

DOI: [10.20472/TE.2019.7.1.007](https://doi.org/10.20472/TE.2019.7.1.007)

CORRELATION INVESTIGATION: THE COGNITIVE REFLECTION TEST AND THE MATH NATIONAL EVALUATION SCORES

ANCA TAMAS

Abstract:

The aim of the paper is to assess the relationship between the Cognitive Reflection Test (CRT) scores and the Math National Evaluation scores. The CRT measures the general tendency of most students to choose the easy, but wrong answer instead thinking in order to discover the correct answer. To the author's knowledge, this is the first analysis of the relationship between CRT scores and Math scores at secondary school level. Quantitative methods were used: statistical analyze and econometric methods, using SPSS, the Pearson correlation. A critical assessment of the literature review was made too. Two versions of CRT were used: the original CRT and the long CRT (LCRT), including 10 items from other CRT alternatives. It was proved that there is a significant, strong and positive correlation between the Math National Evaluation scores and the Cognitive Reflection Test scores, in both versions, the original CRT and the 10 items Long Cognitive Reflection Test. More, it was proved that both CRT and LCRT are predictors not only for students' performance on national evaluation, but on another types of external evaluation, like simulation, as well as on internal evaluations too, like semestrial evaluation. Also, the results proved that the students who previously made this type of problems obtained better results than the ones who were never exposed to these tests. As a recommendation, optional courses focused on logical reasoning could have a positive impact on Math National Evaluation scores. The main limitation of the paper is that the participants were only from the Eastern part of Romania.

Keywords:

Cognitive Reflection Test, Long Cognitive Reflection Test, Math National Evaluation, internal evaluation, external evaluation, students' performance

JEL Classification: C02, I21

Authors:

ANCA TAMAS, Center of International Business and Economics; The Bucharest University of Economic Studies, Romania, Email: anca.tamas@rei.ase.ro

Citation:

ANCA TAMAS (2019). Correlation Investigation: The Cognitive Reflection Test and the Math National Evaluation scores . International Journal of Teaching and Education, Vol. VII(1), pp. 92-103., [10.20472/TE.2019.7.1.007](https://doi.org/10.20472/TE.2019.7.1.007)

Introducing CRT

The goal of the paper is to find out if there is a correlation between the Cognitive Reflection Test (CRT) scores and the Math National Evaluation scores.

The Cognitive Reflection Test, widely known as CRT, was introduced by Frederick in 2005, aiming to measure the general tendency to choose an easy, but wrong answer and not to think deeply to come to the right answer. In his seemingly paper, Frederick proved that the ones scoring high on CRT were more patient and more willing to gamble comparing to those scoring low at CRT. CRT mainly measures the cognitive reflection, the ability of a person not to jump to an obvious intuitive conclusion and is highly correlated with regular national tests such are SAT or ACT, thus could predict performance.¹ CRT is related with System 1- referring to intuitive reflex thinking system and System 2-referring to reflective in deep thinking as found in Kahneman (2011)². The Cognitive Reflective test consists of three questions:

1) A bat and a ball cost £1.10 in total. The bat costs £1 more than the ball. How much does the ball cost?

A. 5 B. 10

A is the correct answer and B is the common misleading incorrect answer.

2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?

A. 100 B. 5

A is the common misleading incorrect answer and B is the correct answer.

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?

A. 24 B. 47

A is the common misleading incorrect answer and B is the correct answer.

The original CRT included no answer choices, but, in this paper, this form of CRT was used.

¹ Frederick, S., (2005, Fall). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), pp. 25-42, DOI: 10.1257/089533005775196732.

² Kahneman, D., (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.

In the original study, the students in 11 universities took the test with an overall score mean of 1.24, MIR scoring best with 2,18 out of 3, 4 universities had more than 50% of the students missing all the correct answers, none of the universities exceeding 50% of students with all the correct answers. Also, the original study found CRT is gender sensitive, the number of male participants with CRT high scores was significant higher comparing to the number of female participants. These results are consistent with those of Zanolla (2018), who concluded that males are more likely than females to achieve a high mark at math tests and they are perceived as more active and participative, while females are seen as more conform to school rules, but less involved and more insecure.³

Considering CRT as a predictor of performance on heuristics and biases tasks is controversially, on one hand, Cokely and Kelley (2009) found a correlation between a good performance on CRT and the proportion of choices consistent with expected value, on the other hand, Campitelli and Labollita (2010) found this correlation very low. Inconsistent results were found in the correlation between CRT performance and the tendency to commit the conjunction fallacy, while Oechssler, Roeder and Schmitz (2009) found a good relation, Obrecht, Chapman and Gelman (2009) found none.

