



EDUCATION AND NEW DEVELOPMENTS 2018

**Edited by
Mafalda Carmo**

Education and New Developments

2018

Edited by
Mafalda Carmo

Edited by Mafalda Carmo, World Institute for Advanced Research and Science (WIARS), Portugal

Published by InScience Press, Rua Tomas Ribeiro, 45, 1ºD, 1050-225 Lisboa, Portugal

Copyright © 2018 InScience Press

All rights are reserved. Permission is granted for personal and educational use only.

Commercial copying, hiring and lending is prohibited. The whole or part of this publication material cannot be reproduced, reprinted, translated, stored or transmitted, in any form or means, without the written permission of the publisher. The publisher and authors have taken care that the information and recommendations contained herein are accurate and compatible with the generally accepted standards at the time of publication.

The individual essays remain the intellectual properties of the contributors.

ISSN (electronic version): 2184-1489

ISSN (printed version): 2184-044X

ISBN: 978-989-99864-8-0

Legal Deposit: 428062/17

Printed in Lisbon, Portugal by GIMA - Gestão de Imagem Empresarial, Lda.

THE MATHEMATICAL SKILLS AND TIME AVAILABLE IN STANDARDIZED TESTS

Marina De Nunzio, Clarissa Sorrentino, & Stefania Pinnelli

Department of History, Society and Human Studies, University of Salento (Italy)

Abstract

The research project sets out to investigate the influence of the “time available” variable in the performance of INVALSI math tests in Italian students of secondary school. The INVALSI tests are standardized tests, based on articulated and rigorous procedures, for Italian school students. Evaluation of students’ performance through INVALSI tests is mandatory by law (Article 51 paragraph 2 of the Decree-Law of 9 February 2012, No. 5 converted into law No. 35). The tests are elaborated by the National Institute for the Evaluation of the Education and Training System (INVALSI), a public law research body of the Ministry of Education, University and Research (Miur). The purpose of the tests is to draw a statistical reference framework on the level of learning in Italy. With the INVALSI tests it is possible to monitor the national education system and compare it with other European institutions. The INVALSI Test carried out at the conclusion of the Third-Degree State Examination of lower secondary school was an instrument for the certification of the students’ learning (Fondazione Giovanni Agnelli 2014, p. 68). A study of 2011 (Chamberlain, Daly, Spalding), aimed at exploring the causes that trigger the examination anxiety, highlighted the importance of time available to complete the test among the causes. The time allotted to carry out a test or, in any case, the perception that students have of it can therefore be an element to be taken into consideration because it could alter the measurement of learning. Specifically, the survey aimed to perform a statistical analysis of the performance results of a sample of 137 students when the time variable varies. In particular, we wanted to establish if the time variable influences a different expression of skills. In first analysis it would seem that the time factor does not have direct consequences on students’ performances, but Rasch analysis has shown that the time difference granted to students influences the emergence of students’ skills in relation to the different difficulty level of each item of the math test. Finally, the survey describes the influence of the variable “luck” in the results of standardized evaluations.

Keywords: *INVALSI, time, mathematics, secondary school, performance.*

1. Introduction

The INVALSI tests are standardized tests for Italian middle school students, based on articulated and rigorous procedures. These tests are mandatory by law (Article 51 paragraph 2 of the Decree-Law of 9 February 2012, No. 5 converted into law No. 35), this means that schools in Italy, in specific classrooms must perform them. The tests are elaborated by the National Institute for the Evaluation of the Education and Training System (INVALSI), a public law research body, of the Ministry of Education, University and Research (Miur). The purpose of the tests is to draw a statistical reference framework on the level of learning in Italy. With the INVALSI tests it is possible to monitor the national education system and compare it with other National and European institutions.

The INVALSI Test carried out at the conclusion of the Third-Degree State Examination was an instrument for the certification of the learnings of the students (Fondazione Giovanni Agnelli, 2014, p. 68).

Although the Italian INVALSI Tests have a census and non-sampling purpose, they have the same evaluation approach, based on standardized structured tests, and the same statistical evaluation model of the OECD Pisa (Program for International Student Assessment) tests. This model is based on the matrix of Rasch (Ray Adams, Comments on Kreiner 2011) that has been significantly criticized in Italy and abroad (A. Angelucci, 2015). The substantial difference between the two tests is that the Pisa OECD tests are aimed at the assessment of skills in context while the INVALSI Tests remaining anchored to the

school program, in both cases are time trials, although the OCSE Pisa tests give two hours of time available.

