

DOI: [10.20472/TE.2020.8.1.006](https://doi.org/10.20472/TE.2020.8.1.006)

COGNITIVE REFLECTION TEST - NEXT GENERATION

ANCA TAMAS

Abstract:

Purpose - The aim of this paper is to investigate the correlation between the short Cognitive Reflection Test and the long Cognitive Reflection Test (CRT), the gender effect and the educational level effect on short CRT and long CRT, as well as the Granger causality between the two ones.

Design / Methodology / Approach - A statistical analyze was provided and econometric measures and tests were used, in EViews and SPSS, like the Pearson Correlation Coefficient and the Granger Causality. A critical assessment of literature review was made. For collecting the data, quantitative methods were used (questionnaires).

Findings - There is a significant, strong and positive correlation between the short Cognitive Reflection Test version (short CRT) and the long version (long CRT), long CRT Granger causes short CRT, girls outperformed boys on each educational level.

Practical implications - Long CRT could be an alternative for short CRT, which is now overexposed and widely known.

Originality / Value - Once thinking system 2 is triggered by completing another questionnaire, the results on CRT in either version increased, even if the nature of the other questionnaire is completely different.

Limitations - The other questionnaire was administrated only for high school students. Future research should be conducted in order to trig the thinking system 2 before administrating the test for secondary school and university students too.

Keywords:

short Cognitive Reflection Test, long Cognitive Reflection Test, thinking system 2

JEL Classification: I21, C02

Authors:

ANCA TAMAS, The Bucharest University of Economic Studies, Romania, Email: anca.tamas@rei.ase.ro

Citation:

ANCA TAMAS (2020). Cognitive Reflection Test - Next Generation. International Journal of Teaching and Education, Vol. VIII(1), pp. 84-96., [10.20472/TE.2020.8.1.006](https://doi.org/10.20472/TE.2020.8.1.006)

Introduction

Cognitive Reflection Test was introduced by Frederick in 2005 and consists of the following three questions:

- 1) 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?
- 2) If it takes 5 machines 5 minutes to make 5 widgets, how long would it take 100 machines to make 100 widgets?
- 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?

As Frederick (2005) stated, the cognitive reflection is the ability of reflecting on or questioning the intuitive answer or the answer which pops up in mind first. Even for respondents that gave the correct answer, the intuitive incorrect answer was considered first and rejected in favor of the reflective answer and never the other way round.¹ It develops the students' mind in a pleasant way, like other different mathematical topics (games, puzzles, fraction posters etc.) (Van Der Walt and Potgieter, 2019). The correlation between CRT and general intelligence or reasoning performance was positive and had a strength between 0.2 and 0.5 (Bialek and Domurat, 2017). CRT is considered the best predictor of rational thinking, based on the 0.49 correlation between them (Toplak et al., 2011). CRT is not another numeracy test, because it predicts different biases (Liberali et al., 2012). CRT measures more than the mathematical thinking, it measures the rational thinking and the open-minded thinking.²

Literature review

Over the time, CRT proved to be a valuable instrument to determine whether a respondent is an intuitive thinker or a reflective thinker. Reflective thinkers are aware of the intuitive answer and of the reflective answer and are able to self assess their performance on CRT test. The intuitive thinkers are aware only of the intuitive answer and unable to assess their poor performance.³ Reflective thinking influences the performance of daily thinking (Pennycook et al., 2015) and is more relevant than numeracy skills (Cokely and Kelly, 2009). The rational thinking in everyday life means how to behave in order to get most of you want with your own limited resources (Stanovich et al., 2016).

¹ Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), pp. 25–42, DOI: 10.1257/089533005775196732.

² Campitelli, G. and Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, 42, pp. 434–447, <https://doi.org/10.3758/s13421-013-0367-9>.

³ Mata, A., Ferreira, M. B. and Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, 105, pp. 353–373, <https://doi.org/10.1037/a0033640>.

The CRT reflective score (the sum of all correct answers) and the CRT intuitive score (the sum of all intuitive answers) are highly negative correlated. Still CRT intuitive score cannot accurately distinguish between the intuitive answers and the mathematical errors.