Different approaches for scoring CRT can be found in Erceg and Bubić (2017):

- CRT – Regular - summing all the correct answers;
- CRT – Intuitive - summing all the intuitive answers;
- CRT - Proportion Intuitive - proportion of intuitive answers in all wrong answers disregarding the correct answers;
- CRT - Reflection - the sum of non-intuitive answers;
- CRT – Calculation - proportions of correct answers in all non-intuitive answers. ⁴

In this study, the CRT-regular was used.

Over the years, CRT was over exposed and widely known and other versions were developed. Toplak, West and Stanovich (2014) came with a 4 items version and a 7 items version, the latter including the original three items from CRT⁵, another version of CRT 4 items version was

³ Zanolla, G. (2018). The gender gap in math. Evidences of a study in the primary school in the Swiss canton of Ticino. *International Journal of Teaching and Education*, VI(1), pp. 103-125, DOI: 10.20472/TE.2018.6.1.007.

⁴ Erceg, N. and Bubić, A. (2017, January 13). One test, five scoring procedures: different ways of approaching the cognitive reflection test. *Journal of Cognitive Psychology* [online], pp. 381-392. Available at: <https://www.tandfonline.com/doi/full/10.1080/20445911.2016.1278004?scroll=top&needAccess=true> [Accessed 18 March 2018].

⁵ Toplak, M. E., West, R. F. and Stanovich, K. E. (2013, October 28). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning* [online], pp. 147-168. Available at: <https://www.tandfonline.com/doi/abs/10.1080/13546783.2013.844729> [Accessed 17 January]

developed by Thomson & Oppenheimer (2016)⁶ and a 6 items one can be found in Primi et al. (2015)⁷. A version of CRT with 10 items, based on these studies and called LCRT (long cognitive reflective test), was used in this study.

Thomson and Oppenheimer (2016) found out that the participants previously exposed on CRT or at least on one of the three problems outperformed those with no prior exposure, but Chandler, Mueller and Paolacci (2014) proved that changing one or all the problems with the similar ones there was no correlation between CRT performance and prior CRT exposure.⁸ Haigh (2016) concluded that, for strongly intuitive participants or for those with high scores on CRT originally, repeated exposure makes no difference.⁹

None of the students involved in this study knew the CRT test, though the students in two schools, School no.1 and School no. 9, had optional classes in the previous years focused on logical reasoning. As Elsdon et. al (2018) said, there is a lack of interest last years in classical mathematical modules, so the teachers from these schools introduced these logical optional courses trying to stimulate the students' interest for math.¹⁰

The method and the sample

In Romania, after the 8th. grade, the students are taking a national evaluation at Romanian language and Mathematics. I wanted to investigate the correlation between the CRT scores and the Math scores at the above mentioned evaluation. I used the original CRT and another version, called LCRT-long CRT, including 10 items from other CRT versions. I also wanted to find out which one was the best predictor for the Math scores in national evaluation-MNE: the original CRT-OCRT, LCRT, the simulation scores-S, the Math mean scores in the 8th. grade-MA and the mean of the scores in the semestrial evaluations in the 8th. grade-SEA.

269 students aged 14 or 15 from 5 secondary schools in my home town, situated in the Eastern part of Romania, 146 girls (54,27%) and 123 boys (45,73%) were involved in this study. At the end of the first semester, they answered at CRT and at LCRT, choosing one of the two possible

⁶ Thomson, K. S. and Oppenheimer, D. M. (2016, January). Investigating an alternate form of the Cognitive Reflection Test. *Judgment and Decision Making*, 11(1), pp. 99-113.

⁷ Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A. and Hamilton, J. (2015, June 14). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT), *Journal of Behavioral Decision Making* [online] Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.1883> [Accessed 12 January 2018].

⁸ Haigh, M., (2016, September 30). Has the standard cognitive reflection test become a victim of its own success?. *Advances in Cognitive Psychology*, [online] Volume 12(3), pp. 145–149. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5225989> [Accessed 12 February 2018].

⁹ Thomson, K. S. and Oppenheimer, D. M. (2016, January). Investigating an alternate form of the Cognitive Reflection Test. *Judgment and Decision Making*, 11(1), pp. 99-113.

¹⁰ Elsdon, M., Busawon, K., Vigouroux, A., Binns, R., Forbes, I. and Railland, V. (2018, June). A Comparative Assessment of Delivery Methods of Mathematics and Technology Enhanced Learning. In: *Proceedings of the 5th Teaching & Education Conference*. [online] Amsterdam: ISES, pp. 27-35. Available at: <http://iises.net/proceedings/5th-teaching-education-conference-amsterdm/table-of-content> [Accessed 16 February 2019], DOI:10.20472/TEC.2018.005.003.

answers (LCRT can be found in the appendix). The other scores were collected at the end of the school year and after the national evaluation.

