



## Students' Perceived Invested Effort in the Italian National Assessment of Mathematics

**Daniel Doz**

University of Primorska, Faculty of Education, Slovenia, [doz\\_daniel@yahoo.it](mailto:doz_daniel@yahoo.it)

**Eleonora Doz**

University of Trieste, Department of Life Sciences, Italy, [doz.eleonora@yahoo.it](mailto:doz.eleonora@yahoo.it)

National assessments of mathematics are often used as an objective measurement of students' knowledge. However, various factors are believed to influence students' achievements on the national assessments, among which students' invested effort. Moreover, students' effort may vary between genders and might be influenced by the presence of external motivational factors, such as grades, or test difficulty. In Italy, no systematic research has studied students' perceived effort on the national assessment of the mathematical knowledge, in particular the relation between the invested effort, the test difficulty and the impact of the external motivational factors, such as grading the national assessment. With the present research, we aimed to investigate students' invested effort in the Italian national assessment of mathematics and its relation to the perceived test difficulty, gender and external motivational factors. Through an online questionnaire with 232 participants, we found that 25.4% of the students put little to no effort. Moreover, when asked about how much more effort would students put if the national assessment was graded, 78.5% of the participants stated they would put more effort than they actually did. The majority of the participants (69.8%) argued that they did not want their standardized test to be graded; however, students who put a greater amount of effort in taking the national assessment would like it to influence their final grade. Lastly, we investigated students' feelings about the test difficulty, finding a negative correlation between students' perception of the difficulty of the test and their invested effort: harder tests are solved with lower effort. Overall, this research suggests that we should carefully interpret the results of the Italian national assessment of mathematics and that policymakers should develop

Keywords: mathematics, gender, effort, national assessment, Italy

### INTRODUCTION

The main aim of national and international assessments is to measure students' knowledge, in order to improve the national educational system (Felda, 2018; Gneezy et al., 2019). Measuring students' knowledge is particularly important in mathematics

**Citation:** Doz, D., & Doz, E. (2021). Students' perceived invested effort in the Italian national assessment of mathematics. *International Journal of Instruction*, 14(3), 893-908. <https://doi.org/10.29333/iji.2021.14352a>

since mathematical skills are used in everyday life and recent studies have shown that mathematical abilities predict future financial and educational success, especially for women (Geary et al., 2013). However, students' knowledge of mathematical contents and competencies are not the only factors determining the outcomes in standardized tests: various factors can affect the test performance and impact the validity and reliability of the tests themselves. Among them, there are students' intelligence (Duckworth et al., 2011), domain-general cognitive abilities (Best et al., 2011), emotional factors (Calvin et al., 2010; Strand et al., 2006), students' motivation (Cheng et al., 2014; Dodeen et al., 2014; Duckworth et al., 2011; Finn, 2015; Timmerman et al., 2017) and effort (Stenlund et al., 2018; Eklöf et al., 2014; Silm et al., 2013).

Students' effort plays an important role in determining the outcomes on national assessments of mathematics: higher effort is related to higher results (DeMars et al., 2013; Finn, 2015). The amount of invested effort can vary with external motivation, gender and perceived difficulty of the test.

For what concerns the influence of external motivation on effort, Cankar (2008) stated, that if students do not get any grade for the national assessment, then they will not invest much effort in solving it. Similarly, it has been noticed that if the national assessments are undemanding, students tend to invest less effort (Hoyt, 2001; Simzar et al., 2015; Eklöf et al., 2014). Moreover, Eklöf and Nyroos (2013) found that the majority of the students felt motivated to do their best on the national science test, yet more than 30% of the students felt not motivated and half of them stated that they could have tried harder. Eklöf and colleagues (2014) found a similar gap between students' actual invested effort and the willingness of trying harder on the Trends in International Mathematics and Science Study (TIMSS) standardized tests.

Secondly, the results of studies that considered the role of gender in students' invested effort have been unclear. Specifically, some researchers found that males believe to have put more effort than females in solving standardized tests (Eklöf & Nyroos, 2013), while others reported that women tend to put more effort than men (Butler & Adams, 2007; Cole et al., 2008; Eklöf, 2007; Wise et al., 2009). Moreover, some researchers did not find any difference between girls' and boys' perceived level of invested effort (Stenlund et al., 2017; Eklöf et al., 2014). Furthermore, some researches have homogeneously shown that boys outperform girls in standardized mathematics tests (Else-Quest et al., 2010; Reilly et al., 2019; INVALSI, 2019), which might be the consequence of gender stereotypes in mathematics (Cvencek et al., 2011; Smeding, 2012; Tomasetto et al., 2011; Wang & Degol, 2017; McGuire et al., 2020; Barth & Masters, 2020; Koul et al., 2016) and self-concept (Goldman & Penner, 2016). Traditional math-gender stereotypes suggest that boys and men are more likely to succeed in mathematics, while girls and women are more likely to succeed in language arts subjects. These stereotypes can negatively influence females' math performance (Spencer et al., 1999) and math learning (Appel et al., 2011). Especially in those situations, where math skills are exposed to judgment; for example, a formal standardized test, the stereotype threat activates and women are more likely to have a poorer performance (Spencer et al., 1999).

