

# What may impact the time of solving tasks on PISA 2015?

Malinowska Małgorzata, Michalczyk Sylwia, Przypański Olga

Warsaw University of Technology  
Faculty of Mathematics and Information Science  
27th April 2017

## Abstract

- On average, during exam students spend the most of their time on mathematical tasks and they receive the worst results from them.
- The type of booklet received does not affect the response time profile. Students spend more time on tasks in the first and third position than on those in remaining two positions, regardless of what type those tasks are.
- The factor that affects the results in each position is the level of stress while taking the exams. Students who are stressed about taking the exam need more time to solve tasks in positions 2 and 4. However, it does not affect their results, which are getting worse in the next sections of the test.

## The study purpose

The main goal of our study was to examine factors that influence the amount of time that students spend on tasks on different parts of a test.

First, we compare results of mathematics, reading and science in ten selected countries. Second, we examine importance of the order in which students received tasks from particular subject. To make our analysis more precise we take into account the difficulty of tasks in every booklet. In the last part we investigate if pre-exam stress influences the amount of time students spend on tasks.

## Description of the data

In our study we analysed data of **PISA 2015**.

PISA (The Programme for International Student Assessment) is a worldwide study by the OECD (Organisation for Economic Co-operation and Development) in member and non-member nations of 15-year-old school pupils' scholastic performance on mathematics, science and reading.

PISA 2015 represents the sixth such study since PISA was first conducted in 2000. Seventy two OECD countries or partner economies participated in PISA 2015.

In PISA 2015, the main mode of assessment has moved from a paper-based delivery to a computer-based delivery. Students may perceive items as being of varying difficulty, or students may apply varying degrees of effort, depending on the position of the item in the test booklet. Rotating the positions of items across different versions of test booklets mitigates this effect. In PISA 2015, computer-based delivery allowed for 66 different test forms.

Let's point out that each booklet consisted of four parts and each of these parts contains one type of task. In what order the students received their assignments from mathematics, science and reading, it depended on booklet number which they obtained.

We selected students from 10 countries: Australia, Austria, Belgium, Brazil, Estonia, Finland, Germany, Greece, Poland, Portugal. Analysed data included information about results of 6 321 students.

We consider only 36 booklets (with numbers from 31 to 66) and items which was solved with the computer-based assessment.

## Comparison of countries' results

In this part we compare the average completion time and the average results in countries for every type of tasks (mathematics, reading, science). We want to investigate if one of these types proves to be more difficult for students and find out the best countries in each subject. The plot below presents our findings.

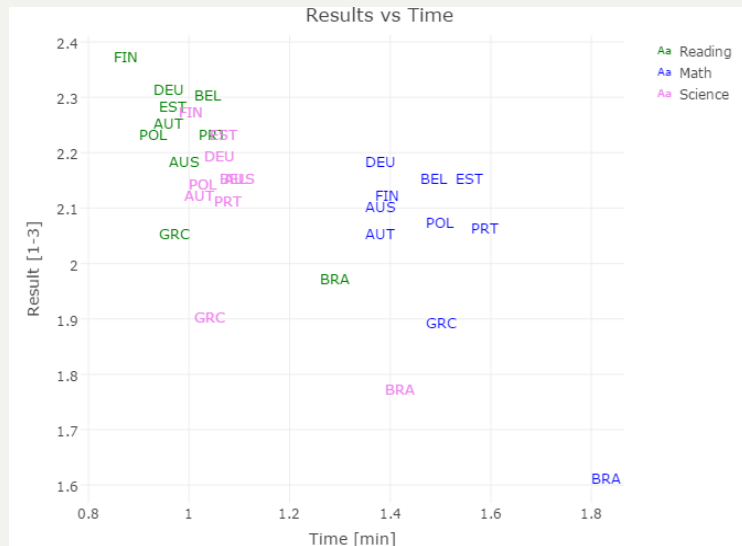


Figure 1. The figure shows the results of three types of tasks for each country.

As we can see Finland, Estonia, Belgium and Germany got the best average results. From the remaining countries, Brazil stands out the most. Not only the time it took for Brazilian students to solve the test was the longest, but also their results were the worst.

It can also be noted that the reading parts of the test were solved in the shortest time and results for reading were the highest. Finnish students were the quickest and also achieved the best results among considered countries in this kind of tasks. Results and completion times of science tasks were only slightly worse.

Mathematical tasks seem to be the hardest for students, since they took them the longest to complete and the average score from them is the worst. Brazil has the longest average completion time and the lowest score from these tasks.

## Booklet analysis

Every student gets different booklet which has some particular order of tasks of different type (mathematics, reading, science).

The main purpose of this part of our analysis is to examine the importance of that order.

We divide all booklets into six groups:

type 1 - numbers from 31 to 36 (S,S,R,R)

type 2 - numbers from 37 to 42 (R,R,S,S)

type 3 - numbers from 43 to 48 (S,S,M,M)

type 4 - numbers from 49 to 54 (M,M,S,S)

type 5 - numbers from 55 to 60 (S,S,M,R) or (S,S,R,M)

type 6 - numbers from 61 to 66 (R,M,S,S) or (M,R,S,S)

A type and order of tasks is specified in parentheses, where M corresponds to mathematics, R to reading and S to science.

