

MATHEMATICS AND SCIENCE ASSESSMENT FOR 12 YEAR OLD STUDENTS – THE ROMANIAN EXPERIENCE: INCREASING THE EFFECTIVENESS OF THE TEACHING-LEARNING ASSESSMENT PROCESS USING INFORMATION AND COMMUNICATIONS TECHNOLOGY

Gabriela Streinu-Cercel & Bogdan Cristescu

National Assessment and Evaluation Centre, Bucharest (Romania)

gabriela.streinucercel@rocnee.eu, bogdan.cristescu@rocnee.eu

To facilitate interdisciplinary standardized assessment, we implemented a national assessment program for 12-13 year-old students in Romania, using tests combining notions from mathematics and science. These tests are evaluated based on clustered codes. Each code presented in the encoding pattern is associated with a descriptor and is followed by a list of examples. The encoding pattern has been computed into an electronic application that lists each student's strong points and weak points in relation with the disciplinary competencies (mathematics, physics and biology) and the six interdisciplinary competencies that we specified. Using the list of strong and weak points expressed in action verbs, and detailed facts for each content assessed, teachers can develop individualized learning plans for their students.

BACKGROUND

We present a national assessment program that offers standardized information regarding students' knowledge and competencies and specific feedback for the teaching – learning – assessing process. To increase the effectiveness of this complex process, we have developed and implemented an interdisciplinary tool that prospectively identifies the areas where individual or collective adjustment is warranted.

Since 2014, in Romania we are employing a national assessment strategy (External National Assessment) for 2nd graders (8 – 9 years old), 4th graders (10 – 11 years old) and 6th graders (12 – 13 years old), complementing the national disciplinary exams already in place. We present an assessment of the 6th graders' ability to operate with and interconnect the content of mathematics and science (physics and biology) that they have already learnt. We are currently (2015) in the second year of applying this strategy; approximately 200,000 students were involved each year in this nationally-applied strategy for personalizing the teaching – learning – assessing process.

The results of this External National Assessment are available to teachers of mathematics, physics and biology, to students and their parents. Using the computer application developed, each student's results for each item are analyzed and an interpretation is provided on individual, class, school, regional, and national level. Teachers use the test's output to pinpoint the individual learning needs of students, to set learning objectives, adjust teaching strategies, initiate potential remedial activities, and plan learning activities.

The novelty behind these tests relies in their ability to provide a diagnostic assessment as well as a formative one. Through the use of these tools, teachers can identify particular skills, attitudes and aptitudes for each student, and make early decisions about future learning

opportunities, thus providing personalized teaching assistance based on individualized learning programs.

CONSTRUCTING THE TEST

The tests were elaborated in several steps: devising the interdisciplinary competencies, projecting the specifications matrix, designing the test, elaborating the specific items and the encoding pattern (Streinu-Cerel & Cristescu, 2014). The tests were designed in accordance with the quality cycle: plan-do-check-act.

Six interdisciplinary competencies (IC) were derived from the general and specific competencies of the school’s curricula in mathematics, physics and biology. We assess the following interdisciplinary competencies (Streinu-Cerel & Cristescu, 2014):

IC1. Identifying data, concepts, specific relations of mathematics/science in an interdisciplinary context

IC2. Processing the following types of data: quantitative, qualitative, specific structural mathematics/science contained in various data sources

IC3. Using concepts, algorithms and procedures of mathematics/science to locally or globally characterize a particular case

IC4. Expressing the quantitative or qualitative characteristics of a particular situation in the specific language of mathematics/science

IC5. Analyzing the characteristics of relationships, phenomena or processes specific to mathematics/science, based on real or hypothetical situations

IC6. Interpretation of problem-situations specific to mathematics/science by integrating knowledge from different fields

Competencies are stated in action terms – Bloom’s cognitive levels (Bloom, 1994) –, as they cross-link stages of the teaching – learning – assessing process, categories of skills, and informational feedback, on skills and cognitive levels. In Romania, the mathematics curricula are designed on six competencies associated with Bloom’s cognitive levels. Subsequently, the national evaluation in mathematics is based on six assessment competencies also associated with Bloom’s cognitive levels. Taking into account the particularities of this national assessment strategy, cognitive levels of the interdisciplinary competencies are those from the school’s curricula in mathematics (Streinu-Cerel & Cristescu, 2014).

Each test contains a total of 15 items of mathematics, physics and biology, introduced in a general context and three specific contexts, as shown in Table 1. The items are not specific to a particular discipline (for example, items of physics include mathematical calculus), but roughly there are five items for each discipline (CNEE, 2012).

