

MATHEMATICS TESTS IN PISA 2003

© Tünde KÁNTOR
(University of Debrecen)

tkantor@math.klte.hu

The aim of this paper is to investigate the role of mathematics tests in the assessment of mathematical literacy in PISA 2003. The article analyses the different types of tests, their advantages and disadvantages, their characteristics, methods of solving mathematics tests and some experiences in solving tests.

Keywords: achievement, control, assessing students, performance measurement of competences

About the PISA assessment

PISA presents students with problems set in real-world situations. The goal of the PISA assessment is to obtain results about how the students can apply their mathematical knowledge and competencies to solve such kind of problems successfully. In traditional mathematics education the students learn arithmetical techniques, completing an arithmetic computation and solve different types of equations, equation-systems, inequalities, geometric properties and relationships, combinatorics, probability and statistics, some relevant concepts, skills and methods of proving theorems.

How is mathematics measured in PISA? - PISA measures mathematics performance in three dimensions:

- Mathematical content
- The processes involved
- The situations in which problems are posed.

Mathematical content. The different problems and questions are related to the following four areas:

1. Space and shape
2. Change and relationships
3. Quantity
4. Uncertainty.

Traditionally this content covers the knowledge of algebra, arithmetic, geometry, functions and probability.

Process. The processes involved need to connect observed phenomena with mathematics. Beginning with a real world problem, students must organise it according to mathematical concepts. They have to transform it into a mathematical form by making simplifying assumptions. They have to identify the relevant mathematical concepts, then to perform mathematical operations, to make a mathematical model, and retranslate the result into the original problem, into the real world. In this process various *competencies* are required:

- Thinking and reasoning
- Argumentation
- Communication
- Modelling
- Problem posing and solving
- Representation
- Using symbolic, formal and technical language and operations.

PISA distinguishes *three competency clusters*:

- Reproduction cluster
- Connections cluster
- Reflection cluster.

The items of the *reproduction clusters* are relatively familiar and essentially require the reproduction of practised knowledge, such as knowledge of facts and common problem representations, recognition of equivalents, recollection of familiar mathematical objects and properties, performance of routine procedures, application of standard algorithms and technical skills, manipulation of expressions containing symbols and formulae in familiar and standard form, and carrying out straight-forward computations.

The connection cluster builds on reproduction to solve problems that are not simply a routine, but they still involve somewhat familiar settings or extend and develop beyond the familiar to a relatively minor degree. Problems typically involve greater interpretation demands, and require making links between different representations of the situation, or linking different aspects of the problem situation in order to develop a solution.

The problems addressed using the competencies in *the reflection cluster* involve more elements than the others, and additional demands typically arise for students to generalise and to explain their results.

Situation. There were four sorts of situations:

- *Personal situations*, which directly relate to students' personal daily activities
- *Educational or occupational situations* which appear in a student's life at school, or in a work situation
- *Public situations* relating to the local and broader community require students to observe some aspects of their broader surroundings.
- *Scientific situations* are more abstract and might involve understanding a technological process, a theoretical situation or explicit mathematical problem ("intra- mathematical" context)

The PISA assessment items were constructed to cover these different dimensions.

About the tests, the types of the tests

A test is a systematic procedure for comparing two or more persons. It has many different types:

- Intelligence test
- Assessment test
- Competence test
- Neuropsychological test

- Behavioural test
- Educational test
- Mathematics Competitions

In mathematics tests there are a *variety of forms*:

1. *Alternative choice* (false or true)
2. *Multiple choice*
 - a) One right answer (for example from five answers)
 - b) More right answers
 - c) Searching for the best answer
 - d) Searching for the wrong answer
3. *Connection of answers and facts*
4. *Choice of response*
 - a) Open constructed response
 - b) Short response
 - c) Long response
 - d) Essays

Standardized tests Standardized test is a test administered and scored in a standard manner. These tests are designed so that the questions, the conditions for scoring procedures are predetermined. There are two types of standardized tests:

- Norm-referenced tests, which are associated with traditional education and knowledge
- Criterion-referenced tests, which are associated with competencies

Advantages and disadvantages of the tests. A well designed, standardized test provides an assessment of students' knowledge, skills or competencies, it gives useful information.

We have to mention the problems of students' and teachers' behaviour to the tests. Many students make occasionally guessing. They choose their answer by chance. They mark their answers on the answer sheet without thinking, without knowing anything about the content, using the trial and error method. To the students' behaviour we have to pay attention in the evaluation of the tests. It is better to choose the multiple-choice form with five answers, it is not so noisy and it does not depend on the luck of the student.

