European Journal of Science and Mathematics Education, 2024, **12**(1), 85-96
ISSN: 2301-251X (Online)

OPEN ACCESS

**Research Article**

# The cognitive reflection test and students' achievements in mathematics and physics

**Daniel Doz** [1*]
0000-0002-6942-6937

**Josip Sliško** [2]
0000-0002-5805-4808

[1] Faculty of Education, University of Primorska, Koper, SLOVENIA
[2] Faculty of Mathematical and Physical Sciences, Benemérita Universidad Autónoma de Puebla, Puebla, MEXICO
[*] Corresponding author: daniel.doz@pef.upr.si

**Citation:** Doz, D., & Sliško, J. (2024). The cognitive reflection test and students' achievements in mathematics and physics. *European Journal of Science and Mathematics Education, 12*(1), 85-96. https://doi.org/10.30935/scimath/13832

**ARTICLE INFO**

Received: 20 Aug 2023
Accepted: 18 Oct 2023

**ABSTRACT**

The cognitive reflection test (CRT) assesses an individual's capacity to restrain impulsive and intuitive responses and to engage in critical reflection on mathematical problems. The literature indicates that several factors influence students' performance on CRT, including gender, age, and prior knowledge of mathematics. In this study, our objective was to investigate the correlation between CRT scores and students' achievements in both mathematics and physics. We conducted our research with a sample of 150 Italian high school students, and the findings revealed a positive predictive relationship between CRT scores and students' performance in both mathematics and physics. Furthermore, we employed an ordinal logistic regression to evaluate the impact of CRT scores, gender, and school level on students' achievements in mathematics and physics. The results showed that both CRT scores and school level had statistically significant effects on predicting these achievements. In contrast, gender emerged as a statistically significant factor only in predicting students' mathematics achievements.

**Keywords:** achievements, cognitive reflection test, mathematics, ordinal logistic regression, physics

## INTRODUCTION

The cognitive reflection test (CRT) (Frederick, 2005) is frequently employed in research as a measure of cognitive and intuitive reflection (Stieger & Reips, 2016). The original CRT by Frederick (2005) comprises three mathematical problems, each of which initially elicits an intuitive but incorrect response. These problems can only be solved through rational deliberation, leading to a non-obvious correct answer. An example of those problems is the following (Frederick, 2005):

> A bat and a ball cost $1.10 in total. The bat costs $1.00 more than the ball. How much does the ball cost? _____ cents.

Typically, participants provide incorrect response of "10 cents," whereas the correct answer is five cents. The underlying concept of CRT is rooted in the existence of two distinct cognitive processes:

(1) a rapid, intuitive approach to problem-solving and

(2) a slower, reflective method (Epstein, 1994; Stanovich & West, 2000).

To successfully solve CRT problems, students must suppress their intuitive instincts and instead employ careful reflection. These theories, often referred to as "dual-coding" or "dual-processing" theories (Fang et al., 2019; Lem et al., 2015), propose the presence of two distinct types of thinking: system 1 (intuitive) and system 2 (deliberative). System 1 is fast, unconscious, associative, and not reliant on working memory. It allows

individuals to quickly access intuitive responses, which can be valid but also prone to errors. Conversely, system 2 is slow, conscious, controlled, and strongly associated with an individual's working memory, as well as their thinking dispositions or mental styles. Working memory is a key feature of analytical system 2 processing. While system 2 processing is necessary for solving complex thinking and reasoning problems, it alone does not guarantee correct responses. Most dual-processing theories posit that system 1 processing generates intuitive responses that subsequent system 2 deliberation may or may not modify. Stanovich et al. (2011) argue that deliberative reasoning requires overriding system 1, and this process demands executive control and an inclination to actively think and resist hasty problem-solving. Therefore, executive control processes play a pivotal role in analytical system 2 processes (see De Neys & Glumicic, 2008; Evans, 2009; Thompson, 2009). System 2 function is linked to general measures of cognitive ability, such as intelligence quotient, while system 1 function is not (Stanovich, 1999). Dual-processing theories of thinking have been primarily developed in the context of deductive reasoning tasks, and there is substantial evidence supporting the relevance of both thinking systems in propositional and syllogistic reasoning (see Barrouillet, 2011; De Neys, 2006; Evans, 2008).

Frederick (2005) noted that CRT scores exhibit a strong correlation with various standardized math tests, such as the American college testing and scholastic aptitude test. Additionally, his research demonstrated that males tend to outperform females on CRT. These gender-related differences have been confirmed in more recent studies as well (Campitelli & Gerrans, 2014; Primi et al., 2018; Ring et al., 2016; Zhang et al., 2016). One possible explanation for these gender differences in CRT performance is that females often experience higher levels of math-related anxiety, which can lead to lower academic achievements (see Louis & Mistele, 2012; Zhang et al., 2016).

Another factor associated with CRT performance is age. Several studies have indicated that older students tend to perform better than their younger counterparts (Stieger & Reips, 2016; Young et al., 2018). Older participants might be more familiar with the test (Stieger & Reips, 2016) or possess superior inhibition skills compared to younger individuals (Gilmore et al., 2015).

