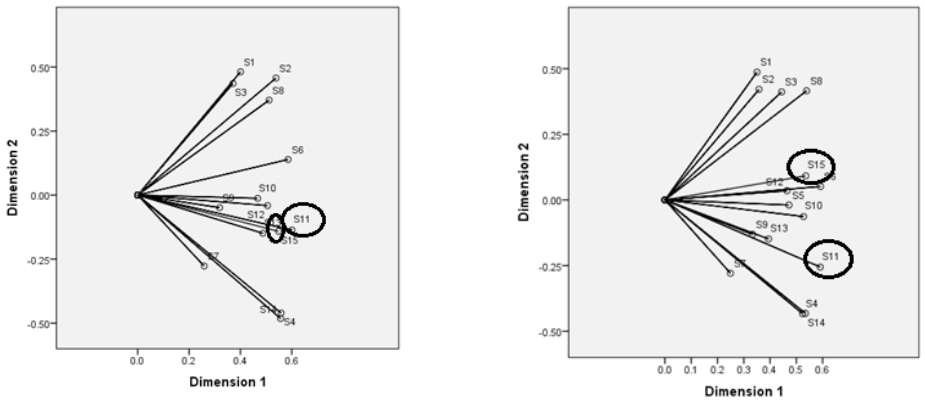


Figure 12. Position of vectors associated with Items 11 and 15 – Test 1 of 2017 (left), Test 2 of 2017 (right).

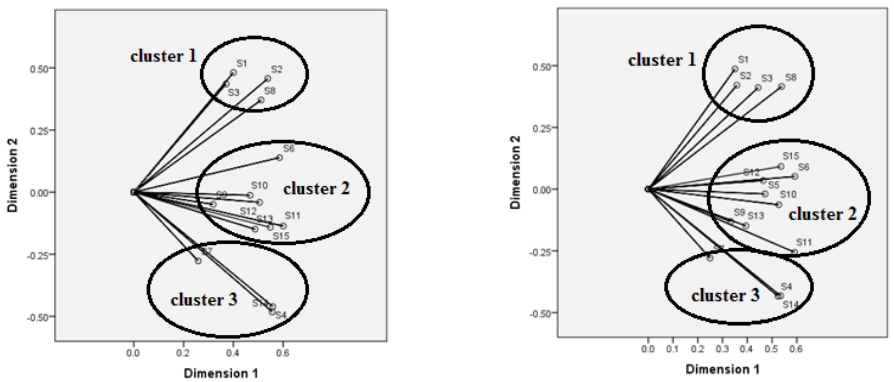


Figure 13. Position of item clusters – Test 1 of 2017 (left), Test 2 of 2017 (right).

Figure 10, Figure 11 and Figure 12 show the position of the vectors associated with different items, while Figure 13 presents the existing item clusters.

Discussion

In the first stage of the study, the items were grouped in clusters by analysing their statements.The existence of three distinct groups of items were identified: group of *Items 1, 2, 3*, and *8*; group of *Items 5, 6, 9, 11, 12, 13,* and *15*; and group of *Items 4, 7* and *14* (see Annex).

Items 1, 2, 3, and *8* are based on procedural and conceptual knowledge, which is a foundation of scientific logical reasoning. Solving them entailed the identification of a cell in a data table, calculating the difference between two quantities specified in a table, correlating the symbol of the measurement unit for length with its name, identifying some data in a graph and making a simple subtraction operation. *Item 3* was connected to *Item 1* by its statement, while *Item 8* had a solving procedure similar to *Item 2*.

The second group (Items 5, 6, 9, 10, 11, 12, 13, 15) was made up of mathematics items (6, 11), biology items (5, 9, 10, 13, 15), and one physics item (12). Item 9 required the students to match an element with the group to which it belonged. The correct answer could be identified by resorting to personal experience, coming from the observation of the surrounding world, and by using intuitive logic. Item 10 aimed at the identification of the cause generating a phenomenon. This was also a problem whose solution was given by the student's capacity of analysing a situation known from personal experience, of establishing relationships between known elements in