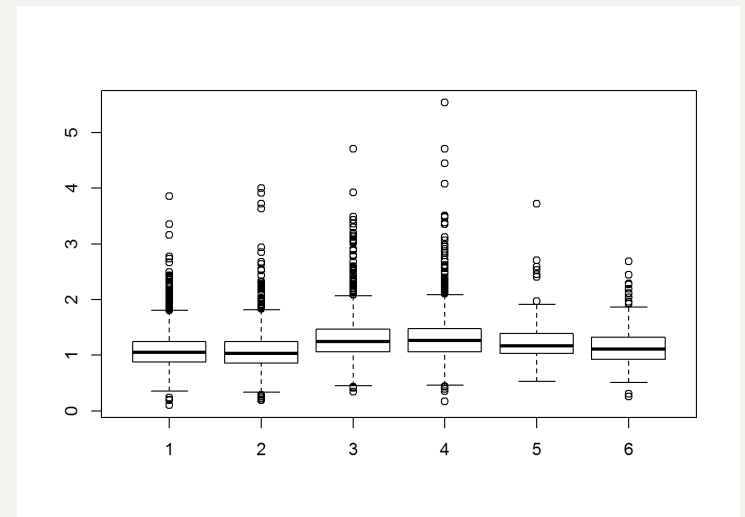


Figure 2. The figure shows the distribution of time for all six groups of booklets.

On the plot we can observe some differences between groups. Types 3 and 4, which both had mathematical tasks in two parts of the test, had the longest solving times.

## Consideration of the difficulty of tasks

We examine if the type of a task influences the response time profile. We want to include the fact that every type of the task has different difficulty.

To do this we divide a completion time for particular task by mean completion time for tasks of that type in its booklet. We do this for all tasks, from all booklets.

Thanks to normalization, we will be able to see a relation between completion times and position of task in a booklet and whether a task was being solved longer than the average for its type in a booklet.

## The most important is to take it easy

We would like to know if the level of stress affects the results obtained on the individual positions.

We use the question number ST118 of the questionnaire, which contains 5 sub-points to prepare the analyzes. In each of these points, the students indicated how much they agree with the statement. The questions were structured in such a way that a pupil who did not care about the examination was disagreed with the statement and the agreement with the statement reflects the student who is very stressed by the exam.

We would like to create variable that specifies the level of the stress by using all sub-points from the question.

To do this we have to turn the factorial answers into number:

- “Strongly disagree” = 1,
- “Disagree” = 2,
- “Agree” = 3,
- “Strongly agree” = 4.

Next, we calculate the average of the responses to items from 1 to 5 and classify the students into three groups:

- The mean between [1;2] = ‘A little nervous’,
- The mean between (2;3] = ‘Under stress’,
- The mean between (3;4] = ‘Very stressful’.

The following charts visualize calculated results.

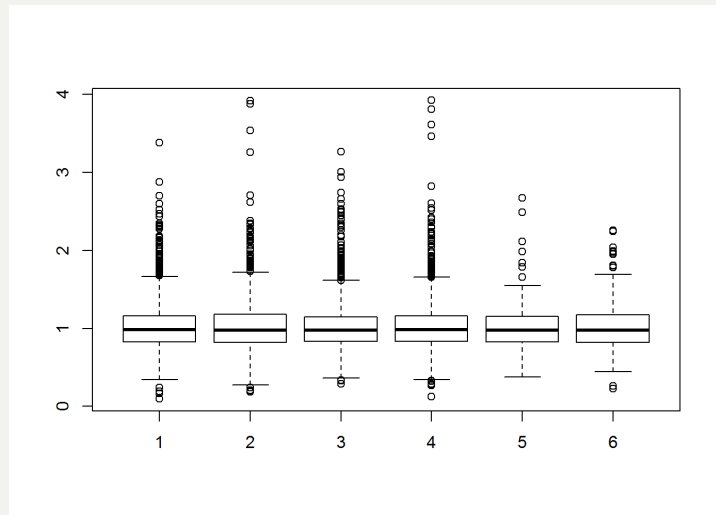


Figure 3: The figure shows the distribution of time for all six groups of booklets after normalization.

We can see that the differences between groups of booklets shrunk.

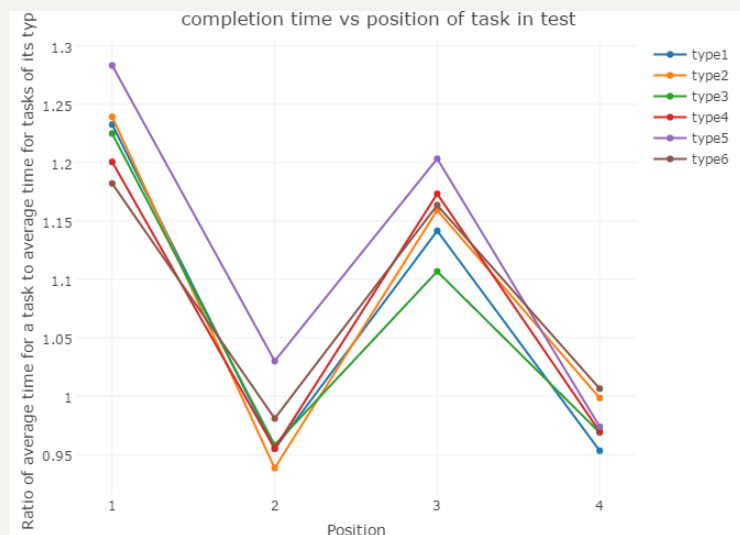


Figure 4: The figure shows the ratio of average time for a task to average time for tasks of its type for each of four positions.