Among the three problems of the original CRT, the bat and the ball problem is solved by one in five respondents in most empirical studies. The main reason is that the respondents are not willing for a cognitive effort and the intuitive thinking is good enough for them.⁴ A similar explanation of the low percentage of respondents who solved the bat and the ball problem in CRT was found by Szolossi et al. (2017). They proved that people substitute a difficult task with an easier one without being aware of the substitution.

There is a substantial research body and there are a lot of empirical studies which prove that people with a good score on CRT would get good scores at numeracy or general ability tests (Campitelli & Labollita, 2010; Cokely & Kelley, 2009; Frederick, 2005; Liberali et al., 2012; Toplak, West & Stanovich, 2011). There is a significant correlation between CRT and academic achievement.⁵

Most of the empirical studies showed that CRT is gender sensitive, males usually perform better than females. Campitelli and Gerrans (2014) explained this difference and proved that for females is more difficult to resist to the intuitive answer compared to males. The usually gender difference is not fully explained so far, though Thomson and Oppenheimer (2016) considered it as a consequence of the numeracy ability, usually higher for males.⁶

The previous exposure at one of the CRT problems increases the overall test score, not only for the original CRT, but also for other CRT versions (Toplak et al., 2011; Thomson and Oppenheimer, 2016; Tamaş, 2018).

Primi et al. (2016) considered that CRT is more suitable for highly educated adults and for the others is a difficult test. On the other hand, Stieger and Reips (2016) considered that CRT is not suitable neither for highly educated adults, because one third of them can actually solve all three problems, nor for lowest educated, because one third of them are not able to solve any of the problems.⁷ The same task could be solved in different ways, depending on the format of the test, according to Mastrogiorgio and Petracca (2014).

⁴ De Neys, W., Rossi, S., and Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, 20, pp. 269–273, <https://doi.org/10.3758/s13423-013-0384-5>.

⁵ Campitelli, G. and Labollita, M. (2010, June). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5(3), pp. 182–191, <https://www.sas.upenn.edu/~baron/journal/10/91230/jdm91230.pdf>.

⁶ Thomson, K. S., and Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11, pp. 99–113, <http://journal.sidm.org/15/151029/jdm151029.pdf>.

⁷ Stieger, S. and Reips, U.-D. (2016, September). A limitation of the cognitive reflection test: Familiarity. *PeerJ*, 4(2):e2395, DOI: 10.7717/peerj.2395.

Aim

The aim of this paper is to investigate the correlation between the short CRT and the long CRT, the gender effect and the educational level effect on short CRT and on long CRT, as well as the Granger causality between the two ones.

Research Hypothesis

The first hypothesis was that once thinking system 2 is triggered, in this case by completing the first questionnaire, the results on CRT will increase, although the nature of the two questionnaires is completely different.

The second hypothesis was that the students at secondary school level will score higher on intuitive problems.

The third hypothesis was that the students at university level will score higher on practical problems.

The sample

600 questionnaires were administrated in Spring term 2019 in one secondary school and two high schools in Bârlad, one of the poorest regions in Romania and in four high schools and a university in Bucharest, the capital and the richest region in Romania. Out of them, 586 were fully completed. The questionnaires were completed on voluntary basis only by respondents who declared they are not familiar with CRT in either form. There was a difference though, the students in high school had two questionnaires to complete, the first one was about their future study place and their future work place (as part of a different study) and the second questionnaire was with the CRT in both versions. The short CRT is the original CRT introduced by Frederick in 2005. A long CRT was considered based on the extended versions of Toplak and Thomson. To assess the short CRT, the CRT score was used.