https://doi.org/10.33225/jbse/19.18.209

order to give an answer based on deduction. Item 13 (biology) was built similarly to Item 10, the logical deductive reasoning being used here with a view to backing a statement with arguments. Item 15 required the identification of the way in which a known cause generated a certain effect. It was a similar situation to that of Items 10 and 13, discussed above. Item 5 was problematic. The analysis of the statement revealed that the distractors were wrongly built, leading to the choice of a different variant from the one considered adequate by the creator. In the data-gathering stage it was found that most students, including those with high total scores, indicated as correct answer a different option from the one specified in the assessor's sheet. It was still needed to determine statistically, if what was noticed was proof of an erroneous item. Item 6 was a simple, even intuitive item, because it operated with elements that could be easily transposed into a concrete situation. Item 11 had a more complex structure: it contained a sequence of connected data and the solution involved a higher degree of abstraction (calculating percentage, parts out of a quantity, identifying connections between quantities). Nevertheless, exercising these abilities of mathematical calculation over a long period (starting as far back as primary school) and establishing a connection between the calculations and a concrete, tangible reality (here quantities of medicinal herbs) placed the item in the area of those whose solving required deductive scientific logical reasoning, which operated in the sphere of concrete, tangible, familiar elements. Item 12 required the drawing of a simple electrical circuit, using symbols as substitutes for reality. It was a first step towards abstract reasoning, nevertheless, connecting symbols to visible reality placed this item in an area in which reasoning operated with elements belonging to the real, known space.

About this second group of items, it is clear that they refered to an axis of concrete-oriented scientific logical reasoning. It is reasonable to administer them at this age, when the development level of reasoning is still oriented to aspects of directly perceivable reality and a lot less to the abstract.

The third group of items (*4, 7* and *14*), illustrated the way the students were able to operate with abstract concepts, such as speed, density (*Items 4* and *14*), or with elements of plane geometry (*Item 7*), and about the way in which they were able to make short demonstrations starting from a series of data, in order to arrive at a final result. In this case, one can speak of an axis of abstract-oriented scientific logical reasoning.

The analysis of Test 2 statements leads similarly to the grouping of the items into the three clusters discussed for Test 1: group of *Items 1, 2, 3*, and *8*; group of *Items 5, 6, 9, 10, 11, 12, 13*, and *15*; and group of *Items 7, 4* and *14*.

Items 1, 2, 3, and *8* had a structure similar to the items administered within Test 1, with *Item 3* having a differentiating element, though: the measurement unit utilized for length was the hectometre, much less commonly used than the metre, which appeared in *Item 3* of Test 1. In the statistical analysis it was expected to find a different perception and a lower performance of the subjects to whom this item was administered than of the subjects who solved *Item 3* of Test 1.

The second group of items, *Item 5* of Test 1 was considered problematic, which was the reason why it was decided to remove it in the subsequent stages of analysis. *Item 5* of Test 2 had an adequate structure, and it allowed for the identification of the answer by resorting to intuitive logic, without the necessity of any zoology knowledge.

Within the same group, *Item 9* of Test 2 apparently had a similar structure to *Item 9* of Test 1, but its solution involved intuition to a much lesser extent. In this case too, statistical analysis could provide information on the subjects' perception of the item and on the equivalence of this item in the two tests.

Giving a correct answer to *Item 13* of Test 2 entailed biology knowledge. Students were here expected to specify two characteristics of an element, which proved that the element belonged to a certain category, while solving *Item 13* of Test 1 was based rather on logical deductions, without the necessity of biology knowledge to solve it. Therefore, this item can be assumed as not equivalent to the one administered in Test 1 either.

As for the other items belonging to this group, no significant differences were identified at this stage.

Items 4, 7 and *14* belonging to the last group were physics and mathematics items and had an invariant character in relation to the same group within Test 1.

In the second stage of the test analysis, where the CATPCA was used, a number of two dimensions was preferred, rather than one dimension, according to the initial assumption, because the intention was to identify as accurately as possible the consistency of the scales and the way in which the dimensions load the items. In addition, it was checked if there were items referring to another dimension, outside the applied model. Analysing the coordinates of the items in relation to the two dimensions, it was found that *Item 5* had extremely low values of the coordinates in both dimensions. It was the very item that was already considered problematic in the first stage of the analysis, the current stage only bringing statistical backing for the original assumption. It was decided to remove this item from the analysis.

As can be seen in Figures 2 and 3, except for *Item 7*, whose coordinates had low values in both dimensions,

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

all the other items showed good values of the centroid coordinates in Dimension 1. The most emblematic item for Dimension 1 was *Item 11* (mathematics), followed by *6* (mathematics), *4* (physics), *2* (mathematics), *8* (physics), and by *Items 10, 13* and *15* (biology). Nevertheless, *Items 1, 2, 3, 4, 8,* and *14* had quite high values in Dimension 2 as well.