Similarly, some researches highlighted that girls feel mathematics as more difficult than boys do (Samuelsson & Samuelsson, 2016; Uwineza et al., 2018), which might be also the consequence of implicit gender stereotypes in mathematics (Muzzatti & Agnoli, 2007). The perceived difficulty of the test is furthermore believed to be correlated to students' effort. In particular, it has been shown that students put less effort in more difficult tests (Malmberg et al, 2013).

No studies have explored students' opinions and feelings about the possibility of introducing grades for their national assessments of mathematics. Their opinions are however important since they can increase or decrease the amount of effort they put in solving tests. In the study of Eklöf and Nyroos (2013) there is evidence of how pupils' opinion on the importance of a test influences their effort. Being in favor of grading a standardized test could mean that students give a lot of importance to the test.

Students' achievements on the Italian national assessment of mathematics INVALSI are being used by several institutions and legislators, in order to develop a strategy to increase the quality of the Italian educational system (INVALSI, 2019). No systematic research in Italy has investigated students' effort in mathematics standardized tests. This topic is however relevant since lower effort relates to lower achievements, which consequently could depict a false image about students' knowledge and the quality of the Italian educational system.

In the present work, we aim to contribute to the literature by investigating some of the factors underlying the students' performances in the Italian national assessment of mathematics.

Firstly, being effort one of the factors that influence students' achievements on standardized tests, we aimed to investigate, how much effort Italian students put in solving it. Since the literature reported mixed results regarding gender differences in the perceived level of effort, such variable was also considered.

Secondly, we aimed to validate the hypothesis that students would put more effort, if an external motivational factor was introduced, such as getting a grade on national assessments of mathematics. Gender was controlled as well.

Thirdly, we intended to investigate students' opinions about the INVALSI being part of their final grading in mathematics and its relationship with effort. We hypothesize that students who are more convinced that national assessments should be part of their grades will consequently invest more effort in solving them.

Finally, we aimed to understand, how the perceived difficulty of the national assessment correlates to the invested effort. Furthermore, since girls are believed to feel mathematics tests as more difficult, gender differences were also considered.

### **Research Questions**

Our research questions are hence the following:

RQ1: How do students assess their invested effort on the Italian national assessment of mathematics? (RQ1.1) Are there any differences in the reported effort between genders? (RQ1.2)

RQ2: Would students put more effort if their assessments were graded and influencing their final grades? (RQ2.1) Is there any difference in extra effort between men and women? (RQ2.2)

RQ3: Did students, who put more effort in solving the standardized tests, want their assessments to be graded? (RQ3.1) Is there any difference in this belief between genders? (RQ3.2)

RQ4: Is the perceived difficulty of the assessments correlated to the students' invested effort? (RQ4.1) Are there any gender differences in such opinion? (RQ4.2)

## **METHOD**

### **Research Design**

The present research uses a quantitative and non-experimental methodology. Since the problem of determining the relation between effort, external motivation and perceived difficulty have not been extensively studied yet, a survey was used to gather data about students' opinions on the topic.

### **Population and Sample**

The population used in the present study was secondary school students from the North-Eastern Italian region Friuli-Venezia Giulia.

Participants of the study were 232 students; 114 were women (49.1%). The mean age was 16.8 ( $SD=1.71$ ,  $min=14$ ,  $max=24$ ). 166 (71.6%) students attended a lyceum, 56 (24.1%) a technical school and 10 (4.3%) a vocational school.

Since schools were closed during the COVID-19 sanitary emergency in Italy, we were unable to reach students directly, hence our sample was chosen in the following way: a group of randomly chosen high school mathematics teachers was contacted to participate in the research and the link to the online questionnaire was sent to them. Teachers were instructed to send the link to their students using one of the following channels used during the COVID-19 pandemic: e-mails or online educational platforms (Google Classroom, Edmodo, Moodle...).

Students' academic results were not impacted by their decision to participate in the study.

### **Instrument**

We developed an online questionnaire with four items:

- What do you think about the INVALSI being part of your final grading in mathematics?
- How much do you agree with the sentence "I put very high effort in writing the

INVALSI assessment”?