Table 1: Sample distribution

	Boys	Girls	Total
Secondary school	85	67	152
High school	162	160	322
University	32	80	112
Total	279	307	586

Source: Author's table based on questionnaires

The results

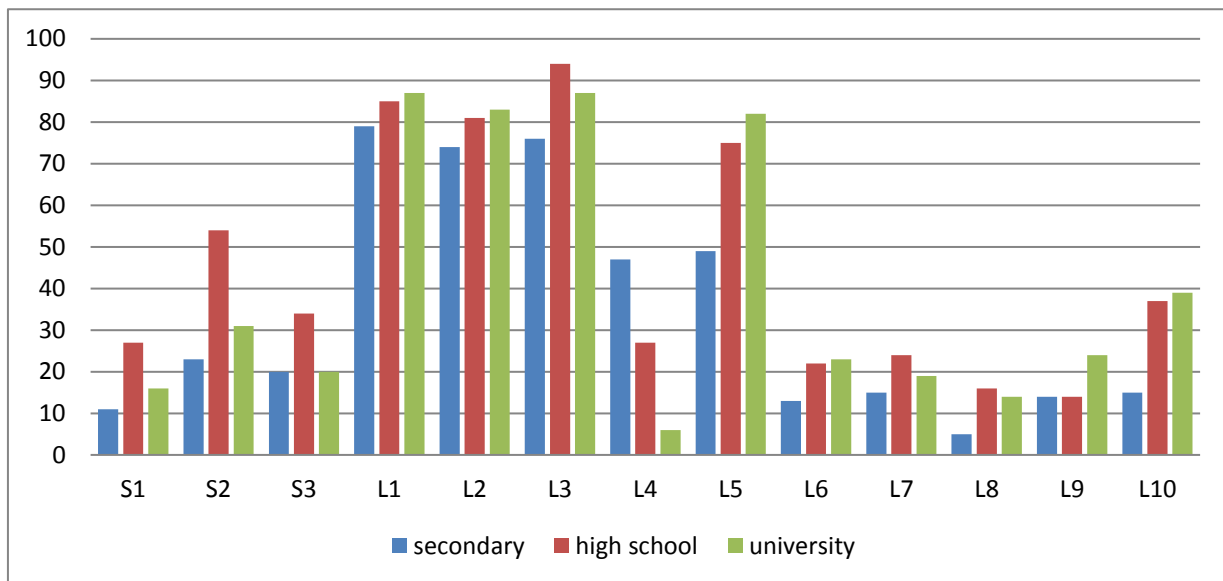
Table 2: The results on short CRT

		Number of correct answers				Percent of correct answers (%) (approximated to the nearest whole number)				CRT score
		Low		High		Low		High		
		0	1	2	3	0	1	2	3	
Secondary school (152)	total	88	46	13	5	58	30	9	3	0.57
	boys	48	28	7	2	56	33	8	3	0.49
	girls	40	18	6	3	60	27	9	4	0.57
High school (319)	total	115	96	58	53	36	30	18	16	1.18
	boys	64	51	31	16	40	31	19	10	1
	girls	51	45	27	37	32	28	17	23	1.32
University (112)	total	64	28	17	3	57	25	15	3	0.63
	boys	17	7	7	1	55	22	22	1	0.8
	girls	47	21	10	2	59	26	13	2	0.58
Overall		267	170	88	61	46	29	15	10	0,91
All boys		129	86	45	19	46	31	16	7	0.82
All girls		138	84	43	42	45	27	14	14	0.9

Source: Author's table based on respondents' answers

Compared with the results in the Frederick's study, the results in this study are in range between 0.57, similar to University of Toledo for secondary school, up to 1.18, similar to University of Michigan for high school. For University, the results are in line with the previous studies, the short CRT score is 0.63 and compared to the lowest University score in the original study, the percents for the extremes are better and the intermediate ones are worse. The boys outperformed the girls, 0.8 compared to 0.58, sustaining the previous studies (Zanolla, 2018).⁸ But for secondary school and for high school, the results are different, in both cases the girls significantly outperformed the boys and this result stands included for all girls and all boys. In light of these results, the short CRT is most suitable for high school.

⁸ 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

Figure 1: The percent of correct answers in short CRT and long CRT

Source: Author's figure based on the respondents' answers

The bat and ball problem from the short CRT involves more Math competences compared to the other two. In order to solve this problem, one either use System 1 of thinking, according to Kahneman and jump to the wrong intuitive answer, either decide to use System 2 of thinking and came up with the correct answer. Most of the previous studies reveal that one in five people come up with the correct answer, this study confirms the results, overall 21% solved this problem.