The Table 1 of factor saturation is of the utmost importance in the present study. The same as in factorial analysis, CATPCA indicates the ratio in which the latent factors/dimensions contribute to the variance of each variable.

The analysis of the factor saturation provides the statistical backing for the existence of the three clusters referred to in the previous section of the paper. The items could be separated in the group of *Items 1, 2, 3,* and *8*, with high saturations in both dimensions, the group of *Items 6, 9, 10, 11, 12, 13*, and *15*, with high saturations in the first dimension and low values of saturation in the second dimension, and the group of *Items 4, 7* and *14*, with high saturations in the first dimension and negative saturations comparable in absolute value to the saturations of the first dimension. All items showed relatively high saturations in Dimension 1, except for *Item 7*, which had the lowest saturation in this dimension and a comparable saturation in Dimension 2.

In fact, these statistical results backed the initial assumption that there was a dominant dimension, a construct which determines the variability of all the responses to the 14 test items, i.e. scientific logical reasoning, with three sides:

- conceptual and procedural knowledge;
- intuitive and deductive scientific logical reasoning, oriented to correlating elements of the surrounding world;
- inductive scientific logical reasoning, oriented to correlating abstract elements.

CATPCA also provides a graphical representation of the students' responses, in relation to the two dimensions that were assumed to exist. The aspect of the graphical representation is similar to a Gaussian distribution (Figure 4).

The cloud of points distributed in the area of very low values both in Dimension 1 and in Dimension 2 refers to students whose performance is low both in the dominant Dimension 1 and in Dimension 2 (Figure 5).

The response patterns associated to these students were as follows:

1006	0	0	0	0	0	0	0	0	0	1	0	0	1	0	1
970	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1
574	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
367	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0

There are students whose responses indicated that they had not assimilated simple procedural aspects such as those that involve solving *Items 1, 2* and *3*. The density of the existing points from this area signifies the small number of students to be found in such a situation.

The other side of the distribution showed students whose performance was high and very high, both in the dominant Dimension 1 and in the Dimension 2 (Figure 6).

Their response patterns were of the type:

669	1	0	1	1	0	1	1	1	1	1	1	1	1	1	1
164	1	1	1	1	0	1	0	0	1	1	1	1	1	1	1
599	1	1	1	1	1	0	0	1	1	1	1	1	1	1	1
099	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1

These students had a good performance both in the area of the items associated with procedural knowledge, with intuitive scientific logical reasoning, and also in the area of *Items 4, 7* and *14*, associated with abstract scientific logical reasoning. The low density of the point cloud from this area indicated a small number of students who had shown this type of behaviour.

The highest density of the point cloud was to be found for the students who had a high performance in the area of conceptual and procedural knowledge, and also in the area of intuitive scientific logical reasoning. They answered correctly to *Items 1, 2, 3* and *8*, and also to *Items 6, 9, 10, 11, 12, 13,* and *15* (Figure 7).

Their response patterns were of the type:

https://doi.org/10.33225/jbse/19.18.209

126	1	1	1	0	0	1	0	1	1	1	0	0	1	0	0
148	1	1	1	0	1	1	0	1	1	1	1	1	1	0	1
337	1	1	1	0	0	1	0	1	1	1	0	1	1	0	1
481	1	1	1	0	1	1	0	1	1	0	0	1	1	0	1

This type of students' distribution according to their responses to items did not come as a surprise, considering their level of cognitive development. It was expected from the majority of 6th-grade students to have a good performance at the first two levels of the assessed construct and a poorer one at the higher level of the construct, that of the abstract scientific logical reasoning.

The analysis of item loading in the two dimensions provided statistical backing, in the case of Test 2, for the existence of the three clusters: group of *Items 1, 2, 3,* and *8*, with high saturations for both dimensions; group of *Items 5, 6, 9, 10, 11, 12, 13*, and *15*, with high saturations in the first dimension and low values of saturation in the second dimension; and group of *Items 4* and *14*, with high saturations in the first dimension and negative saturations comparable in absolute value to the saturations of the first dimension. The first dimension saturated to a relatively high extent all the items, except for *Item 7*, which had, in this test too, a low saturation in Dimension 1 and a comparable saturation in Dimension 2 (Table 2).