- How much do you agree with the sentence “The INVALSI assessments are very hard”?
- How much do you agree with the sentence “I would put more effort in solving the national assessment if it impacted my final score in mathematics”?

Participants reported the extent to which they agreed with the statements using a 7-points Likert scale (1=Absolutely contrary; 7=Absolutely in favor).

We decided to use a short questionnaire instead of an articulated one, in order to maximize the number of participants who would complete it (see, e.g. Toepoel et al., 2009) and maximize their focus.

The content validity of the questionnaire was measured on the initial draft of 17 questions regarding students’ effort, perceived difficulty, students’ agreement to include the INVALSI as one of their grades and extra effort students would put if the INVALSI was part of their final mathematics grade. The draft was reviewed by a group of experts in the field of education. Such group was composed of six experts in the field of mathematics education, i.e. two full professors of mathematics education, two independent researchers (one in the field of mathematics education and one in the field of developmental psychology) and two Ph.D. candidates in mathematics education with at least one publication in the field of mathematics education. Experts were instructed to assess the relevancy of the questions with a 4-items Likert scale (1 = not relevant, 4 = very relevant). The scale was dichotomized into “agreed” (ratings 3 and 4) and “not agreed” (ratings 1 and 2). The item level content validity was determined as the number of “agreed” ratings divided by the number of experts. Items were accepted as content-valid if their item content validity was at least .83 (Yusoff, 2019). Experts critically analyzed each item’s content, in order to review the suitability and relevancy of the statements. The questionnaire tool was assessed using the S-CVI/Ave, calculated by averaging the individual I-CVIs. The questionnaire was valid if the S-CVI/Ave was greater or equal to .90 (Polit & Beck, 2006). Hence, four questions were maintained in the final questionnaire.

#### **Procedure of Data Analysis**

Descriptive and inferential statistical methods were used in the present research. Descriptive statistical methods were used to analyze demographic variables. Students’ opinions, expressed with the aid of 7-points Likert scales, were analyzed considering the medians, due to the ordinal nature of the variables. Frequency tables were constructed for each item. Gender differences were investigated by considering the nonparametric Mann-Whitney U test. Correlations were computed using the nonparametric Spearman’s  $\rho$  correlation coefficient.

The gathered data was analyzed with the statistical software *Jamovi*.

## FINDINGS

### Self-Reported Effort

Students reported they invested a considerable amount of effort in solving the national assessment of mathematics ( $Mdn=5.00$ ).

The difference in the self-reported effort between boys ( $Mdn=5.00$ ) and girls ( $Mdn=4.00$ ) is statistically non-significant ( $U=6053$ ;  $p=.181$ ), which indicates that boys and girls have similar beliefs about the effort they put in taking the national assessment. 25.4% of all students put no or very little effort in solving the national assessment, while 53.0% of the students self-rated their effort to be moderate or more (see Table 1).

Table 1  
Students' self-reported effort on the national assessment of mathematics

How much do you agree with the sentence "I put very high effort in writing the INVALSI assessment"?	Boys	Girls	Total
1 (Absolutely contrary)	9 (7.6%)	5 (4.4%)	14 (6.0%)
2 (Contrary)	5 (4.2%)	9 (7.9%)	14 (6.0%)
3 (More contrary)	17 (14.4%)	14 (12.3%)	31 (13.4%)
4 (Neutral)	18 (15.3%)	32 (28.1%)	50 (21.6%)
5 (More in favor)	30 (25.4%)	25 (21.9%)	55 (23.7%)
6 (In favor)	21 (17.8%)	22 (19.3%)	43 (18.5%)
7 (Absolutely in favor)	18 (15.3%)	7 (6.1%)	25 (10.8%)
Total	118 (100.0%)	114 (100.0%)	232 (100.0%)

### External Motivation

Students reported that they would put more effort if the INVALSI assessments were impacting their final mathematics grades ( $Mdn=6.00$ ). When we computed the difference between boys' ( $Mdn=5.60$ ) and girls' ( $Mdn=5.47$ ) self-reported extra effort they would put if national assessments were influencing their grades, we found statistically non-significant differences ( $U=6321$ ;  $p=.311$ ), which indicates that both boys and girls would put almost the same amount of extra effort if the national assessment was part of their final grading.

In Table 2 we present the frequencies about students' opinion on the extra effort students would put in if the INVALSI assessment was part of their final grading. 78.5% of the students would try harder if grades were assigned.