General context					
Specific context 1		Specific context 2		Specific context 3	
Item 1.	Mathematics	Item 6.	Mathematics	Item 11.	Mathematics
Item 2.	Mathematics	Item 7.	Mathematics	Item 12.	Physics
Item 3.	Physics	Item 8.	Physics	Item 13.	Biology
Item 4.	Physics	Item 9.	Biology	Item 14.	Physics

Item 5.	Biology	Item 10.	Biology	Item 15.	Biology
---------	---------	----------	---------	----------	---------

Table 1: Test design

CALIBRATING THE TEST

In order to obtain accurate information at different levels (student, class, school, regional and national) the tests used in this national assessment program were calibrated using the Item Response Theory, which adopts explicit models for the probability of each possible response to each item. Through IRT the probability of each possible response is derived as a function of ability and some item parameters (Hambleton & Swaminathan, 2000). We have used IRT to obtain the likelihood of ability as a function of the actually observed responses as well as item parameters. The ability value that has the highest likelihood becomes the estimated ability (Partchev, 2004). The IRT model has to be true (correct) and the item parameters known. Calibration studies have been performed in 2012, 2013, and 2014 by giving the items to a representative number of tested students and computing their responses to estimate the item parameters.

We used the Three Parameter Logistic (3PL) model (Partchev, 2004) to correlate students' abilities in mathematics, physics and biology with their test performance. The item i is characterized by the parameters a_i , b_i and c_i and the probability of a random student having the ability θ to respond correctly to item i is given by:

$$P_i(\theta, a_i, b_i, c_i) = c_i + (1 - c_i) \frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}},$$

where a_i is the item discrimination parameter, b_i is the item difficulty parameter and c_i is the probability of a correct response when true ability approaches $-\infty$ (Streinu-Cercel & Cristescu, 2014). Then, for each test we chose samples that were statistically identical (up to negligible random variation). The items of each test were applied on those samples and on data collected this way, both with an initial estimate of the ability level, were used to determine the probability of correct answer. The probability of correct answer was plotted depending on ability level and we determined the item's parameter values.

For each test, the likelihood function has been calculated based on the values already determined for the parameters of the items:

$$L(\theta) = \prod_{i=1}^n P_i^{u_i}(\theta, a_i, b_i, c_i) (1 - P_i(\theta, a_i, b_i, c_i))^{1 - u_i},$$

where $u_i \in [0, 1]$ is the score on item i ; $u_i = 0$ if the student answered incorrectly at item i and $u_i = 1$ if the student answered correctly at item i . For each student in the sample, the ability was set at the particular value that maximizes the likelihood function. We take $\theta^* = l\theta + k$, where l and k are constants, and we obtain a normal distribution of abilities for the students

(mean of 0 and standard deviation of 1). The 3PL model can thus be adjusted to accommodate the linear transformation of ability by taking $a_i^* = \frac{a_i}{l}$, $b_i^* = lb_i + k$ and $c_i^* = c_i$. Since,

$$P_i(\theta^*) = P_i(\theta),$$

the probability of a correct response is invariant to these transformations. For these tests, the difficulty parameter b_i^* varied between -1.8 and $+1.3$.

For each test, we obtained the test characteristic curve by plotting the probability that a student with the ability θ will obtain a certain score to the test, using,

$$\xi(\theta) = \sum_{i=1}^n P_i(\theta)$$

(Streinu-Cercel & Cristescu, 2014). The tests were constructed to provide virtually identical curves, independent of item-specific context.

ENCODING PATTERN - COMPUTER APPLICATION ASSOCIATED

This national assessment program is based on a novel process for result interpretation. The tests are not graded, but are instead evaluated based on clustered codes. Thus, students receive individual feedback (translated into personalized learning plans) but clustered feedback is also provided at class, school, regional and national level, and each teacher can generate personalized teaching plans (Streinu-Cercel & Cristescu, 2014).

Each score in the encoding pattern has an associated descriptor. To ensure that the teachers (mathematics, physics and biology) that assess each test select the most appropriate evaluation code, we provide detailed examples.

We present as an example an item and its encoding pattern from one of the assessor's brochure:

One diorama exhibits 60 birds, mammals and reptiles.

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

The number of birds represents 30% of the number of exhibits, and the number of mammals is equal to the number of reptiles. Determine the number of reptiles presented in this diorama.