The third stage of the analysis was based on the results obtained in the first two stages. The comparative analysis of the orientation of the vectors associated with each item of the two tests confirmed a series of assumptions from the first stage of the analysis regarding *Items 3, 9*, even *13*, and, furthermore, there arose a series of new elements regarding *Items 11* and *15*. Thereby, if in Test 1 of 2017 the vectors of *Item 3* and of *Item 1* had the same orientation, in Test 2 the vector of *Item 3* had a different orientation, which showed that the students' perception was different from that of the students who solved *Item 3* of Test 1 (Figure 10). As anticipated, the measurement unit hectometre had a different impact on the students from the measurement unit metre, frequently used in science classes. It was expected, from the stage of ascertaining item difficulty, to obtain different values of the characteristics of *Item 3* for the two tests, with an obvious impact on their equivalence.

Another issue concerned the group of *Items 9* and *13*. The diagram of Test 2 shows a change in the orientation of the vectors associated with these items, a shift of the vectors towards the group of *Items 4, 7* and *14*, compared to the same items from Test 1 (Figure 11). Hence, the students who took Test 2 perceived these items as having a higher degree of abstraction than *Items 9* and *13* from Test 1.

Item 11 also had a different impact on the students who took Test 2 from those who took Test 1, its vector displaying a shift to the area of *Items 4, 7* and *14* (Figure 12). There was a similarity between *Item 11* of Test 1 and *Item 11* of Test 2, concerning both the data included in the statement and the task. The difference was given by the type of quantities with which the item operated: in the first case concrete, tangible quantities (medicinal herbs), in the second abstract quantities (number of problems selected) and an activity which was not familiar to the students (teachers selecting a number of problems to be solved in a contest). The previous observation shows the importance of the stage of pre-testing the items administered in a test. Even when the test creator follows the rules of test construction, apparently insignificant differences between the statements of items considered equivalent may have an undesirable impact upon the subjects. *Item 15* was in the same situation. In Test 1 the vector was oriented more closely to the group of *Items 4, 7* and *14*, which involved the use of abstract logic, while in Test 2 the vector of *Item 15* had a shift to the group of *Items 1, 2, 3*, and *8* (Figure 12). In both cases, the task was similar: the students were asked to specify causes that generate a certain effect, but the statement that frames the task created differences between the items.

Notwithstanding the differences identified between the items of Test 1 and Test 2, from the point of view of the existing clusters and of their component parts, the two tests had similar structures (Figure 13).

Conclusions

The complementary use of the "construct map" and the CATPCA allows relatively easy exploration of the dimensionality of the items administered in an assessment test. While the "construct map" is a tool that relies heavily on the researchers' intuition, CATPCA is an objective analysis that can bring statistical support to the researchers advanced assumptions in the subjective analysis phase. Furthermore, the method also helps to assess the item disparity in distinct tests, yet assumed equivalent.

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

As far as the study of the 6th Mathematics and Natural Sciences Tests is concerned, the subjective analysis based on the "construct map" had shown that there was only one dimension underlying the construction of the tests, namely logical-scientific reasoning, with three faces :

- Conceptual and procedural knowledge;
- Intuitive and deductive logical-scientific reasoning oriented towards the relation of elements in the surrounding reality;
- Inductive logical-scientific reasoning oriented towards the relationship of abstract elements, to the relationship of concepts.

CATPCA statistically supported the hypothesis of unidimensional items. The CATPCA grouped these items into three clusters behind which lie the three faces of the identified construct. All the statistical analysis led to the identification of an item (*Item 5* of Test 1) that contained a construction error, and was removed from the analysis.

The comparative study of the statements of the items administered in the two parallel forms of the test led to the conclusion that the biology items exhibited great variability from Test 1 to Test 2, while physics and math items were much more stable. CATPCA fully supported this result by means of the comparative analysis of the vectors associated with the items in the two tests. In addition, CATPCA had highlighted the existence of disparities that had not been observed in the subjective analysis phase.

These differences between items assume that they had a negative impact on the equivalence of the two administered tests. This raises questions about fairness in testing, especially given the fact that the results a student obtains affect his future educational path.

The role of the present analysis is to enable the researcher to make a decision on the use of unidimensional or multidimensional IRT models. These patterns can be used to determine the difficulty and discrimination characteristics of the items used, thus forming a complete picture of the analysis of the tests.

References

Brown, N. J. S., & Wilson, M. (2011). A model of cognition: The missing cornerstone of assessment. *Educational Psychology Review*, *23*(2), 221–234.

Chambers, T. (2014). *Three pedagogical approaches to introductory physics labs and their effects on student learning outcomes* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3620074).