Table 2  
Students' opinions about the extra effort they would put

How much do you agree with the sentence "I would put more effort in solving the national assessment, if it will impact my final score in mathematics"?	Boys	Girls	Total
1 (Absolutely contrary)	10 (8.5%)	3 (2.6%)	13 (5.6%)
2 (Contrary)	2 (1.7%)	6 (5.3%)	8 (3.4%)
3 (More contrary)	1 (0.8%)	7 (6.1%)	8 (3.4%)
4 (Neutral)	10 (8.5%)	11 (9.6%)	21 (9.1%)
5 (More in favor)	19 (16.1%)	21 (18.4%)	41 (17.7%)
6 (In favor)	23 (19.5%)	21 (18.4%)	44 (19.0%)
7 (Absolutely in favor)	53 (44.9%)	44 (36.6%)	97 (41.8%)
Total	118 (100.0%)	114 (100.0%)	232 (100.0%)

### National Assessments as Part of Students' Grade

Students are mainly against the grading of their INVALSI mathematics standardized assessments ( $Mdn=2.00$ ). The differences between males' ( $Mdn=2.85$ ) and females' ( $Mdn=2.20$ ) opinions about the inclusion of the national assessments as part of their final mathematics grades are statistically significant ( $U=5305$ ;  $p=.003$ ). Hence, we conclude that females are more convinced that the INVALSI should not influence their final grading in mathematics than males.

In Table 3 frequencies of the students' opinion whether the INVALSI assessment should be part of their final grading are presented. 69.8% of the students are absolutely more contrary to the fact that national assessment should influence their final grading, while 16.9% are more in favor of including the achievements on the INVALSI assessment among their grades.

Table 3  
Students' opinions about the inclusion of the national assessments as part of their final grading

What do you think about the INVALSI being part of your final grading in mathematics?	Boys	Girls	Total
1 (Absolutely contrary)	44 (37.3%)	65 (57.0%)	109 (47.0%)
2 (Contrary)	13 (11.0%)	11 (9.6%)	24 (10.3%)
3 (More contrary)	18 (15.3%)	11 (9.6%)	29 (12.5%)
4 (Neutral)	19 (16.1%)	12 (10.6%)	31 (13.4%)
5 (More in favor)	13 (11.0%)	8 (7.0%)	21 (9.1%)
6 (In favor)	6 (5.1%)	6 (5.3%)	12 (5.2%)
7 (Absolutely in favor)	5 (4.2%)	1 (0.9%)	6 (2.6%)
Total	118 (100.0%)	114 (100.0%)	232 (100.0%)

A positive correlation between the opinion that the national assessment of mathematics should be part of students' grades and the self-reported effort was found ( $\rho=.212$ ;

$p=.001$ ). That is, students who are more in favor of including the INVALSI assessment as part of their grades, also try harder on the test.

Moreover, using the Wilcoxon test we found that students would put more effort if their assessments were grades ( $Mdn=6.00$ ) than if they were not ( $Mdn=5.00$ ),  $W=2842$ ;  $p<.001$ .

### Perceived Difficulty of the Test

Students felt the mathematics standardized assessment was normal-difficult ( $Mdn=5.00$ ). When students' perceived test difficulty was considered, differences among girls ( $Mdn=4.55$ ) and boys ( $Mdn=4.08$ ) was found ( $U=5529$ ;  $p=.017$ ): girls tend to evaluate the INVALSI test as more difficult than boys. Details are presented in Table 4.

Table 4

Students' opinions about the difficulty of the mathematics national assessment

How much do you agree with the sentence "The INVALSI assessments are very hard"?	Boys	Girls	Total
1 (Absolutely disagree)	11 (9.3%)	5 (4.4%)	16 (6.9%)
2 (Disagree)	10 (8.5%)	8 (7.0%)	18 (7.8%)
3 (More disagree)	19 (16.1%)	11 (9.6%)	30 (12.9%)
4 (Neutral)	26 (22.0%)	23 (20.2%)	49 (21.1%)
5 (More agree)	30 (25.4%)	35 (30.7%)	65 (28.0%)
6 (Agree)	15 (12.7%)	26 (22.8%)	41 (17.7%)
7 (Absolutely agree)	7 (5.9%)	6 (5.3%)	13 (5.6%)
Total	118 (100.0%)	114 (100.0%)	232 (100.0%)

We calculated the Spearman's correlation coefficient between the perceived test difficulty and the invested effort. A negative and statistically significant correlation was found ( $\rho=-.147$ ;  $p=.025$ ). That is, the more difficult the INVALSI test is perceived, the fewer effort students put in solving its tasks.