SPSS was used for standard statistics, correlation tests were run and linear regression using MNE as dependent variable and OCRT, LCRT, S, MA, SEA as independent variables. Further investigations were made to find out if there are gender differences to compare schools results, to find out how did the ones in CRT high scores group or in the CRT low scores group in the national evaluation.

The results

In table 1, the average scores were computing for OCRT, LCRT, MNE, S, MA, SEA, considering overall scores and then the averages for each considered schools and for the two categories: boys and girls. For OCRT and LCRT, the CRT-regular (Erceg and Bubić, 2017) was used.

Table 1: The average scores on OCRT, LCRT, MNE, S, MA, SEA

School number	OCRT	LCRT	MNE	S	MA	SEA
1	1,46	7,25	6,46	4,97	7,37	7,01
4	0,63	6,04	5,60	4,44	6,45	6,11
8	1,06	6,34	6,97	5,74	7,11	6,46
9	1,55	6,50	6,88	5,14	6,97	6,85
11	1,15	5,74	6,33	4,81	7,17	6,87
Total	1,14	6,17	6,53	5,12	7,10	6,69
Girls	1,03	5,84	6,56	5,08	7,21	6,79
Boys	1,27	6,58	6,51	5,18	6,97	6,56

Source: Author's table based on own calculations

The means for OCRT are comparable with the ones in the original study in the interval (0.57; 2.18). The average scores for OCRT and LCRT were higher for the two schools where the students had optional courses focused on logical reasoning, which is consistent with Thomson and Oppenheimer (2016), but in contradiction with the ones of Chandler, Mueller and Paolacci (2014), considering LCRT items are different from OCRT items. The OCRT and LCRT averages for boys are higher than the averages for girls, the results confirm the Frederick ones. As for the

averages for MNE and S, School 8 leads, followed by the two schools with higher scores on OCRT and LCRT.

Table 2: The correlation between OCRT, LCRT, MNE, S, MA, SEA

	MNE	OCRT	LCRT	S	MA	SEA
MNE	1	,568**	,545**	,878**	,826**	,805**
OCRT		1	,587**	,580**	,602**	,602**
LCRT			1	,520**	,512**	,519**
S				1	,809**	,786**
MA					1	,932**
SEA						1

** . Correlation is significant at the 0.01 level (2-tailed).

Source: Author's table based on SPSS output

The correlations are significant, positive and strong among all pairs of variables, which might be explained by the fact that both OCRT and LCRT are based on numeracy, therefore those doing well on these tests are doing well on various Math evaluations.

Table 3: OCRT scores compared to NE, S, MA, SEA

Means at	OCRT=0	OCRT=1	OCRT=2	OCRT=3
National evaluation	5,32	6,30	7,67	8,43
Simulation	3,90	4,79	6,22	7,38
Mathematics	6,06	6,72	8,29	8,96
Semestrial evaluations	5,48	6,31	7,93	8,85

Source: Author's table based on own calculations

Table 4: LCRT scores compared to NE, S, MA, SEA

	Means at			
LCRT scores	National evaluation	Simulation	Mathematics	Semestrial evaluations
10	8,33	7,34	9,13	8,97
9	7,77	6,30	8,55	8,35

8	7,36	5,93	7,44	7,13
7	7,09	5,52	7,52	7,09
6	6,54	5,19	7,06	6,64
5	5,87	4,44	6,51	6,05
4	5,56	4,20	6,20	5,89
3	5,44	3,78	5,63	4,81
2	3,64	2,70	5,20	4,90
1	3,77	2,30	5,41	4,32

Source: Author's table based on own calculations

Table 3 and table 4 prove that both OCRT and LCRT are predictors for students' performance on external evaluations, like national evaluation or simulation, as well as on internal evaluations, like semestrial evaluation or Math scores. The higher the scores on OCRT and LCRT are, the higher are the scores on Math internal and external evaluations.

Further on, the correlations are made separate for schools previous having optional courses focused on logical reasoning, called L schools and the rest of the schools, called NL schools.

Table 5: Correlations of OCRT and MNE, S, MA, SEA

	MNE	S	MA	SEA
L schools	,636**	,628**	,599**	,654**
NL schools	,562**	,582**	,612**	,591**

** . Correlation is significant at the 0.01 level (2-tailed).

Source: Author's table based on SPSS output

For the students used with the kind of problems presented in Frederick original test, OCRT is a better predictor for external evaluations comparing to the students who are not used with this kind of problems. The results are consistent with ones of Thomson and Oppenheimer (2016).