A study of 2011 (Chamberlain, Daly, Spalding), aimed at exploring the causes that trigger the examination anxiety, highlighted the importance of time available to complete the test among the causes. The time allotted to carry out a test or, in any case, the perception that students have of it can therefore be an element to be taken into consideration because it could alter the measurement of learning. Moving from these elements, the survey aimed to perform a statistical analysis of the performance results of a sample of 137 students when the time variable varies. In particular, we wanted to establish if the time variable influences a different expression of skills.

2. Research design

Pupils from third-year secondary schools were involved in the research. The proposal to participate in the project was sent to 7 local schools (Lecce and the province) of which 3 of them were available to participate in the project. Among these three schools, 8 classrooms were selected. The identification criteria suggested to the teachers for the classroom selection were: heterogeneous classes for performances thus avoiding excellent or particularly weak classes; heterogeneity by gender; classes that had not already used the INVALSI tests selected for the search. Overall a total of 188 students participated in the investigation. During the analysis of the results, the sample has undergone a reduction. The final non-probabilistic sample included 137 students, specifically 72 females and 65 males. Students who did not take part in the whole investigation procedure and pupils with disabilities were excluded from the initial group. Each of the 8 classes, for a total of 137 students, has been divided into 2 sub-groups, also assorted so as to equally distribute the school performances of the students. Subgroup A and Subgroup B were thus constructed. As indicated below:

Table 1. Distribution of the students in the sub-groups.

	MALE	FEMALE	TOT
SUB-GROUP A	31	32	63
SUB-GROUP B	34	40	74
TOT	65	72	137

3. Instruments

For the evaluation of mathematical performances, 2¹ INVALSI Tests of previous years were selected. In particular, we chose to use the school year 2010/2011 (called α test) and the school year 2011/2012 (called test β). The relative correction and scoring grids have also been used.

4. The Survey on the field

In the month of May 2017, two administrations of INVALSI Tests (Test α and Test β) were proposed in 2 separate subgroups (see Table 2), in 2 different days of the same week, in order to avoid a significant increase in the mathematical skills of the students between the 2 administrations. Each group, therefore, carried out both tests having different time frames available. In particular:

Table 2. Organization of the administration.

	1°ADMINISTRATION TIME: 120 minutes	2° ADMINISTRATION Time: 75 minutes
SUB-GROUP A	Test α	Test β
SUB-GROUP B	Test β	Test α

The administration took place inside the classrooms of the boys, in the presence of a teacher of the class (not necessarily that of mathematics) and at least one researcher. The boys' desks were separated and placed by file; the pupils, as far as possible, were seated so as not to have close companions who were doing the same test (in fact in the classroom there were guys who were taking the α test and guys who were doing the β test).

¹We chose to select two tests to be able to keep the "Test" factor under control, i.e. we wanted to check that the possible effects of the time available in the answers to the questions were not related to a specific test but could be generalized to anyone.

5. Analysis of the results

The first type of analysis carried out took into consideration the average scores obtained by the boys at both tests. Table 3 shows the test averages.

Table 3. Average of the scores of the two tests in the 2 administrations.

	1°ADMINISTRATION 120 minutes	2° ADMINISTRATION 75 minutes
Sub-group A (63 students)	Test α 36,84	Test β 25,9
Sub-group B (74 students)	Test β 26.76	Test α 34.64

At a first glance we immediately notice that in the Sub-group A the performance performed in more time has recorded higher scores significantly compared to the test performed in 75 minutes, this, therefore, would seem to confirm that a decrease in time available to perform the proof corresponds to a lower "possibility" to demonstrate one's own learning.

However, if we look at the averages of the two tests obtained from Subgroup B, we can see that the inverse condition occurs, i.e. that the performance has improved as time passes. What could have caused such contradictory results?

In order to understand what this different situation is due to, it is necessary to compare the two tests, in particular the mean scores of the α test in 120 with those of the β test carried out at the same time.

As indicated in Table 3, the mean α Test scores in the two administrations are generally higher than the mean β Test scores. A t-test detects significant differences between the two tests in both administrations. In the first administration (120 minutes), the mean scores are equal to 36.84 and 26.76 for α and β , respectively, with a difference significantly different from zero ($t = 6.968$; $p = 0.000$). Likewise, in the second administration (75 minutes) the difference between the mean scores of α and β denotes a significant difference between the two tests ($t = -6.062$; $p = 0.000$). Therefore, the α test and the β test are significantly different from each other and the α test is easier (since the results in both modes of administration are significantly higher) than β .

This first data, although interesting because it highlights the "luck" factor in the final evaluation of the students, does not allow to verify if the time of the test has influenced the performance results for the same subgroup of subjects. It is therefore necessary to compare the results to the same test carried out at different times by the two subgroups of students, relying on the fact that the starting samples were constructed by homogeneously distributing the pupils by level of learning, gender and experience of the tests. The test carried out in less time recorded lower average scores. Analyzing the means of the results, we can see that among them there is a difference, but a t-test showed little significance of this difference ($t = 1.35$; $p = 0.17$).