Total score

Code 21: complete and correct reasoning and solving. Correct answer: 21 reptiles

Examples:

- $\frac{30}{100} \cdot 60 = 18$ birds

$60 - 18 = 42$ and $\frac{1}{2} \cdot 42 = 21$, thus there are 21 reptiles

- $\frac{1}{2} \cdot \left(60 - \frac{30}{100} \cdot 60 \right) = 21$ reptiles

- $100\% - 30\% = 70\%$, $\frac{1}{2} \cdot 70\% = 35\%$, that means that there are $\frac{35}{100} \cdot 60 = 21$ reptiles

Etc.

Partial score

Code 11: partially correct reasoning, calculations correct but incomplete

Examples:

- $\frac{30}{100} \cdot 60 = 18$ birds, $60 - 18 = 42$
- $30\% \cdot 60 = 18$ birds

Etc.

Code 12: partially correct reasoning, calculation errors

Examples:

- $\frac{1}{2} \cdot \left(60 - \frac{30}{100} \cdot 60 \right) = \frac{1}{2} \cdot 60 - \frac{30}{100} \cdot 60 = 12$ reptiles
- $\frac{30}{100} \cdot 60 = 20$ birds, $60 - 20 = 40$

Etc.

Code 13: correct answer without justification. 21 reptiles

Zero score

Code 00: incomplete reasoning (correct statements) but not specific enough

Example:

- We can calculate the number of reptiles by subtracting the number of mammals and birds from the total number of exhibits

Code 01: other responses

Code 99: no answer

In Table 2 we present the encoding pattern for the item above that requires basic knowledge of percentage, ratio and proportion as well as basic calculus, with the correspondences among the codes associated to the item, the descriptors for each code and the positive/weak remarks that are used to generate personalized teaching plans.

Code	Descriptor	Positive remarks	Weak remarks
21	- complete and correct reasoning and solving the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus	- usage of percentage, ratio or proportion as well as basic calculus for correct and complete solving of the task	
11	- partially correct reasoning, calculations are correct but incomplete for the task that	- partially correct reasoning for the task that requires basic knowledge of percentage, ratio and proportion as	- the usage of percentage, ratio or proportion wasn't accurate enough to solve

	requires basic knowledge of percentage, ratio and proportion as well as basic calculus	well as basic calculus - correct usage of percentage, ratio or proportion, correct but incomplete calculations	the task
12	- partially correct reasoning, calculation errors for the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus	- partially correct reasoning for the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus	- calculation errors
13	- correct answer without justification for the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus	- correct answer without justification for the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus	- didn't present any steps in reasoning or calculation leading to the correct result for the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus
00	- incomplete reasoning (correct statements) but not specific enough for solving the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus	- correct general statements about percentage, ratio or proportion, but not specific enough for solving the task	- incomplete reasoning; the correct statements written about percentage, ratio or proportion weren't applied to the actual data of the problem
01	- didn't write correct statements about percentage, ratio or proportion		- didn't write correct statements about percentage, ratio or proportion
99	- no answer, no proof of attempting to solve the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus		- didn't show any attempts to solve the task that requires basic knowledge of percentage, ratio and proportion as well as basic calculus

Table 2: Encoding pattern for item no. 11: codes, descriptors, positive remarks, weak remarks

A computer application has been designed to extract from this encoding pattern strong points and weak points for each student in relation with each of the six interdisciplinary competencies, expressed in action verbs, as well as detailed facts for each content assessed. Based on this output, teachers (mathematics and sciences) can design an individualized learning plan for each student (Streinu-Cercel & Cristescu, 2014).

Each student receives a feedback sheet that contains two parts: the first part contains the codes designated after reviewing of her/his paper and the second part contains her/his strong points and weak points.

Strong points (+)	- identification in a table/diagram of quantitative data specific for Math & Sciences (Item 1, Item 12) - processing (comparison and/or calculation using integer numbers) of quantitative data specific for Math & Sciences (Item 2, Item 12)
----------------------	---