### DISCUSSION

The theme of students' effort in national assessments in Italy has not been studied extensively yet. In particular, no systematic research was done, in order to analyze students' perceived level of effort and gender differences; nor if grading boosts effort, and also not the influence of tests difficulty on the invested effort. Such research questions are however important, in order to determine the validity and reliability of the assessment itself.

Firstly, we found that students perceived, on average, that they put a middle amount of effort in solving the national assessment of mathematics. No statistically significant differences between boys and girls were found, suggesting that both boys and girls perceived to have put the same amount of effort. The findings are consistent with those found by some earlier researches (Stenlund et al., 2017; Eklöf et al., 2014).



Even though the majority of students perceived that they put a middle amount of effort in solving the standardized test, 25.4% of the students put little to no effort. Since it has been argued that students' invested effort is correlated to their achievements (DeMars et al., 2013; Finn, 2015) and the overall test validity (DeMars et al., 2013), lower effort could lead to lower students' achievements on the INVALSI mathematics tests, which could give educators and policymakers distorted or even false information about the quality of the Italian educational system.

Secondly, we aimed to investigate students' opinions on the national assessments being part of the students' final grading and on the fact, that in this case, students put more effort in solving them. We found that students are mainly against the inclusion of national assessments in their final grades in mathematics. In particular, 47.0% of the students are completely contrary to the inclusion of the achievements on the INVALSI test among the other grades; 69.8% are overall contrary.

When gender was considered, we found a statistically significant difference in opinions between boys and girls: we found that girls are less inclined to make the INVALSI test be part of their final grading. For instance, there were 37.3% of male students that were absolutely contrary to considering the INVALSI as part of their grades, while 57.0% of the female students believed the same. Overall, 63.6% of the male population in the sample and 76.2% of the females in the group were contrary to include the INVALSI as part of the final grading.

Some researchers have found that despite girls have higher school grades than boys in mathematics, they perform worse on standardized tests (Matějů & Smith, 2015; Xhomara, 2018), which could explain why girls try to avoid that their performances on standardized tests would be part of their final grading. On the other hand, this phenomenon could be explained by considering the theory of stereotypes (Cvencek et al., 2011; Smeding, 2012; Tomasetto et al., 2011): women feel less competent in mathematics and do not want that their grade (which is not the straight outcome of their competencies in mathematics, but also class participation, homework, hard work, effort) would be influenced by the standardized test.

Results confirmed the hypothesis that students, who are more convinced that the national assessments should influence their final grading, invest more effort in solving them.

Moreover, our results demonstrated that if the assessments were graded, both boys and girls would put considerably more effort than they did. In particular, 78.5% of the students would put more effort in solving the national assessment, confirming the hypotheses of Simzar et al. (2015). These findings suggest that students' effort largely relies on the external motivation factors, such as grading the assessments.

The findings show that greater caution in interpreting the results of national assessments is required. The fact that the majority of students would put more effort, if the INVALSI assessment would be graded, means that they did not show their real abilities and knowledge, hence the results of the INVALSI assessment might be partially or completely deviating.

As a consequence of our research, we claim that if the assessments were part of the students' final grading, they would try harder and put more effort in solving it. Introducing national assessments that influence students' grades would hence increase their invested effort, therefore increasing the validity and the reliability of the results of the assessments themselves.

Even though grades could motivate students to try harder, we showed that students would be mostly contrary to it. That is, considering the INVALSI assessments as part of the students' final grading could help them to be more motivated and put more effort, giving hence a clearer picture of the Italian educational system; on the other hand, it might induce students to be more anxious or even sabotage the assessments (see DeMars et al., 2013).

Lastly, we aimed to understand whether the students' perception of the difficulty of the national assessment can affect their effort. We found that participants rated the assessment to be "normally" hard. Statistically significant differences between boys and girls were found. Specifically, girls rated the test as more difficult than boys. This could be explained by considering gender stereotypes. Gender stereotypes in mathematics do affect the students' opinion about his/her mathematical competencies, i.e. his/her self-concept (Goldman & Penner, 2016). Lower self-concept is correlated to a poorer academic outcome (Awan et al., 2011; Chiu & Klassen, 2010) and a feeling that the test is harder (Flore et al., 2018; Lindberg et al., 2010).

The testing of the relation between the perceived test difficulty and the students' effort revealed a negative correlation, which confirms the results of previous researches (Malmberg et al., 2013). That is, the more difficult a test is, the fewer effort participants will be willing to make. Hard tests are hence solved with lower effort than easier ones.

## **CONCLUSIONS**

The aim of the present research was to investigate students' effort in the national assessment of mathematics in Italy. Participants put a medium amount of effort and no significant differences between genders were found. A quarter of the students' population put no or little effort in solving the national assessment, hence probably invalidating or misleading the information that teachers, principals, parents and policymakers get.