CRT is linked to both cognitive abilities and rational-thinking skills (Toplak et al., 2011). Furthermore, CRT measures mathematical ability and the disposition toward actively open-minded thinking (Campitelli & Gerrans, 2014). Indeed, CRT has been shown to predict students' mathematics achievement (Frosch & Simms, 2015; Gómez-Chacón et al., 2014; Gómez-Veiga et al., 2018) as well as their overall understanding of science. For example, Shtulman and McCallum (2014) explored the relationship between CRT and students' comprehension of science across various domains (geology, mechanics, astronomy, evolution, thermodynamics, etc.) and found a positive correlation between students' science understanding and CRT scores.

Moreover, studies have confirmed the role of cognitive reflection in studying and understanding physics (Gette & Kryjevskaia, 2019; Kryjevskaia et al., 2020; Speirs et al., 2021; Wood et al., 2016). It is believed that the cognitive challenges presented by CRT are analogous to those encountered by physics learners (Kryjevskaia et al., 2020; Wood et al., 2016). Students often provide naive or intuitive answers to certain physics questions based on their prior experiences, which are scientifically incorrect. Research has highlighted that students with higher cognitive reflection skills (i.e., higher CRT scores) are more likely to arrive at correct answers by engaging in analytical processes effectively. Therefore, there is a positive relationship between cognitive reflection skills and students' reasoning abilities in physics (Gette & Kryjevskaia, 2019; Kryjevskaia et al., 2020). As such, CRT has been suggested as a valuable diagnostic tool for identifying students who may struggle with physics learning, leading to the development of instructional interventions aimed at strengthening their reasoning skills (Gette & Kryjevskaia, 2019; Kryjevskaia et al., 2020).

Given that the original CRT consists of just three items, its reliability has been reported as relatively low, with Cronbach's alpha values ranging between .60 and .74 (Blacksmith et al., 2019; Campitelli & Gerrans, 2014; Primi et al., 2018; Stieger & Reips, 2016). This limitation has led to the development and use of various adapted versions of CRT, which include more items and exhibit higher reliability (Toplak et al., 2014). Despite ongoing methodological debates, the original CRT remains widely utilized in research (Stieger & Reips, 2016). Furthermore, CRT has transcended academic research and gained popularity on the internet, making it familiar to a broader audience (Brañas-Garza et al., 2019; Stieger & Reips, 2016).

In this current study, our objective was to make a contribution to the existing literature by examining the relationship between CRT and the academic achievements of high school students across various school levels, ranging from grade 9 to grade 13. We aimed to investigate the impact of CRT on both mathematics and physics performance. While previous research has extensively explored the influence of CRT on students' performance in either mathematics or physics, there has been limited investigation into its simultaneous effects on the academic achievements of high school students in both subjects. Our study sought to ascertain how effectively CRT predicts students' performance in mathematics and physics. To accomplish this, we considered students' gender and school levels to determine whether these variables had any influence on students' academic performances.

## EMPIRICAL RESEARCH

### Research Hypotheses

Drawing upon the previously mentioned studies, we aimed to test the following hypotheses:

**H1.** Male students exhibit superior performance on CRT compared to female students.

**H2.** Students in grade 13 demonstrate higher performance on CRT in comparison to grade 9 students.

**H3.** The scores achieved by students on CRT serve as predictors for their achievements in mathematics.

**H4.** The scores obtained by students on CRT are predictive of their achievements in physics.

### Methodology

In the present research, the non-experimental quantitative research method was applied.

### Participants

The study involved 150 high school students hailing from North-Eastern Italy. All of these students were enrolled in a scientific lyceum following the "applied sciences curriculum." Out of the total participants, 44 individuals (29.3%) were female, while 106 (70.7%) were male. It is worth noting that this gender distribution is reflective of scientific lyceums with the "applied sciences curriculum"; however, caution should be exercised when extending these findings to the entire population of high school students.

The students were distributed across different school levels, as follows: 64 students (42.7%) from level 9, 44 students (29.3%) from level 11, and 42 students (28.0%) from level 13. The average age of the participants was mean [M]=16.0 (standard deviation [SD]=1.75) years, with the youngest participant being 14 years old and the oldest 20 years old. The median age was 16. According to school records, the socioeconomic status of the sample predominantly fell within the middle class category. Importantly, all participants were typically developing.

### Measures

#### *Cognitive reflection test*

In this study, we utilized a modified version of CRT originally developed by Toplak et al. (2014), with language adjustments tailored to the Italian context. For instance, the term "dollar, $" was translated to "Euro, €." The instrument can be found in **Appendix A** (also refer to Etcheverry et al., 2020). This revised version expanded upon the original three-item CRT by Frederick (2005) by incorporating four additional questions. Consequently, the instrument comprised a total of seven questions. An example question is, as follows: "A pen and a notebook cost €1.10 in total. The notebook costs one Euro more than the pen. How much does the pen cost?" In each question, participants were required to inhibit their intuitive response to make way for the correct answer. CRT scores were calculated by summing the points awarded for each item (zero points for an incorrect response and one point for a correct response), resulting in a potential score range from zero to seven.