Derbentseva, N., Safayeni, F., & Canas, A. J. (2007). Concept map: Experiments on dynamic thinking. *Journal of Research in Science Teaching, 44*(3), 448-465.

DeMars, C. (2010). *Item Response Theory: Understanding statistical measurement*. Oxford, UK: Oxford University Press.

Ding, L., & Beichner, R. J. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Special Topics Physics Education Research, 5*(2), 020103.

Ding, L., Chabay, R., Sherwood, B., & Beichner, R. J. (2006). Evaluating an electricity and magnetism assessment tool: Brief electricity and magnetism assessment. *Physical Review Special Topics Physics Education Research, 2*(1), 010105.

Engelhardt, P. V., & Beichner, R. J. (2004). Students' understanding of direct current resistive electrical circuits. *American Journal of Physics, 72*(1), 98-115.

Engelhardt, P. V. (2009). An Introduction to Classical Test Theory as Applied to Conceptual Multiple-choise Tests, *Getting Started in Physics Education Research*, *2*(1), 1-40.

Eshach, H. (2014). Development of a student-centered instrument to assess middle school students' conceptual understanding of sound. *Physical Review Special Topics Physics Education Research*, *10*(1), 010102.

George, D., & Mallery, P. (2009). *SPSS for Windows step by step: A simple guide and reference*, *16.0 update* (9th ed.). Boston, MA: Pearson Education.

Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991) *Fundamentals of Item Response Theory*. Newbury Park, CA: Sage Publications.

Hambleton, R. K., & Jones, R. W. (1993). Comparison of classical test theory and item response theory and their applications to test development. *Educational Measurement: Issues and Practice, 12*(3), 38-47.

IBM Corp. (2011). IBM SPSS Statistics for Windows, (Version 20.0). [Computer software]. Armonk, NY: IBM Corp.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York, NY: Springer-Verlag New York.

Liu, Y., & Maydeu-Olivares, A. (2012). Local Dependence Diagnostics in IRT Modeling of Binary Data. *Educational and Psychological Measurement*, *73*(2), 254-274.

National Center for Assessment and Examination. (2017). *Raport Național EN VI 2017 Matematică și Științe – Analiza rezultatelor Evaluării Naționale la finalul clasei a VI-a. [National Report EN VI 2017 Mathematics and Sciences - The Analysis of the National Evaluation Results at the end of 6th grade].* Retrieved from http://rocnee.eu/sites/default/files/2017-12/Raport__ENVI_2017_Mate_stiinte.pdf

National Research Council. (2000*). How people learn: Brain, mind, experience, and school: Expanded Edition*. Washington, DC: The National Academies Press.

https://doi.org/10.33225/jbse/19.18.209

Novak, J. D., & Canãs, A. J. (2007). Theoretical origins of concept maps, how to construct them, and uses in education. *Reflecting Education*, *3*(1), 29-42.

Opariuc-Dan, C. (2012). Analiza Componentelor Principale pentru Date Categoriale (CATPCA) [Categorical Principal Components Analysis (CATPCA)]. *Psihologia Resurselor Umane 10*(2), 103-117.

Opariuc-Dan, C. (2014). Utilizarea hărţii constructelor în demersul de elaborare a probelor psihologice [Using the construct map in the endeavor to elaborate psychological tests]. *Psihologia Resurselor Umane*, *12*(2*)*, 186-194.

Önder, F. (2016). Development and validation of the photoelectric effect concept inventory. *European Journal of Physics, 37*(5), 055709.

Osterlind, J. S. (1998). *Constructing test items: Multiple-choice, constructed-response, performance, and other formats*. Boston, MA: Kluwer Academic Publishers.

Sabella, M. S., & Redish, E. F. (2007). Knowledge organization and activation in physics problem solving. *American Journal of Physics, 75*(11), 1017–1029.

Sava, F. A. (2011). *Analiza datelor în cercetarea psihologică* [*Data* Analysis in Psychological Research] (2nd ed.). Cluj-Napoca, Romania: Editura ASCR.

Wang, J., & Bao, L. (2010). Analyzing force concept inventory with item response theory. *American Journal of Physics*, *78*(10), 1064–1070.

Wilson, M. (2005). *Constructing measures: An item response modeling approach*. New Jersey, US: Lawrence Erlbaum Associates Publishers.

Xitao, F. (1998). Item response theory and classical test theory. An empirical comparison of their item/person statistics. *Educational and Psychological Measurement, 58*(3), 357-381.