For the next investigation, the following notations were considered: L=low scores, 0 in OCRT, less than 5 in LCRT, MNE, S, MA, SEA; IL=intermediate low, 1 in OCRT, greater or equal to 5, less than 7 in LCRT, MNE, S, MA, SEA; IH= intermediate high, 2 in OCRT, greater or equal to 7, less than 9 in LCRT, MNE, S, MA, SEA; H=high, 3 in OCRT, greater or equal to 9 in LCRT, MNE, S, MA, SEA.

Table 6: Percents of students with lower and higher scores in OCRT and LCRT

		L %	IL %	IH %	H %
OCRT	Girls	36,05	35,37	16,32	11,56
	Boys	23,57	39,83	20,32	15,44
LCRT	girls	20,40	40,81	30,62	7,48
	boys	13,82	32,52	33,33	19,51

Source: Author's table based on own calculations

The boys outperformed the girls in three out of four categories, intermediate lower scores (IL) being the exception, therefore Frederick (2005) findings are partially sustained.

For the final question, regression in SPSS was performed, considering MNE as the dependent variable and OCT, LCRT, S, MA, SEA as independent variables.

Testing normality using the Kolmogorov-Smirnov test, Sig is 0.091, which is greater than 0.05, therefore the data are fairly normally distributed. Further on, if we run statistics, the Skewness is -0.192, which is greater than -1 and in absolute value (0.192) is less than the double of the standard deviation value (0.594), therefore there is no normality problem.

To test heteroscedasticity, Glejser Test was used in SPSS and, since Sig for all independent variables is greater than 0.05, there is no heteroscedasticity problem.

The SPSS output for linear regression is shown below. It shows that there is a very strong correlation between MNE and the independent variables S, MA and LCRT as seen in model 3, the correlation coefficient $R=0,902$ and the adjusted squared R value is high, therefore 82% of the MNE variation is explained by the variables S, MA, LCRT.

Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	2,303	,152		15,123	,000		
	S	,826	,028	,878	29,899	,000	1,000	1,000
2	(Constant)	,964	,229		4,201	,000		
	S	,571	,043	,607	13,298	,000	,345	2,896
3	MA	,373	,051	,335	7,340	,000	,345	2,896
	(Constant)	,794	,236		3,364	,001		

S	,547	,043	,582	12,596	,000	,330	3,029
MA	,349	,051	,313	6,825	,000	,334	2,994
LCRT	,075	,029	,082	2,601	,010	,706	1,417

a. Dependent Variable: MNE

The coefficients for S and MA are significant at .01 level, while the coefficient for LCRT is at .05 level. Among all three variables, S is the most important, followed by MA and at a lower level LCRT.

There are no collinearity problems, since the tolerance values are greater than .10 and the VIF values are less 10.

Conclusions

In this study it was analyzed the correlation between two versions of the cognitive reflection tests and the scores at Math National Evaluation. The two versions were OCRT - the original CRT, with options for answers and LCRT - a ten items test based on Toplak (2014) and Thomson & Oppenheimer (2016), extending CRT tests. The Math National Evaluation is an exam for Romanian children aged 14 or 15 who graduated the 8th grade. Another analysis in this study was conducted to find out which is a better predictor for Math National Evaluation scores among the CRT scores, the simulation scores, the Math scores and the Math semestrial scores. Math simulation is organized at national level, usually in February and the Math National Evaluation in June. Semestrial Math Evaluation is an internal evaluation and Math scores is a weighted mean of one's Math scores. The findings are consistent with the results of Frederick and Thomson & Oppenheimer, in respect with the averages for OCRT and the gender sensitivity of OCRT scores. Also the results of Toplack and Haigh were sustained, the students previously exposed to CRT like problems got higher scores on OCRT, as well as on LCRT. The results proved there is a strong positive correlation between OCRT and LCRT scores and Math National Evaluation scores. The linear regression analysis revealed that Simulation scores, Math means and LCRT are better predictors for Math National Evaluation scores. Further research including children from different regions should be done in the following years.