The translation and adaptation of the questionnaire from English to Italian were carried out by an English teacher. Subsequently, three high school mathematics teachers independently reviewed and assessed the

**Table 1.** Descriptive statistics for students' mathematics & physics grade, & CRT achievements

| | Mean | Standard deviation | Minimum | Maximum | Median | Skewness | Kurtosis | Shapiro-Wilk W |
|---|---|---|---|---|---|---|---|---|
| Mathematics | 7.67 | 1.94 | 2 | 10 | 8 | -.691 | -.224 | .900*** |
| Physics | 7.55 | 1.60 | 2 | 10 | 8 | -.764 | 1.57 | .916*** |
| CRT | 3.19 | 1.89 | 0 | 7 | 3 | .293 | -.773 | .944*** |

Note. ***p<.001

translated version. They reached unanimous consensus on the final Italian translation with a perfect inter-rater agreement of 100% (Cohen's κ=1), affirming the suitability of the instrument for use in the research.

The reliability of this adapted CRT version was measured to be α=.551, which is comparable to the Cronbach's alpha values reported in other studies (.60; Campitelli & Gerrans, 2014; cf. Stieger & Reips, 2016).

### *Mathematics & physics achievement*

The achievement measures consisted of the students' final grades in mathematics and physics, which were determined at the conclusion of the first school term. These final grades represent numerical scores that evaluate their performance and knowledge in these two subjects. The grades in mathematics and physics were obtained by reviewing the official school records and were graded on a scale ranging from one to 10 points for each subject, where scores from six to 10 were considered passing grades.

### Procedure

CRT was administered to the participants during their regular school classes. They were given specific instructions to complete the test within a time limit of 20 minutes. Following the completion of CRT, participants were asked to respond to demographic questions. In a separate session, we contacted the mathematics and physics teachers to obtain the students' academic achievements in those respective subjects.

### Data Analysis

All data were analyzed using Jamovi statistical software, incorporating a combination of descriptive and inferential statistical methods. The normality of the data distribution was assessed using Shapiro-Wilk W test. To identify differences between two categories, Mann-Whitney U test was employed, while differences among three or more groups were assessed using Kruskal-Wallis $\chi^2$ test. Subsequently, for post-hoc comparisons, Dwass-Steel-Critchlow-Fligner (DSCF) pairwise comparisons test was applied. Correlations were calculated using Spearman's $\rho$ coefficient. Whenever possible, effect sizes were presented in the form of the point-biserial $r$ coefficient for the Mann-Whitney test and $\varepsilon^2$ coefficient for Kruskal-Wallis test.

Furthermore, to evaluate the predictive capacity of various factors on students' achievements in mathematics and physics, an ordinal logistic regression model was employed, following the methodology proposed by Liu and Koirala (2012) and Warner (2008). The fit of the model was assessed using McFadden's pseudo-$R^2$ coefficient (McFadden, 1979), where pseudo-$R^2$ values between .2 and .4 indicate an excellent fit of the model (Zhou et al., 2018).

## RESULTS

In **Table 1**, we provide descriptive statistics for students' mathematics and physics grades, along with their scores on CRT.

As can be observed in **Table 1**, both mathematics and physics grades, as well as students' performance on CRT, exhibited significant deviations from a normal distribution. Therefore, non-parametric statistical tests were employed.

### Gender Differences in Cognitive Reflection Test

In **Table 2**, we present the descriptive statistics for mathematics and physics achievements, as well as CRT performance, categorized by gender.

**Table 2.** Gender differences in mathematics & physics grades & CRT

| | Gender | Mean | Standard deviation | Minimum | Maximum | Median |
|---|---|---|---|---|---|---|
| Mathematics | Male | 7.55 | 1.97 | 2 | 10 | 8 |
| | Female | 7.95 | 1.85 | 5 | 10 | 8 |
| Physics | Male | 7.50 | 1.51 | 2 | 10 | 8 |
| | Female | 7.66 | 1.58 | 6 | 10 | 7.5 |
| CRT | Male | 3.39 | 1.81 | 0 | 7 | 3 |
| | Female | 2.73 | 2.00 | 0 | 5 | 2 |

**Table 3.** Students' achievements in mathematics & physics & CRT achievements in different years of high school

| | School level | Mean | Standard deviation | Minimum | Maximum | Median |
|---|---|---|---|---|---|---|
| Mathematics | 9 | 8.03 | 2.06 | 2 | 10 | 9 |
| | 11 | 8.29 | 1.46 | 5 | 10 | 9 |
| | 13 | 6.36 | 1.53 | 5 | 10 | 6 |
| Physics | 9 | 7.88 | 1.98 | 2 | 10 | 8 |
| | 11 | 6.98 | 1.05 | 5 | 10 | 7 |
| | 13 | 7.64 | 1.27 | 6 | 10 | 7 |
| CRT | 9 | 2.80 | 1.60 | 0 | 6 | 3 |
| | 11 | 3.57 | 2.19 | 0 | 7 | 3 |
| | 13 | 3.40 | 1.89 | 0 | 7 | 3.5 |

When considering students' gender, it was observed that there were no significant differences in students' mathematics grades (U=2068; p=.269; r=.113) or physics grades (U=2281; p=.831; r=.022) between genders. However, a significant difference emerged in CRT performance (U=1843; p=.041; r=.210), where boys outperformed girls, indicating higher scores on CRT among boys compared to girls.

## Age Differences in Cognitive Reflection Test

**Table 3** presents students' achievements in mathematics and physics, as well as their scores on CRT, categorized by different grades.