Annex

Item statements Test 1 (2017)

To solve tasks 1-5, read the following text:

Looking for information about the natural reserves they could visit, the students discovered that there is a lily-of-the-valley reserve in the area, as well as the Comana Adventure Park. The table below gives information on the peony, butcher's broom and lily-of-the-valley reserves, also about Balta Comana.

Characteristics \ Reserve	Peony Reserve	Butcher's Broom Reserve	Lily-of-the-valley Reserve	Balta Comana
Surface area (ha)	232	250	164	1206
Average altitude (m)	80	75	60	40

Code 1 0 9

1. *Encircle the letter corresponding to the correct answer.*
According to the information given in the table, the average altitude of the ground where the **Peony Reserve** lies equals:

a) 40 m
b) 60 m
c) 80 m
d) 232 m

Code 1 0 9

2. *Encircle the letter corresponding to the correct answer.*
According to the information given in the table, the surface area of the **Butcher's Broom Reserve** is bigger than the surface area of the **Lily-of-the-valley Reserve** by:

a) 15 ha
b) 68 ha
c) 86 ha
d) 164 ha

Item statements Test 2 (2017)

To solve tasks 1-5, read the following text:

Looking for information about the natural reserves they could visit, the students discovered that there are eleven natural reserves in the Domogled-Valea Cernei National Park. The table below gives information about the surface area of four of them, which the students decided to visit, as well as the length of the route taken when visiting each reserve.

Characteristics \ Reserve	Vârful lui Stan	Valea Ţesna	Ciucevele Cernei	Domogled Reserve
Surface area (ha)	232	250	164	1206
Length of route covered by students (hm)	80	75	60	40

Code 1 0 9

1. *Encircle the letter corresponding to the correct answer.*
According to the information given in the table, the length of the route covered by the students when visiting the **Vârful lui Stan Reserve** is:

a) 50 hm
b) 60 hm
c) 80 hm
d) 120 hm

Code 1 0 9

2. *Encircle the letter corresponding to the correct answer.*
According to the information given in the table, the surface area of the **Ciucevele Cernei Reserve** is bigger than the surface area of the **Valea Ţesna Reserve** by:

a) 30 ha
b) 40 ha
c) 1006 ha
d) 1046 ha

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Code 1 0 9

3. *Encircle the letter corresponding to the correct answer.*
The measurement unit for average altitude, as given in the table, is:

a) millimetre
b) metre
c) hectare
d) tape

Code 21 11 12 13 00 01 99

4. A boat covered a distance of 1.5km on the waters of **Balta Comana**, at an average speed of 0.5m/s. Calculate the duration of covering the distance. Give the result in minutes.

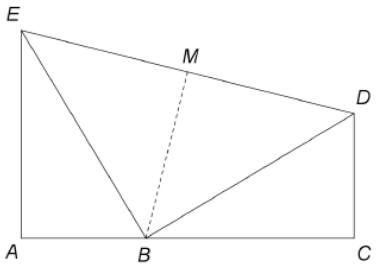
Code 1 0 9

5. There is an area of secular trees in the **Comana Natural Park**. While visiting the park, the 6th-grade students Radu and Mihaela noticed, on the trunk of an oak tree, several invertebrates: ladybirds, spiders, ants, a capricorn beetle. They took pictures and even managed to videotape these invertebrates.
Encircle the letter corresponding to the correct answer.
The correct statement regarding invertebrate animals is:

a) the frog breathes through skin and lungs
b) the ladybird has a pair of chitinous elytra
c) the cockchafer has four membranous wings
d) the spider's body consists of head, thorax and abdomen

To solve tasks 6-10, read the following text:

Once in the Comana National Park, the students were given a map of tourist routes, with stopovers, bird watching points, and also information on that week's weather.
The stopovers on the tourist route covered by the students are represented in the given drawing by points A, B, C, D and E. In the drawing, points A, B and C are collinear, <BAE and <BCD are right angles, AB = CD = 3cm, and BC = AE = 5cm.



Code 2 1 0 9

6. Calculate the length of the AC segment.

***Code** 21 11 12 13 00 01 99*

7. Consider point M midpoint of segment DE. Determine the DBM measure angle.

Code 1 0 9

3. *Encircle the letter corresponding to the correct answer.*
The measurement unit for the length of the route covered by the students, as given in the table, is:

a) hectare
b) metre
c) tape
d) hectometre

Code 21 11 12 13 00 01 99

4. The group of students who visited the **Domogled Reserve** covered the route with a length of 6km at an average speed of 0.4m/s. Calculate the duration of covering this distance. Give the result in minutes.