For the teachers a test is a model. They teach the subjects as these subjects are tested. We have to know that tests are not valid near the bottom of the scores domain.

Characteristics of the tests

- Reliability
- Validity
- Effectiveness.

Mathematics tests in Hungary

In the years of the PISA tests (2000, 2003) in Hungary at schools tests were not used for assessment in mathematics. In other subjects (foreign languages or biology) the teachers used that form, so for the students tests were not unknown, but they were not expert in filling and solving mathematics tests. In the subject mathematics only a few students with high ability met with the tests on some competitions: American Mathematical Competitions

(AHSME, AIME, AMC 10, AMC 12 grade 9-12), Zrínyi Competitions (grade 1-8), Kangaroo Competitions (grade 9-12), Gordiusz Competitions (grade 9-12) from the years nineties. We found correlation between the traditional and of this kind of mathematics competitor's competencies. Since 2005 some parts of the mathematical final examination have test form (grade 12). So our students were not accustomed to solve test problems, they did not know the methods of solving tests. If we look at the history of Hungarian education we find that some educators applied tests since the World War I (M. Éltés) and at a few schools in physics in the nineteen-thirties (in Debrecen: G. Barra and L. Tóth). Barra wanted to introduce the test assessment at the University entrance examinations too (1948), but in that time it was impossible, in consequence of the Hungarian state politics, his ideas were refused.

Methods of solving tests

Students like to solve tests. Competitors told us that it is a joy, a challenge, a focus of mathematical experiment. The methods and tactics of solving test problems are different from the typical mathematical problem solving methods. The role of clever guessing and systematic trials, the method of trial and error increases. It is not necessary to write down precisely the solutions, the thoughts, the assumptions are important. They can work forward and backward too. They have to make decisions quickly and relative well. It is important to make computations without mistakes. The knowledge of general mathematical theorems and the exploitation of the symmetry principle are useful. We found another problem in filling the answer sheets, they made formal mistakes, and they have to pay more attention to right marking.

The method of substitution of the choices is a special method to solve multiple-choice test problems. We shall present and compare two different methods for solving a test problem, a traditional method and the above mentioned substitution method.

E.g. *Problem:* For which k are the roots of the equation $x^2 + 2(k+2)x + 9k = 0$ equal? - (A) 4 (B) 1, 4 (C) 0, 4 (D) -1, 4 (E) 2, -4

Solutions:

1. *Standard method.* First we calculate the value of the discriminant: $D = 4(k+2)^2 - 36k$. The roots are equal if $D = 0$, i.e. $k_1 = 0$, $k_2 = 4$. The answer B is right.

2. *A test method. Method of checking, trial and error.*

It seems to be better to try at first the value $k = 4$, so we substitute the value $k = 4$ in the equation. It is right, because $(x + 6)^2 = 0$.

Possible answers could be: A, B, C D.

We have to check another value too. If we have luck we immediately can get the only right answer, because only one answer is right. We can recognize that $k = 0$ is a wrong value and we drop answer C.

I recommend to try the value $k = 1$. It is right, because $(x + 3)^2 = 0$. So we got the proper answer: B.

Some experiences in solving tests

Preparing our students for the AMC we solved the tests of the previous years and drew their attention to these methods. We said something about guessing too. We advised at first to read attentively the text of the problem. It is necessary to solve the problem carefully if there are wrong answers between

the offered answers. There is a penalty of wrong answers! Sometimes we can find among them the results of the possible mistakes (for example mixing the concepts of perimeter or area). “Distractor” is a test – maker language for wrong answer. The use of this word suggests that the answers are there to trick the students. Suppose they are asked to find the diameter of a circle meeting some condition. Now we suppose that the answers include both 5 and 10. The problem asked for the size in terms of diameter, but they figured it out in terms of radius. Students did a lot of good thinking and hard work to prove that the radius is 5. If after all their work, they make the small slip of forgetting to convert, they get the problem wrong.

Other example: students are asked for the percentage change if a price increases by 20 % and then by 20% again. Suppose both 40% and 44% are among the answers. Students reason that percents add and choose 40. In both cases students have been distracted by a ”distractor”, but the situations were very different. In the first case students were mostly right. In the second case students were using a completely wrong approach. Ambiguous answers will be marked wrong. To guess before eliminating three of the five choices they will get, on average, lower scores. In our practice the scoring system was often changed:

- a) *Short calculations(standard method)*
- b) *Guessing (clever or by chance)*
- c) *Intuition*
- d) *Computation of the possible cases*
- e) *Scale drawings and measuring*

It is important to determine aids which the students could use in solving test problems. At the competitions it was allowed to use a ruler and a compass. We observed that the competitors made scaled figures, they determined the necessary measures and this way they could guess the right result.