Significant differences were observed in students' achievements across different school levels for both mathematics ($\chi^2$[2]=29.8; p<.001; $\varepsilon^2$=.200) and physics ($\chi^2$[2]=17.4; p<.001; $\varepsilon^2$=.117). Specifically, post-hoc analysis using DSCF test revealed that students from school level 9 achieved similar scores in mathematics as their peers from school level 11 (W=.68; p=.880). However, they demonstrated higher achievements in physics (W=-5.80; p<.001). On the other hand, students from level 11 outperformed students from level 13 in mathematics (W=-7.13; p<.001) but had lower achievements in physics (W=3.06; p=.077).

In contrast, students' performances on CRT did not exhibit significant differences among the different school levels ($\chi^2$[2]=4.15; p=.126; $\varepsilon^2$=.028). This indicates that students from level 9 achieved similar scores on CRT as level 11 students (W=2.43; p=.199) and level 13 students (W=2.39; p=.209), while level 11 students had comparable CRT scores to level 13 students (W=-.291; p=.977).

## Cognitive Reflection Test

An analysis of individual responses is presented in **Table 4** (refer to **Appendix A** for details). As indicated in **Table 4**, the item that elicited the most intuitive responses from students was the first one, namely, the "pen and notebook" question. A closer analysis of the frequencies of the obtained scores has shown that the proportions are not homogeneous ($\chi^2$[7]=37.0; $p$<.001).

## Relation Between Students' Achievements & Cognitive Reflection Test

In order to examine the relationship between cognitive reflection abilities and students' achievements, an ordinal logistic regression was utilized. The analysis aimed to predict students' mathematics grades while considering the independent variables of CRT abilities, gender, and students' class. In the initial step of the model, which sought to predict mathematics achievement based on CRT results, statistical significance was observed ($\chi^2$[1]=21.0; p<.001; $R^2$=.037). Specifically, CRT (B=.370; SE=.084) was found to significantly predict students' mathematics grades (Z=4.40; p<.001).

**Table 4.** Correct & incorrect answers in CRT

| Item | Wrong answer (intuitive) | Wrong answer (other) | Correct answer | Missing answer |
|---|---|---|---|---|
| 1 | 89 (59.3%) | 6 (4.0%) | 46 (30.7%) | 9 (6.0%) |
| 2 | 54 (36.0%) | 12 (8.0%) | 84 (56.0%) | 0 (0.0%) |
| 3 | 51 (34.0%) | 12 (8.0%) | 72 (48.0%) | 15 (10.0%) |
| 4 | 17 (11.3%) | 52 (34.7%) | 66 (44.0%) | 15 (10.0%) |
| 5 | 42 (28.0%) | 55 (36.7%) | 39 (26.0%) | 14 (9.3%) |
| 6 | 52 (34.7%) | 18 (12.0%) | 74 (49.3%) | 6 (4.0%) |
| 7 | 31 (20.7%) | 3 (2.0%) | 95 (63.3%) | 21 (14.0%) |

In the second step, gender was added to the model to assess changes in predictions. This extended model demonstrated a good fit ($\chi^2$[2]=24.8; p<.001; $R^2$=.044), and the difference from the previous model was statistically significant ($\chi^2$[1]=3.88; p=.049). Gender (with differences favoring females; B=-.624; SE=.324) significantly influenced students' mathematics grades (Z=-1.96; p=.050).

The third step involved adding students' school levels to further explore the prediction. This expanded model also exhibited a good fit ($\chi^2$[4]=66.7; p<.001; $R^2$=.119), with a significant difference compared to the second step ($\chi^2$[2]=41.92; p<.001). Notably, there was no significant difference between year-9 and year-11 students (B=-.128; SE=.359; Z=-.357; p=.721). However, a significant difference was found between year-9 and year-13 students in their mathematics achievements (B=-2.360; SE=-.409; Z=-5.788; p<.001).

In summary, in this model, CRT significantly predicted students' mathematics achievements (B=.472; Z=5.451; p<.001; OR=1.603; 95% confidence interval [CI] [1.358, 1.908]). Gender also had a significant impact, favoring girls (B=-.673; Z=-1.971; p=.049; OR=.510; 95% CI [.259, .990]), while students' class (specifically, year-9 and year-13) significantly influenced mathematics achievements (B=-2.360; Z=-5.766; p<.001; OR=.095; 95% CI [.042,.207]). Overall, this model explained approximately 12.0% of the variance in the outcome (McFadden's pseudo-$R^2$=.119).

Similarly, the impact of these factors on students' physics achievements was examined. In the initial step, aiming to predict physics achievement using CRT results, statistical significance was observed ($\chi^2$[1]=17.9; p<.001; $R^2$=.034). CRT (B=.323; SE=.078) significantly predicted students' physics grades (Z=4.13; p<.001).

In the second step, gender was included to assess any changes in predictions. This extended model exhibited a good fit ($\chi^2$[2]=18.1; p<.001; $R^2$=.034), with no statistically significant difference from the previous model ($\chi^2$[1]=.161; p=.689). Gender (with no significant impact; B=-.133; SE=.333) did not influence students' physics grades (Z=-.40; p=.698).

The third step involved adding students' school levels to further explore the prediction. This expanded model also demonstrated a good fit ($\chi^2$[4]=48.8; p<.001; $R^2$=.092), with a significant difference compared to the second step ($\chi^2$[2]=30.70; p<.001). A significant difference was found between year-9 and year-11 students (B=-2.064; SE=.386; Z=-5.346; p<.001), and a significant difference was also observed between year-9 and year-13 students in their physics achievements (B=-1.108; SE=.384; Z=-2.889; p=.004).