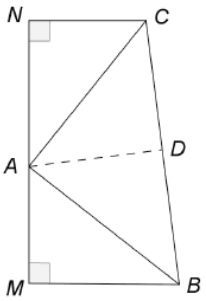
Code 1 0 9

5. The first research data on the fauna (insects, birds) from the area of the **Domogled-Valea Cernei National Park** appeared as early as 1794. Studies showed that over 45% of the total number of butterfly species existing in the country (over 1500 species) live in this park.
Encircle the letter corresponding to the correct answer.
One of butterflies' adaptations to flight is:

a) development by metamorphosis (egg-caterpillar-nymph-adult)
b) body differentiation into head, thorax and abdomen
c) presence of wings, covered by fine scales
d) presence of three pairs of thin legs

To solve tasks 6-10, read the following text:

The educational programs organized by the Administration of the Domogled-Valea Cernei National Park include trips. The students were given a map of the tourist routes, with stopovers and observation points for protected species.
The stopovers on the tourist route covered by the students are represented in the given drawing by points A, B, C, M, and N. In the drawing, points M, A and N are collinear, <AMB and <ANC are right angles, AM = NC = 4cm, and AN = MB = 5cm.



Code 2 1 0 9

6. Calculate the length of the MN segment.

Code 21 11 12 13 00 01 99

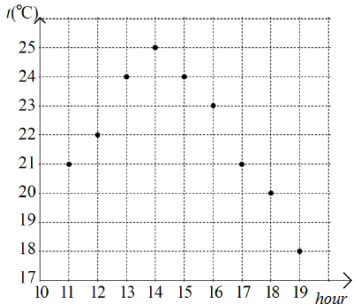
7. Consider point D on side BC so that AD ⊥ BC. Determine the BAD measure angle.

https://doi.org/10.33225/jbse/19.18.209

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Code 1 0 9

8. During their park trip, the students recorded in a table the values of the temperature indicated by a thermometer, between 11:00 and 19:00. The measurements were carried out hourly. The data gathered were afterwards represented in the diagram below.

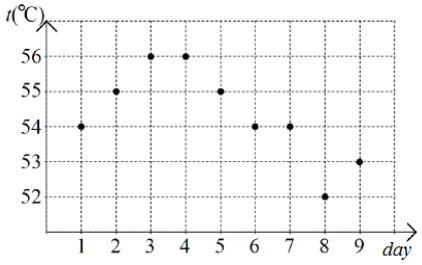


Encircle the letter corresponding to the correct answer.
The temperature measured at 19:00 was lower than the temperature measured at 15:00 by:

a) 18°C
b) 6°C
c) 4°C
d) 1.3°C

Code 1 0 9

8. In the Domogled-Valea Cernei National Park is located the Steam Grotto. It is a thermal cave out of whose cracks comes out steam. The temperature of the cave water was measured for 9 consecutive days, at the same time of each day. The data thus gathered were represented in the diagram below.



Encircle the letter corresponding to the correct answer.
The temperature recorded on the 9th measurement day was lower than the temperature recorded on the 5th day by:

a) 53°C
b) 4°C
c) 2°C
d) 0.9°C

Code 21 11 12 00 99

9. Guided by their biology teacher, the students made observations on some birds seen in the park. Match each example of bird from Column A with the bird group to which it belongs, from Column B. Write the corresponding letter in the space before each figure in Column A. One of the bird groups allows no matching.

Code 21 11 12 00 99

9. Birds make themselves felt by their songs and by their coloured plumage. The students could observe in the park birds such as the wood grouse, the peregrine falcon, the golden eagle, the grey-headed woodpecker, etc. Match each group of birds from Column A with one of its characteristics from Column B. Write the corresponding letter in the dotted space before each figure of Column A. One of the characteristics allows no matching.

Code 2 1 0 9

10. At the request of the park administration, a team of researchers carried out a study on the water animals from Balta Comana. They found that, in the last five years only, the number of pond turtles has decreased by half. Indicate two possible causes of this phenomenon.

Code 2 1 0 9

10. On the territory of the Domogled-Valea Cernei Park there can be found Adam's Cave. Large colonies of bats live here, as well as some invertebrates. Write two arguments in favour of the idea that intense visiting of the cave by groups of tourists could affect the life of the animals living here.