Taking into account the β Test carried out in 120 minutes by the subgroup B and by the subgroup A in 75 minutes we can see that the difference between the average scores of the two tests is even smaller. Also in this case a t-test showed the non-significance of this difference ($t = 0.67$; $p = 0.49$).

In summary, therefore, through the analysis of the means it emerged that the time allotted to carry out an INVALSI Test does not seem to affect the performance of the students.

However, a further analysis was carried out (both for the α test and for the β test) which took into account the difficulty coefficients of the individual test items. In particular, the objective of the survey was to check whether these coefficients were subject to variations as the time allocated for the performance of the test varies. It was therefore proceeded to calculate separately the difficulty parameters of the various items, both for the α test and for the β , depending on whether these were carried out in 120 or 75 minutes (see Table 4). These coefficients have been obtained through an analysis of the performances of the boys elaborated through the Rasch Model. The underlying median data indicate the parameters calculated with the Rasch and those estimated directly by the INVALSI, indicated in the technical reports of the Italian Ministry of Education relating to specific tests.

Table 4. Difficulty parameters for each item of the α and β Test and evaluated according to the Rasch Model.

TEST α	TEST β
<ul style="list-style-type: none"> ● Average of difficulty coefficients of INVALSI items $\rightarrow -0.63$ ● Average of difficulty coefficients of Test α items, test performed in 120 min $\rightarrow -0.59$ ● Average of difficulty coefficients of Test α items, test performed in 75 min $\rightarrow -0.42$ 	<ul style="list-style-type: none"> ● Average of difficulty coefficients of INVALSI items $\rightarrow -0.07$ ● Average of difficulty coefficients of Test β items, test performed in 120 min $\rightarrow 0.02$ ● Average of difficulty coefficients of Test β items, test performed in 75 $\rightarrow 0.15$

The Rasch analysis showed that the difference is due to the different difficulty level of each item. In relation to both tests, comparing the average of the difficulty rates of the items in 120 and 75 minutes, it should be noted that the difficulty increases when the testing time decreases. Furthermore, the distribution of the difference in values has zero mean ($p = 0.012$) which means that the variation is not attributable to the case, so the difference is significant (non-Gaussian distributions based on the Kolmogorov-Smirnov test; for paired samples) (see Table 5).

Table 5. Comparison between the means of the difficulty parameters of the tests carried out in 120 or 75 minutes with the Invalsi estimates.

<p>TEST α:</p> <ul style="list-style-type: none"> ● Alpha 120 vs INVALSI: ($p = 0.73$) null hypothesis not to be rejected. ● Alpha vs 75 INVALSI: ($p = 0.04$) null rejected hypothesis that is significant difference.
<p>TEST β:</p> <ul style="list-style-type: none"> ● Beta 120 vs INVALSI: ($p = 0.27$) null hypothesis not to be rejected ● Beta 75 vs. invalted comparison: ($p = 0.0002$) null rejected hypothesis that is significant difference.

From the comparison of the difficulty coefficients of the 120 and 75 minute tests with the INVALSI estimates (statistical significance analysis performed with sign test), it emerged that the estimates of internal difficulties of the INVALSI items overlap more than those estimated for the test carried out in 120, however, differ significantly from those carried out in 75 at least for the sample of pupils taken into consideration, this means that, conversely, with more time the children examined can achieve the expected results for a standard sample INVALSI in both tests.

6. Discussion

As already amply described in the previous paragraph, the analysis of the test scores averages revealed conflicting data within the 2 sample subgroups. In particular, in fact, if we considered exclusively what happened in the performance of Sub-Group A, we would try to confirm what was hypothesized when the research was started, ie that the time allowed to carry out an INVALSI test affects the outcome of the performance itself. But the results emerged in Subgroup B contrast with this because they manifest an opposite situation. What to say then about the relationship between time and performance? As has emerged through the analysis carried out with the t-test it is not possible to compare the performance of students of the same subgroup in the two tests (alfa and beta) because the two tests are significantly different.

The analysis carried out by comparing the results of the 2 subgroups on the basis of each test showed that the variations between the results of the youngsters exist (more time has produced higher scores / marks) but are not significant.

Going to consider the internal components of the tests, ie the index of difficulty of the items, it emerged that as time decreases, the difficulty coefficients of the tests increase, even significantly. Moreover, the difficulty coefficients of the items estimated by the INVALSI (which are calibrated on tests to be carried out in 75 minutes) are superimposable to those calculated through the Rasch Analysis starting from the results of the tests carried out in 120 minutes by the students examined in this research. . This may mean that the students of the classes under consideration in this study do not reach the standards established in 75 minutes (therefore they show lower performances compared to those of the sample classes), but they manage to do so having more time available. This would seem to confirm the starting hypothesis that the test time is a significant variable from the perspective of the extrinsic of one's own abilities in the mathematical field.