In summary, in this model, CRT significantly predicted students' physics achievements (B=.430; Z=5.323; p<.001; OR=1.537; 95% CI [1.315, 1.807]). Gender had no statistically significant impact (B=-.128; Z=-.389; p=.697; OR=.880; 95% CI [.461, 1.676]), while students' class (year-9 and year-13) significantly influenced physics achievements (as detailed earlier). This model accounted for approximately 9.0% of the variance in the outcome (McFadden's pseudo-$R^2$=.092).

## DISCUSSION & CONCLUSIONS

The relationship between CRT and several factors, such as standardized math tests (Frederick, 2005), students' gender (Campitelli & Gerrans, 2014; Primi et al., 2018; Ring et al., 2016; Zhang et al., 2016), and age (Gilmore et al., 2015; Stieger & Reips, 2016; Young et al., 2018), has been explored in various research studies. In the present research, we investigated the impact of these factors on both mathematics and physics achievements. Specifically, we hypothesized that CRT, along with gender and students' school levels, would be significant predictors of students' math and physics achievements.

Firstly, we examined differences between boys and girls. In the present study, boys and girls exhibited similar mathematics and physics achievements. Although this result aligns with some studies (Hyde et al., 1990; Riegle-Crumb & Moore, 2014), other research has shown that high school boys tend to achieve higher scores in mathematics and physics (Akpotor & Egbule, 2020). One possible explanation for the absence of gender differences in mathematics and physics achievement could be attributed to the non-homogeneous distribution of the sample. Specifically, there were more male participants than female students, making it challenging to generalize the results to the entire student population. Additionally, participants were students in scientific lyceums, who might be more motivated to pursue scientific careers and might have a greater affinity for science compared to students attending non-scientific lyceums or other types of high schools. Consequently, the generalizability of the results cannot be guaranteed.

Furthermore, the results indicated that boys outperformed girls in CRT. This study thus corroborates findings present in the existing literature (Campitelli & Gerrans, 2014; Primi et al., 2018; Ring et al., 2016; Zhang et al., 2016). While additional research is needed to fully comprehend the underlying reasons for these findings, some studies have suggested that girls tend to experience higher levels of anxiety than boys (Zhang et al., 2016), which could potentially hinder their actual abilities. Thus, we confirm **H1**, and additional qualitative and quantitative research is required to further investigate this phenomenon.

Our second hypothesis aimed to explore whether students in grade 13, i.e., older students, would have higher scores on CRT than students in grade 9, i.e., younger students. However, the results did not support this hypothesis, as students' achievements in grades 9, 11, and 13 were similar, with no statistically significant differences among them. These results contrast with those found in other studies (Stieger & Reips, 2016; Young et al., 2018). One potential explanation for these differing findings is the type of school attended by the participants. Since the participants were students in a scientific lyceum, a pre-selection of students may have led to a lack of differences in reasoning abilities due to age. Additionally, since students in this school type engage with mathematics, physics, computer sciences, and science on a daily basis, they may have developed strong reasoning skills from the early years of high school. Nevertheless, increasing the sample size could potentially reveal statistically significant differences in students' CRT scores among the three considered grades. Thus, we reject **H2**.

Furthermore, descriptive statistics have highlighted that the majority of students had significant difficulties in solving CRT. These results are not surprising when compared to earlier research findings (Frederick, 2005; Gómez-Veiga et al., 2018) since CRT is known to be challenging, and more incorrect intuitive answers are expected (Gómez-Veiga et al., 2018). The difficulty of these problems lies in the involvement of two distinct cognitive processes in solving CRT: an intuitive process, which is incorrect, and a reflective process, which is correct (Epstein, 1994; Stanovich & West, 2000). The intuitive process is fast, and students may answer it without engaging in complex mathematical operations. In contrast, the reflective process is slower and requires inhibition (Campitelli & Gerrans, 2014), making it more challenging to apply. Moreover, despite the absence of time constraints during the test (see measures description), researchers observed that students tend to complete the test quickly, leaving little time for reflection (cf. Travers et al., 2016).

To establish the validity of the third and fourth hypotheses, which suggest that students' scores on CRT predict their achievements in mathematics and physics, we initially examined the correlation between these variables. Our findings revealed a positive and statistically significant correlation between students' mathematics achievements and scores on CRT test, consistent with previous research (Frosch & Simms, 2015; Gómez-Chacón et al., 2014; Gómez-Veiga et al., 2018). Further analysis showed that students with different CRT levels achieved varying levels of success in mathematics. Specifically, students with a low level of cognitive response had the lowest achievements among the three groups of students. To rigorously test the impact of various factors on students' mathematics achievements, we conducted an ordinal logistic regression analysis. We chose this approach because the variables were not continuous or normally distributed, making linear regression unsuitable. The model we constructed considered CRT scores as predictors of students' mathematics achievements while also accounting for the influence of gender and students' class. The overall model explained approximately 12.0% of the variance, and all three factors had a statistically significant impact on the dependent variable. Thus, we can confirm **H3**.

These findings align with prior research, suggesting that students with higher CRT scores are more likely to excel in mathematics compared to those with lower scores (Frosch & Simms, 2015; Gómez-Chacón et al., 2014; Gómez-Veiga et al., 2018).