Code 21 11 12 13 00 01 99

11. Medicinal herbs are cultivated on a farm near the park and the total harvest is 160kg of medicinal herbs. Camomile comes to 60% of the harvest, pot marigold is half of the rest of the harvest, and what remains is buckthorn. Calculate the quantity of buckthorn harvested.

Code 21 11 12 13 00 01 99

11. One day with no trip scheduled, the students took part in a mathematics competition. 120 problems were proposed and the teachers selected a number of them for the final phase. The selection was carried out in two stages: in the first stage, 70% of the proposed problems were excluded, and in the second stage a quarter of the problems left were selected. Calculate the number of problems selected for the final stage.

Code 21 11 12 00 99

12. In order to ensure good lighting for the workbench in the woodcraft workshop, they use a light bulb fed by a source of electric power. Using the symbols of circuit elements, draw the chart of an electric circuit made up of a battery, a light bulb, connecting wires and a switch.

Code 21 11 12 00 99

12. In order to ensure night lighting in the area around a chalet, they use a light bulb fed by a power generator. Using the symbols of circuit elements, draw the chart of an electric circuit made up of a power generator, a light bulb, connecting wires and a switch.

Code 2 1 0 9

13. The area of Balta Comana, located on the lower reach of the river Neajlov, is of the utmost importance for the preservation of biological diversity.
Indicate two arguments in favour of the idea that this area can be considered a "bird paradise".

Code 2 1 0 9

13. In the area of the park, near courses of water, in brooks and mountain tarns, there can be observed different species of amphibians: the crested newt, the salamander, the agile frog, the tree frog.
Indicate two characteristics of the salamander proving the fact that it is an amphibian, not a reptile.

Code 21 11 12 13 00 01 99

14. In the woodcraft workshop there can be manufactured wooden birdhouses and wooden toys for children. For one of the toys, there was used a piece of wood with the volume of 600cm^3 and density of 750g/dm^3. Calculate the mass for the piece of wood. Express the result in kilograms.

Code 21 11 12 13 00 01 99

14. According to the Regulations of the Domogled-Valea Cernei National Park, within the park all constructions must observe local architectural traditions, being built of traditional materials, such as stone or wood. For one of the buildings there were used beams of dry wood with a density of 0.8g/cm^3. Calculate the mass of a wood beam with a volume of 0.15m^3. Express the result in kilograms.

https://doi.org/10.33225/jbse/19.18.209

ITEM DIMENSIONALITY EXPLORATION BY MEANS OF CONSTRUCT MAP AND CATEGORICAL PRINCIPAL COMPONENTS ANALYSIS
(P. 209-226)

ISSN 1648–3898 /Print/
ISSN 2538–7138 /Online/

Code 2 1 0 9	***Code 2 1 0 9***
15. In the period 2009-2011 there was carried out a project of ecological rehabilitation of Balta Comana. The project aimed at the preservation of biological diversity, but also at informing visitors and increasing awareness of the importance of protecting the environment in general and the water habitats in particular. Write two examples of the way in which plastic bottles thrown into the water can endanger the life of aquatic organisms.	15. The reptile species from the Domogled-Valea Cernei National Park, such as the common tortoise, the Aesculapian snake, the horned viper are increasingly seldom observed in the park, as a result of a decrease in their number. That is the reason why these species are protected by law. Write two possible causes of the decrease in the number of individuals belonging to these species.

Received: *January 20, 2019* Accepted: *March 25, 2019*

Gabriela Deliu	1PhD Student at University of Bucharest, Faculty of Physics, Doctoral School Department, Atomistilor Street 405, Bucharest-Magurele, Romania; 2Teacher of Physics at "Emil Racoviță" Highschool Brașov, Armoniei Street 6, Brasov, Romania. E-mail: gabinan_bv@yahoo.com
Cristina Miron	PhD Assoc. Prof. at University of Bucharest Romania, Faculty of Physics, Department of Matter Structure, Atmospheric and Earth Physics and Astrophysics, Atomistilor Street 405, Bucharest-Magurele, Romania. E-mail: cmiron_2001@yahoo.com
Cristian Opariuc-Dan	PhD Lecturer at "Ovidius" University, Faculty of Law and Administrative Sciences, Department Administrative Sciences, Mamaia Avenue 124, Constanta, Romania. E-mail: copariuc@gmail.com

https://doi.org/10.33225/jbse/19.18.209