About the PISA items

The items of PISA tests have a variety of formats described over. In many cases students were required to construct a short response in their own words. Sometimes they had to write down their calculations in order to demonstrate some of the methods and thought processes they used in producing an answer. Other questions required students to write an explanation of their results, which again exposed aspects of the methods and thoughts processes they had employed to answer the question. Such kind of open-constructed responses cannot be machine-scored.

For other items requiring students to construct a response, but the answer was given in numeric or in other fixed form. There were items that required students to select one of or more responses from a number of given answers. This format category includes both standard multiple-choice items, for which students were required to select one correct response from a number of given response options, and complex multiple-choice items for which students were required to select a response from given optional responses to each of a number of propositions or questions. Students were given credits for each item that they answered with an acceptable response. In other cases it was possible to give partial credits for the category “ is not quite good”.

85 mathematics items were used in PISA 2003. Each student was given a test booklet with four clusters of items resulting in two hours of individual assessment time. It was possible to construct a scale of mathematical

performance to associate each assessment item with a point score on this scale according to its difficulty and to assign each student a point score on the same scale representing his or her estimated ability. To facilitate the interpretation of the scores assigned, the scale was constructed to have an average score among OECD countries of 500 points. The students' score in mathematics were grouped into six proficiency levels.

Items with relatively high difficulty: item VI (668,7- ?) and item V (606,6-668,7). Items with moderate difficulty: item IV (544,4-606,6) and item III (482,4-544,4). Items with relatively low difficulty: item II (420,4 - 482,4) and item I (358,5-420,4).

For the students it required to reach Level 2. Level 2 means that students can recognise information from texts of the problems, they can manipulate with algebraic tools, terms and algorithms and can apply them for computational skills.

Students' performance in the four areas

1. Spaces and shapes

- *Carpenter* (Question 1(687), Level 6)
- *Staircase* (Question 2 (421), Level 2)
- *Number cubes* (Question 3(503), Level 3)

2. Change and relationships

- *Walking* (Question 5-Score 3 (723), Question 5 - Score 2 (666), Question 5- Score 1 (605), Question 4 (611), Levels 4-6)
- *Growing up* (Question 8 (574) , Question 7- Score 2 (525), Question 7- Score 1 (420) , Levels 2-4)

3. Quantity

- *Exchange rate* (Question 9(406), Question 10(439), Question 11 (586), Levels 1-4)
- *Skateboard* (Question 12- Score 1 (464), Question 12 -Score 2 (496), Question 13 (570), Question 14(554), Levels 2-4)

4. Uncertainty

- *Test scores* (Question 6 (620), Level 5)
- *Robberies* (Question 15-Score 1 (577), Question 15-Score 2 (694), Level 4-6)
- *Exports* (Question 17 (427), Question 18 (565), Levels 2-4)

Space and shapes

- *Carpenter*. This problem was a *complex multiple-choice item* situated in an *educational context*, since it was a kind of quasi-realistic problem. It fitted the *connections competency cluster*, as a non-routine problem. The students needed the competence to recognise that the shapes A, C, D have the same perimeter, 32 metres, but the perimeter of parallelogram B requires more than 32 metres timber. They needed to decode their visual information. (*Level 6*)
- *Staircase*. This problem was a *short-open constructed response item* situated in a *daily life context*. This item had redundant information. The total depth was unnecessary and confusing for the students, but they could make use the representation. This item belonged to the *reproduction competency cluster*. (*Level 2*)
- *Number cubes*. This problem was a *complex multiple-choice item* in a *personal context*. The problem posed requires some spatial insight or visualisation technique. This item was not a routine problem; it belonged

to *connections competency cluster*. It required spatial perception, a connection between space and plane, model and reality. (*Level 3*)

Countries were ranked in descending order of percentage of the students in levels. On the scale of proficiency on space and shape we found the following order: 27. *Hungary* (from 41 countries). A quarter of students fail to reach Level 2 in Hungary. Level 6 reach 4,5% of students.