Furthermore, the findings suggest that students mostly believe that national assessments should not anyhow influence their final grade in mathematics. However, they would be ready to invest more effort in solving the national assessment, if it (would) impacted their grades. That is, if students were externally motivated with grades, they would put more effort in solving national assessments. More effort would also mean a clearer picture of the Italian educational system.

In addition, participants rated the INVALSI difficulty as neutral. The perceived difficulty does slightly negatively correlate with the students' effort. Hence, the more difficult is the test, the fewer effort students put.

Gender was found not to play a significant role in the effort that students put in taking the tests; however, the perceived level of difficulty of the test is higher for women than for men, as well as the belief that the INVALSI assessments should not be part of the students' final grading. Such differences could be explained by considering the theory of gender gap and gender stereotypes in mathematics, yet further research is required, in order to investigate the phenomenon.

Taken together, this research provides evidence that the achievements on the INVALSI tests are not completely reliable. Hence, we suggest that legislators introduce a method to control the invested effort and the difficulty of the national assessments. Furthermore, policymakers should develop a strategy to improve students' effort in solving mathematics standardized tests.

### **LIMITATIONS AND RECOMMENDATIONS**

Our empirical research has some limitations. Firstly, our results are based solely on explicit measures. There could be a discrepancy between what students claim they do and what they actually do; hence, more objective measures are needed. Secondly, we did not investigate some factors, that are believed to affect students' effort on national assessments, such as their personality, beliefs about their competencies, motivation and attitude in mathematics (see e.g., Hemmings & Kay, 2010). Lastly, the question of whether gender stereotypes in mathematics influence effort, opinions about mathematics and perceived difficulty of the test is left unanswered by the present study. Further research could investigate if implicit gender stereotypes are causally related to the effort.

Nevertheless, our work showed that some students tended to invest little or no effort in solving the national mathematics assessment. Being it related to lower achievement, we recommend legislators, school principals, teachers and parents to interpret correctly the results of national assessments, since, as found by our work, some students tend to put no or very little effort.

A method of weighting correctly the students' results, who put low or no effort in solving the national assessment, should be hence developed. We suggest that the INVALSI institute would create a questionnaire for students, where they could express their feelings about and during the assessment itself, in particular to measure their effort in solving the exercises (see, e.g. Yunus & Ali, 2009; or Chouinard et al., 2007). Nevertheless, more objective measurements about students' effort in solving national assessments should be developed.

For what it concerns the idea of grading national assessments, greater anxiety could be derived. Hence, additional research is needed, in order to investigate the phenomenon and, consequently, develop a more efficient way of assessing students' knowledge of mathematics.

Students' perceived difficulty is negatively related to students' effort. In order to correctly interpret students' mathematical achievements, the INVALSI Institute could include not only a questionnaire about the students' effort, but also the perceived difficulty of the test, in order to adjust the interpretation of the found results. Indeed,

students might solve difficult tests with less or no effort, imagining that the test is an excessively big challenge, misrepresenting consequently the picture of the quality of the national educational system. However, because the found correlation is weak, we suggest further investigations to replicate the relationship between perceived difficulty and effort, using a larger sample and investigating those variables during the solving of the national assessment.

## REFERENCES

- Appel, M., Kronberger, N., & Aronson, J. (2011). Stereotype threat impairs ability building: Effects on test preparation among women in science and technology. *European Journal of Social Psychology, 41*(7), 904-913.
- Awan, R. U. N., Noureen, G., & Naz, A. (2011). A Study of Relationship between Achievement Motivation, Self Concept and Achievement in English and Mathematics at Secondary Level. *International Education Studies, 4*(3), 72-79.
- Barth, J. M., & Masters, S. (2020). Changes in Math and Science Interest over School Transitions: Relations to Classroom Quality, Gender Stereotypes, and Efficacy. *International Journal of Gender, Science & Technology, 12*(1).
- Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and individual differences, 21*(4), 327-336.
- Butler, J., & Adams, R. J. (2007). The impact of differential investment of student effort on the outcomes of international studies. *Journal of Applied Measurement, 8*, 279-304.
- Calvin, C. M., Fernandes, C., Smith, P., Visscher, P. M., & Deary, I. J. (2010). Sex, intelligence and educational achievement in a national cohort of over 175,000 11-year-old schoolchildren in England. *Intelligence, 38*(4), 424-432.
- Cankar, G. (2008). Nacionalno preverjanje znanja kot pripomoček za izboljšanje poučevanja. *Sodobna pedagogika/Posebna izdaja, 59*, 130-141.
- Cheng, L., Klinger, D., Fox, J., Doe, C., Jin, Y., & Wu, J. (2014). Motivation and test anxiety in test performance across three testing contexts: The CAEL, CET, and GEPT. *Tesol Quarterly, 48*(2), 300-330.
- Chiu, M. M., & Klassen, R. M. (2010). Relations of mathematics self-concept and its calibration with mathematics achievement: Cultural differences among fifteen-year-olds in 34 countries. *Learning and Instruction, 20*(1), 2-17.
- Chouinard, R., Karsenti, T., & Roy, N. (2007). Relations among competence beliefs, utility value, achievement goals, and effort in mathematics. *British journal of educational psychology, 77*(3), 501-517.
- Cole, J. S., Bergin, D. A., & Whittaker, T. A. (2008). Predicting student achievement for low stakes tests with effort and task value. *Contemporary Educational Psychology, 33*, 609-624.