	<ul style="list-style-type: none"> - identification of measurements in physics with the same unit (Item 3) - representation of an electrical circuit (Item 4) - association of life forms with their environment (Item 5) - adequate reasoning using information from the text in order to calculate the perimeter of a triangle (Item 6) - using of parallelism properties, properties of rectangle, isosceles and/or equilateral triangle together and/or properties of symmetry to deduce relations and to prove congruence of two triangles (Item 7) - calculations with rational numbers (Item 8, Item 14) - correct reasoning using the relation between the distance and the time of movement (Item 8) or between mass and density (Item 14) - identification of amphibians' adaptations (types of movement) to the environment (Item 9) - identification of type of breathing for four groups of life forms (Item 10) - adequate reasoning, analyzing data and usage of percentage, ratio or proportion as well as basic calculations for correct and complete solving of the task (Item 11)
Weak points (-)	<ul style="list-style-type: none"> - errors in analyzing the correct operating of an electrical circuit (Item 4) - errors in transformations of measurement units (Item 8, Item 14) - didn't prove the knowledge of evolutionary characteristics of studied life forms (Item 13) - indicates a living organism with an impact on environment without explaining its role (Item 15)

Table 3: Example of Part 2 of one feedback sheet regarding strong points and weak points

The computer application is user friendly. The list of possible codes is predefined and checkboxes are provided; thus, the teacher only needs to tick the codes for each student. The computer application automatically generates the list of strong/weak points. Once all the data regarding student's answers have been recorded, the application permits many types of data interpretation at different levels: student, class, teacher, school, regional and national. One of the important things about data analysis is that the application also enables extraction of data specific to any content and each competency. For example, if we are interested in the *IC.1 Identifying data, concepts, specific relations of mathematics/science in an interdisciplinary context* we can see if a student is able to identify data that he/she can use in mathematics tasks but he/she is not able to identify data that can be used in biology tasks; in this case, the area where further work is needed is not at the competencies level but at disciplinary level. On the other hand, if one student has issues on all contents with *IC.5 Analyzing the characteristics of relationships, phenomena or processes specific to mathematics/science, based on real or hypothetical situations*, the approach should be multidisciplinary.

The data that the computer application provides can be used at teacher's level if most of the students of one teacher are showing the same difficulties, at school level for deciding the

strategy for teachers' professional development, and at regional and national level in order to implement educational policies.

Using the test characteristic curve and the distribution of test scores, we can also determine the students' ability in mathematics and science and compute their distribution among the school population.

The report summarizing the results of national assessments applied in 2014 at the end of the 2nd, 4th and 6th grades was published in 2015 (CNEE, 2015). The feedback from Mathematics and Sciences teachers from Romania was that the procedure is labor-intensive (reviewing tests, putting together individualized learning plans, remedial worksheets, etc.), but that, when correctly implemented, it can have a substantial impact on increasing the performance in students.

CONCLUSIONS

We have developed a state-of-the-art interdisciplinary tool that, associated with a computer application, can provide individual feedback for each student, as well as clustered feedback at class, school, regional and national level. By using clustered codes instead of classical grading systems, we are now able to provide personalized learning and teaching plans. This national assessment strategy can thus identify the areas where adjustment is needed to increase the effectiveness of the teaching – learning – assessing process.

This national assessment strategy has the potential to generate a database for remedial teaching-learning-assessment activities, as well as actions recommended specifically for each type of performance descriptor from the encoding pattern.

Given the formative nature of this assessment strategy, we expect that students who were evaluated in 2014 at the end of their 6th grade through this strategy will have better results at the end of the secondary school (8th grade), when they will pass their certificate assessment. These results will become available in 2016, and a statistical analysis will be performed at that point.

References

- Bloom, B. S. (1994). *Bloom's taxonomy: A forty-year retrospective*. In K. J. Rehage et al. (Eds.), *Yearbook of the national society for the study of education*, 93(2). Chicago, IL: National Society for the Study of Education.
- CNEE (2012). Ghid de elaborare a subiectelor pentru evaluările prin probe transdisciplinare, la finalul clasei a VI-a, www.rocnee.eu/Files/Ghid_evaluare_clasa_a_VI-a.pdf
- CNEE (2015). Evaluările Naționale la finalul claselor a II-a, a IV-a și a VI-a. Raportul administrării 2014, www.rocnee.eu/index100.htm, (pp. 173-211), Anexa 3B, 535-565.
- Hambleton, R. K., & Swaminathan, H. (2000). *Item response theory – Principles and applications*. Boston: Kluwer Nijhoff Publishing.
- Partchev I. (2004). *A visual guide to item response theory*. Friedrich-Schiller Universität Jena.
- Streinu-Cercel G., & Cristescu B. (2014) Interdisciplinary (Mathematics, Physics and Biology) national assessment in Romania – personalized teaching and learning plans. In Union of Mathematicians of Macedonia (Eds.), *Proc. of the 5th. Congress of Mathematicians of Macedonia* (Section: History and Education of Mathematics and Informatics, pp. 98-107). Ohrid, Republic of Macedonia.