Change and relationships

- *Walking*. *Question 5* was an *open-constructed response* situated in a *personal context*. There were two levels of partial credits. The mathematical routine needed to solve the problem with a substitution in an algebraic formula. They needed to do some conversions between units of measures. The lower partial credit part of this item belonged to the *connections competency cluster* at the top part of *Level 4*, the higher level of partial credit illustrated *Level 5*. Full credit illustrated the upper top of *Level 6*. - *Question 4* was an *open-constructed item* situated in a *personal context*, an item about relationship: number of steps per minute and pace length. The students needed to carry out the actual calculation in order to get full credit. This item belonged to the *reproduction competency cluster* (*Level 5*)
- *Growing up*. *Question 8* was an *open-constructed response item*. The communication of the students' observations was central to the question. It belonged to the *connections competencies cluster*. This item required mathematical insight and some reasoning and communication of the results (*Level 4*). In *Question 7* the students had to compare characteristics of two data sets and interpret these data sets and draw conclusions. It belonged to the *reproduction competencies cluster*. (*Level 1, 2*). *Question 6* was a *closed-constructed response item* situated in a *scientific context*: the growth curves of young males and females over a period of ten years. The students had to translate the question into mathematical context, but there they carried out only a subtraction. It belonged to the *reproduction competency*. (*Level 2*)

Countries were ranked in descending order of percentage of the students in levels. On the scale of proficiency on change and relationship we found the following order: 21. *Hungary* (from 41 countries). The result of this part was better than in PISA 2000.

Quantity

- *Exchange rate*. *Question 11* was an *open-constructed response item* situated in *public context*. The competences needed to solve this problem were not trivial, but they got all the information. It belonged to the *reflection competency cluster*. The students needed problem solving competencies, some communication, some argumentation, and reasoning (*Level 4*). *Question 10* was a *short-constructed response item*, situated in *public context*. It was a routine procedure and application of a standard algorithm. It belonged to the *reproduction competency cluster* (*Level 2*). *Question 9* was a short constructed response item in a public context. All the required information was explicitly presented. There were clearly defined familiar questions. It belonged to the reproduction competency cluster (*Level 1*).
- *Skateboard*. *Question 13* was a *multiple-choice item* in the *personal context*. Students had to make combinatorial computations. If they were

not accustomed to compute such kind of problems, they could apply the strategy of systematic listing. It belonged to the *reproduction cluster (Level 4)*. *Question 14* was a *short constructed response*. Students had to look at the table with prizes, they had to make combinations and do some computations. It belonged to the *connections competency cluster*. *Question 12* was a *short constructed response in a personal context*. Students had to find strategies for getting the minimum and the maximum. This strategy was part of the practised knowledge. It belonged to the *reproduction competency cluster (Level 2, 3)*.

Countries were ranked in descending order of percentage of the students in levels. On the scale of proficiency on quantity scale we found the following order: 24. *Hungary* (from 41 countries). Under Level 1 were 21% of students, and 2,5% of Hungarian students reach Level 6.

Uncertainty

- *Robberies*. *Question 15* was an *open-constructed response item situated in a public contest*. This item involved the analysis of a graph and data interpretation. It belonged to the *connections competency cluster*. The essential competencies for solving this problem were: understanding and decoding of a graphical representation in a critical way. The partial credit response illustrated *Level 4*, full credit response illustrated *Level 6*. Students had to explain in detail that the statement was unreasonable.
- *Test scores*. *Question 16* was an *open-constructed item in educational context*. Knowledge of this area of mathematics is essential in an information society. It belonged to the *connections cluster*. Students applied their statistical knowledge in a problem situation (*Level 5*).
- *Exports*. *Question 17* was a *closed-constructed response item situated in a public context*. The mathematical content was to read data from a bar diagram. It belonged to the reproduction cluster (*Level 2*). *Question 18* was a *multiple-choice item situated in a public context*. Students needed to combine the information of the two graphs in a relevant way. It belonged to the *connections cluster (Level 4)*.

Countries were ranked in descending order of percentage of the students in levels. On the scale of proficiency on uncertainty scale we found the following order: 22. *Hungary* (from 41 countries). Level 1-2 reach 21,2% of students, and 1,6% of the Hungarian students reach Level 6.

Overall performance in mathematics PISA assessment in Hungary

The average of Hungary is a little bit under the expected OECD average (490 points). At each level of proficiency on the mathematics scale we found the following order: 26. *Hungary* (from 41 countries).

Reference

A Profile of Student Performance in Mathematics OECD 2004, Learning for Tomorrow's World- First Results from PISA 2003