- Cvencek, D., Meltzoff, A. N., & Greenwald, A. G. (2011). Math–gender stereotypes in elementary school children. *Child development, 82*(3), 766-779.
- DeMars, C. E., Bashkov, B. M., & Socha, A. B. (2013). The Role of Gender in Test-Taking Motivation under Low-Stakes Conditions. *Research & Practice in Assessment, 8*, 69-82.
- Dodeen, H. M., Abdelfattah, F., & Alshumrani, S. (2014). Test-taking skills of secondary students: the relationship with motivation, attitudes, anxiety and attitudes towards tests. *South African Journal of Education, 34*(2).
- Duckworth, A. L., Quinn, P. D., Lynam, D. R., Loeber, R., & Stouthamer-Loeber, M. (2011). Role of test motivation in intelligence testing. *Proceedings of the National Academy of Sciences, 108*(19), 7716-7720.
- Eklöf, H. (2007). Test-taking motivation and mathematics performance in TIMSS 2003. *International Journal of Testing, 7*, 311–326.
- Eklöf, H., & Nyroos, M. (2013). Pupil perceptions of national tests in science: perceived importance, invested effort, and test anxiety. *European journal of psychology of education, 28*(2), 497-510.
- Eklöf, H., Pavešič, B. J., & Grønmo, L. S. (2014). A cross-national comparison of reported effort and mathematics performance in TIMSS Advanced. *Applied Measurement in Education, 27*(1), 31-45.
- Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: a meta-analysis. *Psychological bulletin, 136*(1), 103.
- Felda, D. (2018). Preverjanje matematičnega znanja. *Journal of Elementary Education, 11*(2), 175-188.
- Finn, B. (2015). Measuring motivation in low-stakes assessments. *ETS Research Report Series, 2015*(2), 1-17.
- Flore, P. C., Mulder, J., & Wicherts, J. M. (2018). The influence of gender stereotype threat on mathematics test scores of Dutch high school students: A registered report. *Comprehensive Results in Social Psychology, 3*(2), 140-174.
- Geary, D. C., Hoard, M. K., Nugent, L., & Bailey, D. H. (2013). Adolescents' functional numeracy is predicted by their school entry number system knowledge. *PloS one, 8*(1), e54651.
- Gneezy, U., List, J. A., Livingston, J. A., Qin, X., Sadoff, S., & Xu, Y. (2019). Measuring success in education: the role of effort on the test itself. *American Economic Review: Insights, 1*(3), 291-308.
- Goldman, A. D., & Penner, A. M. (2016). Exploring international gender differences in mathematics self-concept. *International Journal of Adolescence and Youth, 21*(4), 403-418.