7. Conclusions

It is not easy today to establish the correlation between the time variable and the mathematical abilities, but it is clear that a relationship between them exists. It is not possible to deduce this only from the analysis of the scores or marks achieved by the students because the method of attribution of the same, for subsequent "boxing", reduces the evidence of this data. But the analysis of the difficulty coefficients of the tests shows the existence of an inverse correlation between the time allocated for the performance of the test and the difficulty of the same. This does not call into question the validity of the INVALSI Tests, which, it is clear, have a census scope, but raises some perplexities from the point of view of the evaluation of mathematical abilities.

In this regard it is necessary to make a further statement: the analysis carried out showed that the tests chosen for this research have an objective difference in relation to their difficulty coefficient. And this is even more evident if we notice that the same subgroup of youngsters shows extremely different performances in the two Tests: in Subgroup A the sufficient or more than sufficient performances (from 6 up) pass from 79% (Test α) to 37% (Test β) of the sample and for Subgroup B from 31% (Test β) to 73% (Test α).

This finding raises a question: starting from the assumption that these Tests are National Tests or have been examination tools respectively in the years 2011 and 2012 and that "the INVALSI Test carried out at the conclusion of the State Exam of the Secondary Secondary School also serves to certify the learnings of the students "[Cf. Agnelli Foundation; 2014, page 68], which of the values (scores / grades) achieved by the students can be considered the value of the learnings of the students? We recall, in fact, that the vote obtained by the students at the National Trial until 2017 has contributed to the final evaluation of the examination of the first cycle of education and today is equally significant as it is a prerequisite for admission to state examinations. Moreover, an analysis of the coefficients of difficulty of the National Tests carried out from 2010 to 2017 (these data are included in the Technical Reports published by the INVALSI annually) shows a considerable variability between the Tests, so a student with average skills can get extremely different results while working at most, in fact, in addition to his skills depending on the test he is called to perform, the time available and the luck factor play a considerable role. It is highlighted that time management is a critical element for the emergence of skills in those boys with average grades that make up the majority of students. As it is known from the 2017/2018 school year, following the D.M 741/2017, the Tests are granted more time, it goes from 75 to 90 minutes but will be tests to be performed on the computer. This novelty in the administration apparently seems to have grasped the problem of the variable time available for the conduct of the tests and therefore respond to the need for more calm and reflection, the other will risk to nullify this data having to deal with the computer tool, with the connection of the school network in addition to the familiarity of the instrument itself in solving a mathematics task.

In conclusion, considering the relationship that the variable time has in the emergence of mathematical skills in relation to the difficulties of the tests, considering also that this element affects the boys which perform at an average level, furthermore considering that the elaboration of solution-problem processes can follow creative and complex processes, that requires the implementation of a specific competency. It is believed that freeing the INVALSI tests from such rigorous time margins, as well as helping to lower individual anxiety levels, can contribute to a more realistic expression of learning processes.

References

- Angelucci A. (2015). *OCSE PISA e INVALSI: una prospettiva europea*. Roars. In <https://www.roars.it/online/ocse-pisa-e-invalsi-una-prospettiva-europea/>
- Borkovec T.D., Ray W.J., Stoeber J., (1998). Worry: a cognitive phenomenon intimately linked to affective, physiological, and interpersonal behavioral processes. *Cognitive Therapy and Research*. vol 22, pp. 561-576.
- Caponi B., Falco G., Focchiatti R., Cornoldi C., Lucangeli D. (2006). *Didattica meta cognitiva della matematica. Nuove prospettive e strumenti*. Erickson, Trento.
- Chamberlain S., Daly A.S., Spalding V. (2011). Assessment and Qualifications Alliance, *Pastoral Care in Education*, vol 29, pp. 193-205.
- Di Nuovo S., Magnano P. (2012). *Competenze trasversali e scelte formative. Strumenti per valutare meta cognizione, motivazione, interessi e abilità sociali nella continuità tra livelli scolastici*, Erickson, Trento.
- Faust M.W., Ashcraft M.H., Fleck D.E. (1996). Mathematics anxiety effects in simple and complex addition. *Mathematical Cognition*. vol. 2, pp. 25-62
- Fondazione Giovanni Agnelli (2014). *La valutazione della scuola. A che cosa serve e perché è necessaria all'Italia*. Editori Laterza, Bari.
- Ray Adams, Comments on Kreiner (2011). Is the foundation under PISA solid? A critical look at the scaling model underlying international comparisons of student attainment, in <https://www.oecd.org/pisa/47681954.pdf>