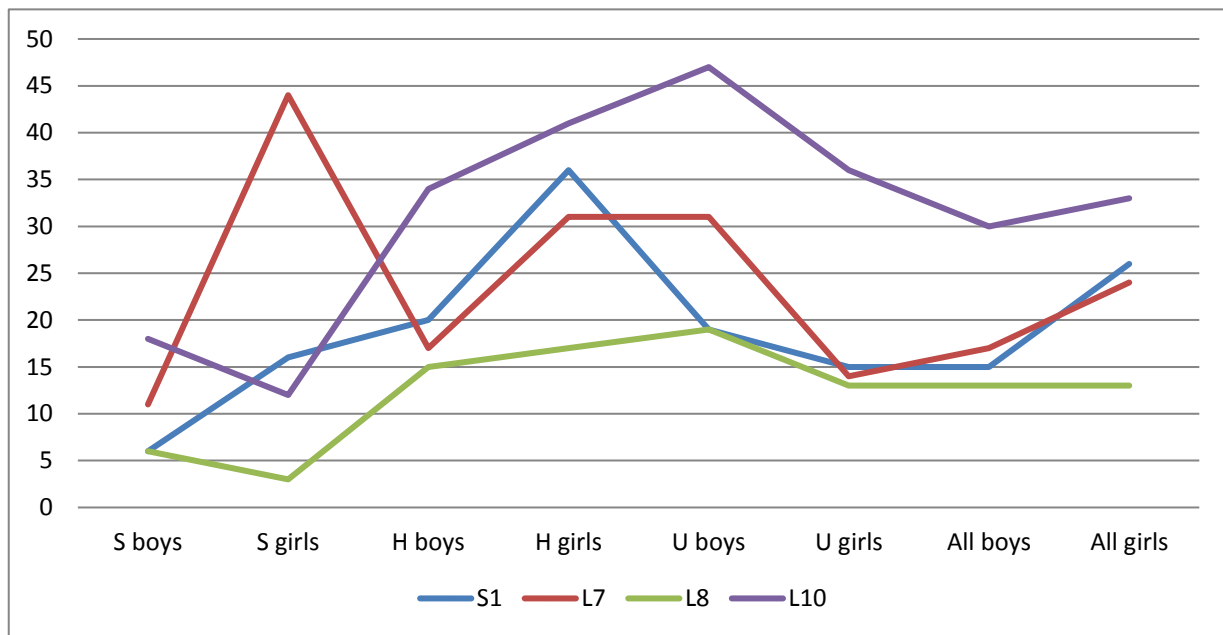
The short CRT Math problem, namely the bat and ball problem, has the following corresponding Math problems in the long CRT: L7, L8, L10.

L7: The probability that tall athletes to win medals is three time larger compared to short athletes. If the team won 60 medals by now, how many medals were won by short athletes?

L8: If John drinks a barrel of water in 6 days and Mark drinks a barrel of water in 12 days, how many days would take to the two to drink a barrel of water together?

L10: One day in 2018, Simon decided to invest 8000\$ in the shared market. Six months later, on 17th of July, his actions fall with 50%. Yet, between 17th of July to 17th of September, his actions value increased with 75%. How much did Simon earned?

To solve S1, one should decide to use an arithmetic method and to have abilities to work with decimals, L7 is a replica of S1 using whole numbers, L8 involves fractions instead of decimals and L10 involves percents. The ability of working with fractions was the lowest and the ability of working with percents was the highest in all categories. The best score for girls was for L7 for secondary level, while the best score for boys was for L10 for university level.

Figure 2 : The percent of correct answers for Math problems in short CRT and long CRT