Observing the plot, we can conclude that a type of tasks seems to have no effect on response time profile i.e. students take a test similarly, independently of a group of a booklet. Students take much longer time to solve tasks in the beginning of a test, and that time decreases towards the end. It does not matter if the first part is mathematics, reading or science.

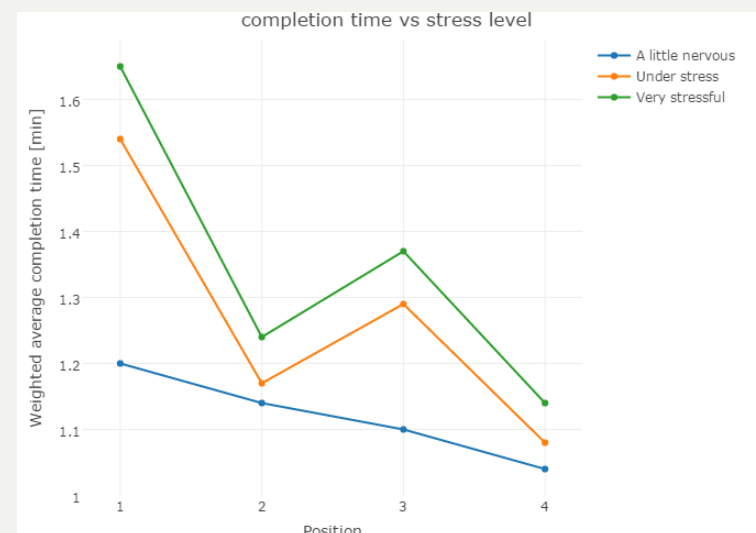


Figure 5: The figure shows average completion time on each position for three groups of stress level.

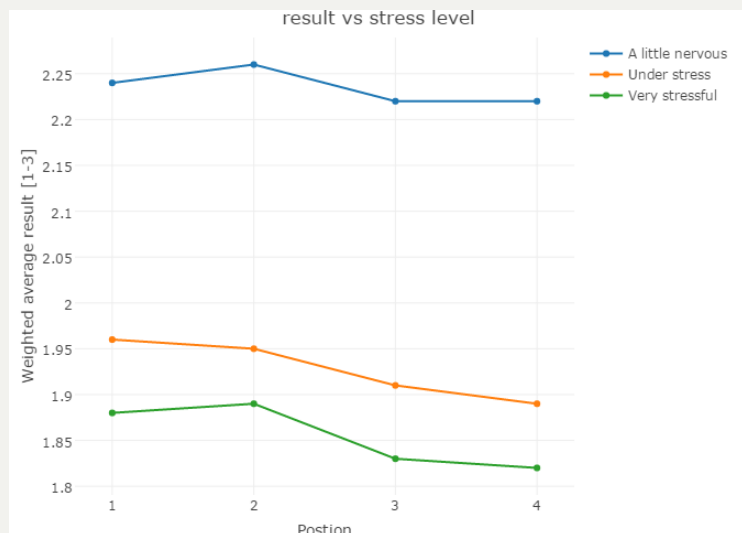


Figure 6: The figure shows average score on each position for three groups of stress level.

It can be observed that students who have declared that they do not stress the exam stand out from the rest. The time at which the tasks are solved is shorter and the successive positions tend to decrease. This is also evident in the results obtained - they get much better results from other students.

On the other hand, stressful students solve tasks longer on positions 1 and 3 than on positions 2 and 4. This could mean that their stress decreases during the exam. They need to spend more time on tasks when they come back after a break (the third position).

However, it seems that the time spent on solving tasks does not affect results in individual positions.

We can suppose that students who are relaxed are the well-prepared ones. They solve tasks relatively quickly and get good results regardless of task positions. On the other hand, stressful students need more time and get worse results.

In contrast, students who can not control stress during exams need a lot more time to work on positions 1 and 3 than on positions 2 and 4. Their average score at each position is about one point lower than for students who are relaxed.

We can conclude that students who are able to dominate stress during examinations can focus on solving tasks better, and thus gain better results than the rest.

However, the question remains whether the level of stress is the result of a student's level of knowledge and preparation for the test, or vice versa - the student's score is a result of how stressful he is. The answer to this question requires greater sociological knowledge, so we leave them without answer.

## Results

Our analysis showed that the main factor associated with the time that students need to solve tasks in the PISA 2015 study is the stress caused by taking exams.

Students who were less stressed received very good results and the time they needed to solve the tasks was relatively small and did not depend on the task position in the test.