Similarly, we observed a positive and statistically significant correlation between students' scores on CRT and their achievements in physics. Additionally, students with lower CRT levels had the lowest achievements in physics, although no statistically significant difference was found in physics achievements between students with middle and high CRT levels. Using ordinal logistic regression, we demonstrated that students with higher CRT levels also performed better in physics. Gender did not have a statistically significant impact on students' physics achievements, whereas students' school levels played a significant role in predicting physics achievement. The model explained approximately 9.0% of the variance. These results are consistent with previous research findings, confirming **H4** (Shtulman & McCallum, 2014).

Based on the outcomes of this study and students' performance on CRT, we recommend that educators incorporate the teaching of inhibiting automatic responses into their instructional strategies. To achieve this, educators could present students with problems featuring counterintuitive answers and provide training to develop their reflective skills. These skills can be applied to problem-solving in mathematics and physics, helping students reduce biased automatic responses that may lead to errors and misconceptions. For example, enhancing students' cognitive reasoning skills could aid in reducing errors related to necessary and sufficient conditions or addressing misconceptions related to gravity (e.g., the erroneous idea that gravity depends on magnetism; Williamson & Willoughby, 2012).

In summary, the present study concludes that cognitive reflection skills significantly influence students' achievements in mathematics and physics. Our empirical research aimed to measure the impact of cognitive reflection skills on these achievements, and consistent with existing literature, we found that students with higher levels of cognitive reflection also excel in mathematics and physics. Therefore, CRT may serve as a valuable diagnostic tool to assist educators in planning future lessons and fostering critical thinking skills in students with lower CRT scores (Shtulman & McCallum, 2014; cf. Szaszi et al., 2017; Toplak et al., 2011). Students with lower CRT scores could benefit from specific training programs aimed at enhancing their cognitive and reflective skills (cf. Bull & Lee, 2014). Additionally, CRT offers promise as a diagnostic tool for understanding students' challenges in physics (cf. Wood et al., 2016).

## Limitations & Future Directions

The present research is not without limitations. As previously mentioned, one limitation is the relatively small sample size, which might have precluded the detection of certain effects that could become evident in a larger study. Another limitation pertains to the reliability of the used instrument. While international research has reported Cronbach's alpha coefficients for CRT test ranging from .60 to .74 (Campitelli & Gerrans, 2014; cf. Blacksmith et al., 2019; Primi et al., 2018; Stieger & Reips, 2016), the Cronbach's alpha coefficient in our study is lower. This suggests that the instrument employed may not be the most suitable, warranting further research to develop a more reliable questionnaire. Furthermore, our research focused exclusively on assessing the impact of CRT, gender, and school levels on students' mathematics and physics achievements. We did not consider other variables that could potentially have a more significant influence on predicting students' mathematics and physics performance, such as mathematical anxiety (e.g., Ma & Xu, 2004), self-efficacy (e.g., Skaalvik et al., 2015), and working memory (e.g., De Smedt et al., 2009). Future studies should consider exploring these variables to gain a more comprehensive understanding of their impact. Additionally, the pseudo-$R^2$ coefficients found in our study (.119 and .092) are relatively low (e.g., Zhou et al., 2018). Therefore, caution should be exercised when attempting to generalize the obtained results.

Despite these presented limitations, our study highlights the influence of students' cognitive reflection levels on their achievements in both mathematics and physics. Notably, students with lower levels of cognitive reflection also tend to have lower attainments in these subjects. This could be attributed to the fact that many mathematical and physical problems necessitate the inhibition of heuristics and intuitive reasoning (cf. Tversky & Kahneman, 1974) in favor of a more reflective approach. Educators should consider strategies to help students reduce their reliance on intuitive and rapid problem-solving and encourage them to adopt a more reflective mindset. This could potentially be achieved through various forms of group and teacher-assisted self-regulated learning (Sliško, 2017).

**Author contributions: DD:** collected data, analyzed data, wrote manuscript & **JS:** conceived & supervised work. Both authors approved the final version of the article.

**Funding:** The authors received no financial support for the research and/or authorship of this article.

**Ethics declaration:** The authors declared that all participants (& their parents, in case they were minors) gave their informed consent. All participants took part on a voluntary basis & were not financially remunerated for their participation in the research. The study was carried out following the ethical standards of the 1964 Declaration of Helsinki (L. 18.02.1989, n. 56), Italian law for data privacy (D. Lgs. 196/2003), & European data protection law (European General Data Protection Regulation–GDPR UE 2016/67).