Source: Author's figure based on the respondents' answers

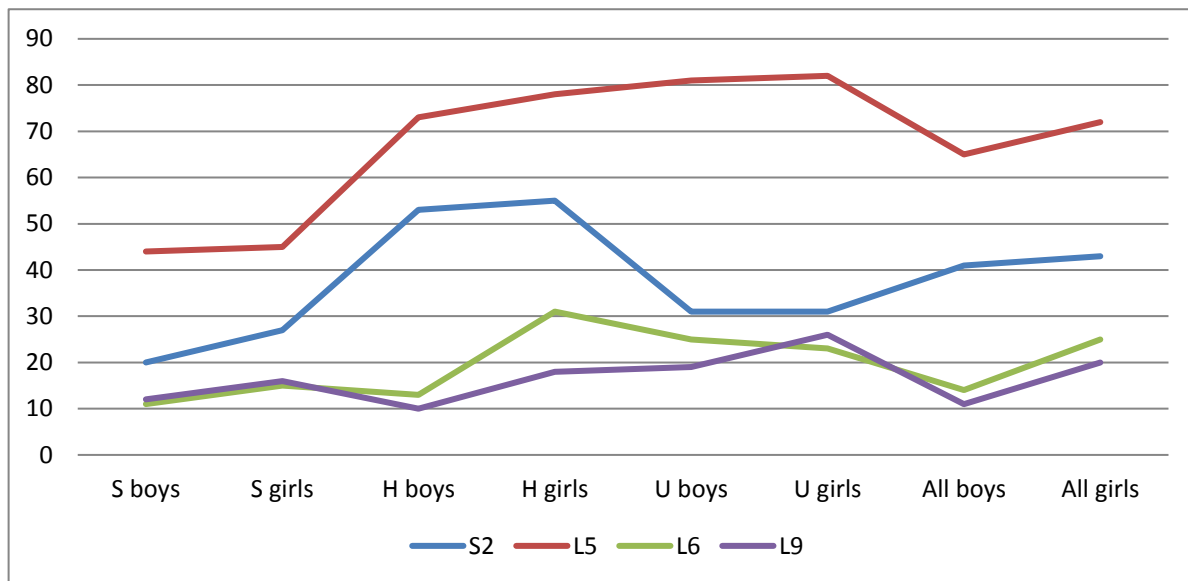
The widgets problem S2 in the short CRT is a practical problem and has 3 corresponding problems in the long CRT: L5, L6, L9.

L5: If three dwarfs can pack three toys in one hour, how many dwarfs are needed to pack six toys in two hours?

L6: Jerry got the 15th higher mark and the 15th lower mark in his class in the same test. How many pupils are in Jerry's class?

L9: A man buys a pig with 60\$ and resells it with 70\$. Afterwards the man buys again the pig with 80 \$ and resells it again with 90\$. How much did the man win?

The results sustained the findings in the previous studies, once the respondents are exposed to a similar problem, the scores are getting higher. L5 is a replica of S2, with dwarfs instead of machines and toys instead of widgets, therefore S2 was like a signal for L5, where everybody scored higher. For the other two practical problems, girls scored better than boys.

Figure 3: The percent of correct answers for practical problems in short CRT and long CRT

Source: Author's figure based on the respondents' answers

Finally, the lily problem S3 of the short CRT is a logical problem and has 4 corresponding problems in the long CRT: L1, L2, L3, L4.

L1. If you are running a race and you pass the person in the second place, what place are you in?

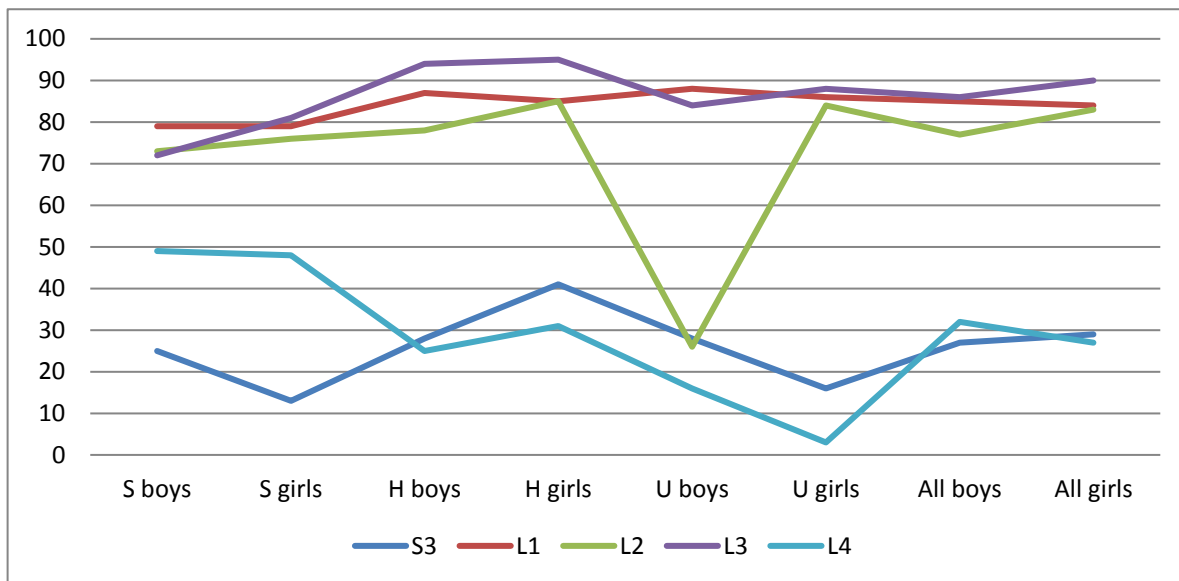
L2. A farmer had 15 sheep and all but 8 died. How many sheep has the farmer now?