- Hemmings, B., & Kay, R. (2010). Prior achievement, effort, and mathematics attitude as predictors of current achievement. *The Australian Educational Researcher*, 37(2), 41-58.
- Hoyt, J. E. (2001). Performance funding in higher education: The effects of student motivation on the use of outcomes tests to measure institutional effectiveness. *Research in Higher Education*, 42, 71-85.
- INVALSI (2019). Retrieved from: [https://invalsi-areaprove.cineca.it/docs/2019/Rapporto\\_prove\\_INVALSI\\_2019.pdf](https://invalsi-areaprove.cineca.it/docs/2019/Rapporto_prove_INVALSI_2019.pdf) (22. 6. 2020).
- Koul, R., Lerdpornkulrat, T., & Poondej, C. (2016). Gender compatibility, math-gender stereotypes, and self-concepts in math and physics. *Physical Review Physics Education Research*, 12(2), 020115.
- Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: a meta-analysis. *Psychological bulletin*, 136(6), 1123.
- Malmberg, L. E., Walls, T. A., Martin, A. J., Little, T. D., & Lim, W. H. (2013). Primary school students' learning experiences of, and self-beliefs about competence, effort, and difficulty: Random effects models. *Learning and individual differences*, 28, 54-65.
- Matějů, P., & Smith, M. L. (2015). Are boys that bad? Gender gaps in measured skills, grades and aspirations in Czech elementary schools. *British Journal of Sociology of Education*, 36(6), 871-895.
- McGuire, L., Mulvey, K. L., Goff, E., Irvin, M. J., Winterbottom, M., Fields, G. E., Hartstone-Rose, A. & Rutland, A. (2020). STEM gender stereotypes from early childhood through adolescence at informal science centers. *Journal of applied developmental psychology*, 67, 101109.
- Muzzatti, B., & Agnoli, F. (2007). Gender and mathematics: Attitudes and stereotype threat susceptibility in Italian children. *Developmental psychology*, 43(3), 747.
- Polit, D. F., & Beck, C. T. (2006). The content validity index: are you sure you know what's being reported? Critique and recommendations. *Research in nursing & health*, 29(5), 489-497.
- Reilly, D., Neumann, D. L., & Andrews, G. (2019). Investigating gender differences in mathematics and science: Results from the 2011 Trends in Mathematics and Science Survey. *Research in Science Education*, 49(1), 25-50.
- Samuelsson, M., & Samuelsson, J. (2016). Gender differences in boys' and girls' perception of teaching and learning mathematics. *Open Review of Educational Research*, 3(1), 18-34.
- Silm, G., Must, O., & Täht, K. (2013). Test-Taking Effort as a Predictor of Performance in Low-stakes Tests. *TRAMES: A Journal of the Humanities & Social Sciences*, 17(4).

- Simzar, R. M., Martinez, M., Rutherford, T., Domina, T., & Conley, A. M. (2015). Raising the stakes: How students' motivation for mathematics associates with high-and low-stakes test achievement. *Learning and individual differences, 39*, 49-63.
- Smeding, A. (2012). Women in science, technology, engineering, and mathematics (STEM): An investigation of their implicit gender stereotypes and stereotypes' connectedness to math performance. *Sex roles, 67*(11-12), 617-629.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of experimental social psychology, 35*(1), 4-28.
- Stenlund, T., Eklöf, H., & Lyrén, P. E. (2017). Group differences in test-taking behaviour: An example from a high-stakes testing program. *Assessment in Education: Principles, Policy & Practice, 24*(1), 4-20.
- Stenlund, T., Lyrén, P. E., & Eklöf, H. (2018). The successful test taker: exploring test-taking behavior profiles through cluster analysis. *European Journal of Psychology of Education, 33*(2), 403-417.
- Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK national picture. *British Journal of Educational Psychology, 76*(3), 463-480.
- Timmerman, H. L., Toll, S. W., & Van Luit, J. E. (2017). The relation between math self-concept, test and math anxiety, achievement motivation and math achievement in 12 to 14-year-old typically developing adolescents. *Psychology, Society, & Education, 9*(1), 89-103.
- Toepoel, V., Das, M., & Van Soest, A. (2009). Design of web questionnaires: The effects of the number of items per screen. *Field Methods, 21*(2), 200-213.
- Tomasetto, C., Alparone, F. R., & Cadinu, M. (2011). Girls' math performance under stereotype threat: The moderating role of mothers' gender stereotypes. *Developmental psychology, 47*(4), 943.
- Uwineza, I., Rubagiza, J., Hakizimana, T., & Uwamahoro, J. (2018). Gender attitudes and perceptions towards mathematics performance and enrolment in Rwandan secondary schools. *Rwandan Journal of Education, 4*(2), 44-56.
- Wang, M. T., & Degol, J. L. (2017). Gender gap in science, technology, engineering, and mathematics (STEM): Current knowledge, implications for practice, policy, and future directions. *Educational psychology review, 29*(1), 119-140.
- Wise, S., Pastor, D. A., & Kong, X. (2009). Correlates of rapid-guessing behavior in low stakes testing: Implications for test development and measurement practice. *Applied Measurement in Education, 22*(2), 185-205.
- Xhomara, N. (2018). Relationship between gender and academic performance of reading, writing and literature, mathematics and science. *Polis, 17*(1), 96-121.
- Yunus, A. S., & Ali, W. Z. W. (2009). Motivation in the Learning of Mathematics.

*European Journal of Social Sciences*, 7(4), 93-101.

Yusoff, M. S. B. (2019). ABC of content validation and content validity index calculation. *Educational Resource*, 11(2). 49-54