Appendix

LCRT - long cognitive reflection test

1. You are running in a race and you surpass the competitor currently on second place. You are now on
 - A. First place
 - B. Second place
2. A farmer had 15 sheeps and all but 8 died. How many sheeps does the farmer still have?
 - A. 8
 - B. 7

3. Emily' father has three daughters. The name of the first two ones are April and May. The name of the third daughter is
- A. June B. Emily
4. How many cubic meters are in a hole 3 meters long, 3 meters width, 3 meters height?
- A. 0 B. 27
5. If 3 elves can pack 3 toys in an hour, how many elves are needed to pack 6 toys in 2 hours?
- A. 3 B. 6
6. Jerry received both the 15th highest and 15th lowest mark in the class. How many students are in the Class?
- A. 29 B. 30
7. In a team the probability that the tall athlets to win medals is threee times bigger than the the probability that the short athlets to win medals. If the team got 60 medals by far, how many medals the short athlets have won?
- A. 20 B. 15
8. If John can drink one barrel of water in 6 days, and Mary can drink one barrel of water in 12 days, how long would it take them to drink one barrel of water together?
- A. 9 B. 4
9. A man buys a pig for 60\$, then sells it for 70\$, afterwards buys the same pig for 80\$ and finally sells it for 90\$. How many dolars does the man won after all these transactions?
10. Simon decided to invest \$8,000 in the stock market on a day in early 2017. Six months after he invested, on July 17, the stocks he had purchased were down 50%. Fortunately for Simon, from July 17 to October 17, the stocks he had purchased went up 75%. At this point, Simon has:
- A. Won money B. Lost money

References

- Campitelli, G., and Labollita, M. (2010, June). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5, pp. 182–191.
- Chandler, J., Mueller, P., and Paolacci, G. (2014, March). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior Research Methods*, 46, pp. 112–130, doi: [10.3758/s13428-013-0365-7](https://doi.org/10.3758/s13428-013-0365-7).

- Cokely, E. T., and Kelley, C. M. (2009, February). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgment and Decision Making*, 4(1), pp. 20–33.
- Elsdon, M., Busawon, K., Vigouroux, A., Binns, R., Forbes, I. and Railland, V. (2018, June). A Comparative Assessment of Delivery Methods of Mathematics and Technology Enhanced Learning. In: *Proceedings of the 5th Teaching & Education Conference*. [online] Amsterdam: IISES, pp. 27-35. Available at: <http://iises.net/proceedings/5th-teaching-education-conference-amsterdm/table-of-content> [Accessed 16 February 2019], DOI:10.20472/TEC.2018.005.003.
- Erceg, N. and Bubić, A. (2017, January 13). One test, five scoring procedures: different ways of approaching the cognitive reflection test. *Journal of Cognitive Psychology* [online], pp. 381-392. Available at: <https://www.tandfonline.com/doi/full/10.1080/20445911.2016.1278004?scroll=top&needAccess=true> [Accessed 18 March 2018]. <https://doi.org/10.1080/20445911.2016.1278004>
- Frederick, S., (2005, Fall). Cognitive Reflection and Decision Making. *Journal of Economic Perspectives*, 19(4), pp. 25-42, DOI: 10.1257/089533005775196732.
- Haigh, M., (2016, September 30). Has the standard cognitive reflection test become a victim of its own success?. *Advances in Cognitive Psychology*, [online] Volume 12(3), pp. 145–149. Available at: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5225989> [Accessed 12 February 2018].
- Kahneman, D., (2011). *Thinking, fast and slow*. New York: Farrar, Straus and Giroux.
- Obrecht, N. A., Chapman, G. B., and Gelman, R. (2009, July). An encounter frequency account of how experience affects likelihood estimation. *Memory & Cognition*, 37(5), pp. 632–643, doi: 10.3758 / MC.37.5.632.
- Oechssler, J., Roider, A., and Schmitz, P. W. (2009, May 13). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization*, 72, pp. 147–152, doi:10.1016/j.jebo.2009.04.018.
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A. and Hamilton, J. (2015, June 14). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT), *Journal of Behavioral Decision Making* [online] Available at: <https://onlinelibrary.wiley.com/doi/abs/10.1002/bdm.1883> [Accessed 12 January 2018].
- Thomson, K. S. and Oppenheimer, D. M. (2016, January). Investigating an alternate form of the Cognitive Reflection Test. *Judgment and Decision Making*, 11(1), pp. 99-113.
- Toplak, M. E., West, R. F. and Stanovich, K. E. (2013, October 28). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. *Thinking & Reasoning* [online], pp. 147-168. Available at: <https://www.tandfonline.com/doi/abs/10.1080/13546783.2013.844729> [Accessed 17 January].
- Zanolla, G. (2018). The gender gap in math. Evidences of a study in the primary school in the Swiss canton of Ticino. *International Journal of Teaching and Education*, VI(1), pp. 103-125, DOI: 10.20472/TE.2018.6.1.007.