**Declaration of interest:** The authors declared no competing interest.

**Data availability:** Data generated or analyzed during this study are available from the authors on request.

## REFERENCES

Akpotor, J., & Egbule, E. (2020). Gender difference in the scholastic achievement test (SAT) among school adolescents. *World Journal of Education, 10*(1), 97-101. https://doi.org/10.5430/wje.v10n1p97

Bartlett, J. E., & Charles, S. (2021). Power to the people: A beginner's tutorial to power analysis using Jamovi. *PsyArXiv*. https://doi.org/10.31234/osf.io/bh8m9

Blacksmith, N., Yang, Y., Behrend, T. S., & Ruark, G. A. (2019). Assessing the validity of inferences from scores on the cognitive reflection test. *Journal of Behavioral Decision Making, 32*(5), 599-612. https://doi.org/10.1002/bdm.2133

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2019). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics, 82*, 101455. https://doi.org/10.1016/j.socec.2019.101455

Bull, R., & Lee, K. (2014). Executive functioning and mathematics achievement. *Child Development Perspectives, 8*(1), 36-41. https://doi.org/10.1111/cdep.12059

Campitelli, G., & Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition, 42*(3), 434-447. https://doi.org/10.3758/s13421-013-0367-9

De Smedt, B., Janssen, R., Bouwens, K., Verschaffel, L., Boets, B., & Ghesquière, P. (2009). Working memory and individual differences in mathematics achievement: A longitudinal study from first grade to second grade. *Journal of Experimental Child Psychology, 103*(2), 186-201. https://doi.org/10.1016/j.jecp.2009.01.004

Epstein, S. (1994). Integration of the cognitive and the psychodynamic unconscious. *American Psychologist, 49*(8), 709-724. https://doi.org/10.1037/0003-066X.49.8.709

Etcheverry, P. T., Ignjatov, J. S., & de Lourdes Juárez, E. (2020). Influencia de la escolaridad en el desarrollo del razonamiento lógico y la reflexión cognitiva en estudiantes de bachillerato [Influence of schooling on the development of logical reasoning and cognitive reflection in high school students]. *UNIÓN-Revista Iberoamericana de Educación Matemática, 16*(60), 212-232.

Fang, S. C., Hsu, Y. S., & Lin, S. S. (2019). Conceptualizing socio-scientific decision making from a review of research in science education. *International Journal of Science and Mathematics Education, 17*(3), 427-448. https://doi.org/10.1007/s10763-018-9890-2

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using G*power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods, 41*, 1149-1160. https://doi.org/10.3758/BRM.41.4.1149

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives, 19*(4), 25-42. https://doi.org/10.1257/089533005775196732

Frosch, C., & Simms, V. (2015). Understanding the role of reasoning ability in mathematical achievement. In Euroasianpacific joint conference on cognitive science. In *Proceedings of the EuroAsianPacific Joint Conference on Cognitive Science* (pp. 633-638). https://doi.org/10.13140/RG.2.1.1107.2727

Gette, C. R., & Kryjevskaia, M. (2019). Establishing a relationship between student cognitive reflection skills and performance on physics questions that elicit strong intuitive responses. *Physical Review Physics Education Research, 15*(1), 010118. https://doi.org/10.1103/PhysRevPhysEducRes.15.010118

Gilmore, C., Keeble, S., Richardson, S., & Cragg, L. (2015). The role of cognitive inhibition in different components of arithmetic. *ZDM, 47*(5), 771-782. https://doi.org/10.1007/s11858-014-0659-y

Gómez-Chacón, I. M., García-Madruga, J. A., Vila, J. Ó., Elosúa, M. R., & Rodríguez, R. (2014). The dual processes hypothesis in mathematics performance: Beliefs, cognitive reflection, working memory and reasoning. *Learning and Individual Differences, 2*9, 67-73. https://doi.org/10.1016/j.lindif.2013.10.001

Gómez-Veiga, I., Vila Chaves, J. O., Duque, G., & García Madruga, J. A. (2018). A new look to a classic issue: Reasoning and academic achievement at secondary school. *Frontiers in Psychology, 9*, 400. https://doi.org/10.3389/fpsyg.2018.00400

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*(2), 139-155. https://doi.org/10.1037/0033-2909.107.2.139

Kryjevskaia, M., Stetzer, M. R., Lindsey, B. A., McInerny, A., Heron, P. R., & Boudreaux, A. (2020). Designing research-based instructional materials that leverage dual-process theories of reasoning: Insights from testing one specific, theory-driven intervention. *Physical Review Physics Education Research, 16*(2), 020140. https://doi.org/10.1103/PhysRevPhysEducRes.16.020140

Lem, S., Kempen, G., Ceulemans, E., Onghena, P., Verschaffel, L., & Van Dooren, W. (2015). Combining multiple external representations and refutational text: An intervention on learning to interpret box plots. *International Journal of Science and Mathematics Education, 13*(4), 909-926. https://doi.org/10.1007/s10763-014-9604-3

Liu, X., & Koirala, H. (2012). Ordinal regression analysis: Using generalized ordinal logistic regression models to estimate educational data. *Journal of Modern Applied Statistical Methods, 11*(1), 242-254. https://doi.org/10.22237/jmasm/1335846000

Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student gender in mathematics and science. *International Journal of Science and Mathematics Education, 10*(5), 1163-1190. https://doi.org/10.1007/s10763-011-9325-9

Ma, X., & Xu, J. (2004). The causal ordering of mathematics anxiety and mathematics achievement: A longitudinal panel analysis. *Journal of Adolescence, 27*(2), 165-179. https://doi.org/10.1016/j.adolescence.2003.11.003

McFadden, D. (1977). *Quantitative methods for analyzing travel behavior of individuals: Some recent developments*. https://elischolar.library.yale.edu/cgi/viewcontent.cgi?article=1706&context=cowles-discussion-paper-series

Primi, C., Donati, M. A., Chiesi, F., & Morsanyi, K. (2018). Are there gender differences in cognitive reflection? Invariance and differences related to mathematics. *Thinking & Reasoning, 24*(2), 258-279. https://doi.org/10.1080/13546783.2017.1387606

Riegle-Crumb, C., & Moore, C. (2014). The gender gap in high school physics: Considering the context of local communities. *Social Science Quarterly, 95*(1), 253-268. https://doi.org/10.1111/ssqu.12022

Ring, P., Neyse, L., David-Barett, T., & Schmidt, U. (2016). Gender differences in performance predictions: Evidence from the cognitive reflection test. *Frontiers in Psychology, 7*, 1680. https://doi.org/10.3389/fpsyg.2016.01680

Shtulman, A., & McCallum, K. (2014). Cognitive reflection predicts science understanding. In *Proceedings of the Annual Meeting of the Cognitive Science Society*.

Skaalvik, E. M., Federici, R. A., & Klassen, R. M. (2015). Mathematics achievement and self-efficacy: Relations with motivation for mathematics. *International Journal of Educational Research, 72*, 129-136. https://doi.org/10.1016/j.ijer.2015.06.008

Sliško, J. (2017). Self-regulated learning in a general university course: Design of learning tasks, their implementation and measured cognitive effects. *Journal of European Education, 7*(2), 12-24.

Speirs, J. C., Stetzer, M. R., Lindsey, B. A., & Kryjevskaia, M. (2021). Exploring and supporting student reasoning in physics by leveraging dual-process theories of reasoning and decision making. *Physical Review Physics Education Research, 17*(2), 020137. https://doi.org/10.1103/PhysRevPhysEducRes.17.020137

Stanovich, K. E., & West, R. F. (2000). Individual differences in reasoning: Implications for the rationality debate? *Behavioral and Brain Sciences, 23*(5), 645-665. https://doi.org/10.1017/S0140525X00003435

Stieger, S., & Reips, U. D. (2016). A limitation of the cognitive reflection test: Familiarity. *PeerJ, 4*, e2395. https://doi.org/10.7717/peerj.2395

Szaszi, B., Szollosi, A., Palfi, B., & Aczel, B. (2017). The cognitive reflection test revisited: Exploring the ways individuals solve the test. *Thinking & Reasoning, 23*(3), 207-234. https://doi.org/10.1080/13546783.2017.1292954

Toplak, M. E., West, R. F., & Stanovich, K. E. (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition, 39*(7), 1275-1289. https://doi.org/10.3758/s13421-011-0104-1

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Assessing miserly information processing: An expansion of the cognitive reflection test. *Thinking & Reasoning, 20*(2), 147-168. https://doi.org/10.1080/13546783.2013.844729

Travers, E., Rolison, J. J., & Feeney, A. (2016). The time course of conflict on the cognitive reflection test. *Cognition, 150*, 109-118. https://doi.org/10.1016/j.cognition.2016.01.015

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science, 185*(4157), 1124-1131. https://doi.org/10.1126/science.185.4157.1124

Warner, P. (2008). Ordinal logistic regression. *Journal of Family Planning and Reproductive Health Care, 34*(3), 169-170. https://doi.org/10.1783/147118908784734945

Williamson, K. E., & Willoughby, S. D. (2012). Student understanding of gravity in introductory astronomy. *Astronomy Education Review, 11*(1), 10105. https://doi.org/10.3847/AER2011025

Wood, A. K., Galloway, R. K., & Hardy, J. (2016). Can dual processing theory explain physics students' performance on the force concept inventory? *Physical Review Physics Education Research, 12*(2), 023101. https://doi.org/10.1103/PhysRevPhysEducRes.12.023101

Zhang, D. C., Highhouse, S., & Rada, T. B. (2016). Explaining sex differences on the cognitive reflection test. *Personality and Individual Differences, 101*, 425-427. https://doi.org/10.1016/j.paid.2016.06.034

Zhou, C., Kuttal, S. K., & Ahmed, I. (2018). What makes a good developer? An empirical study of developers' technical and social competencies. In *Proceedings of the 2018 IEEE Symposium on Visual Languages and Human-Centric Computing* (pp. 319-321). IEEE. https://doi.org/10.1109/VLHCC.2018.8506577

## APPENDIX A

1. A pen and a notebook cost €1.10 in total. The notebook costs one Euro more than the pen. How much does the pen cost? _____ cents [correct answer: five cents & intuitive answer: 10 cents]
2. If it takes five machines five minutes to make five widgets, how long would it take 100 machines to make 100 widgets? _____ minutes [correct answer: five minutes & intuitive answer: 100 minutes]
3. In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? _____ days [correct answer: 47 days & intuitive answer: 24 days]
4. If Luigi can drink one barrel of water in six days, and Maria can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together? _____ days [correct answer: four days & intuitive answer: nine]
5. Sandro received both the 15$^{th}$ highest and the 15$^{th}$ lowest mark in the class. How many students are in the class? _____ students [correct answer: 29 students & intuitive answer: 30]
6. A man buys a pig for €60, sells it for €70, buys it back for €80, and sells it finally for €90. How much has he made? _____ € [correct answer: €20 & intuitive answer: €10]
7. Simone decided to invest €8,000 in the stock market one day early in 2008. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simone, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simone has:
   a. broken even in the stock market,
   b. is ahead of where he began, or
   c. has lost money [correct answer: c because the value at this point is €7,000 & intuitive response: b]