L3. Emily's father has three daughters. The name of the first two are April and May. What is the name of the third one?

L4. How many cube meters are in a hole 3 meters long, 3 meters width, 3 meters high?

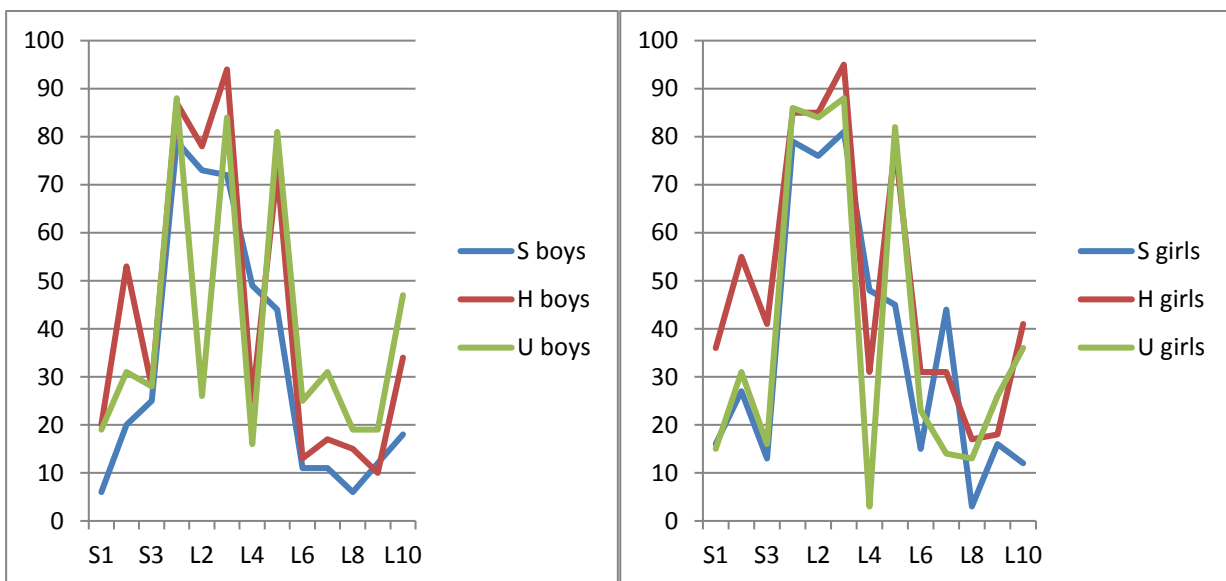
L1, L2, L3 were more intuitive compared with the others, to solve them System 1 according to Kahnemann performed well and, therefore, got higher scores. L4 is the only problem where boys outperformed the girls. Surprisingly, at university level, boys got the lowest scores on L2 problem, yet because it was a difficult problem, but mainly because the boys of any age didn't read the problem carefully, they mainly scan the numbers and then rely on System 1.

Figure 4 : The percent of correct answers for logical problems in short CRT and long CRT

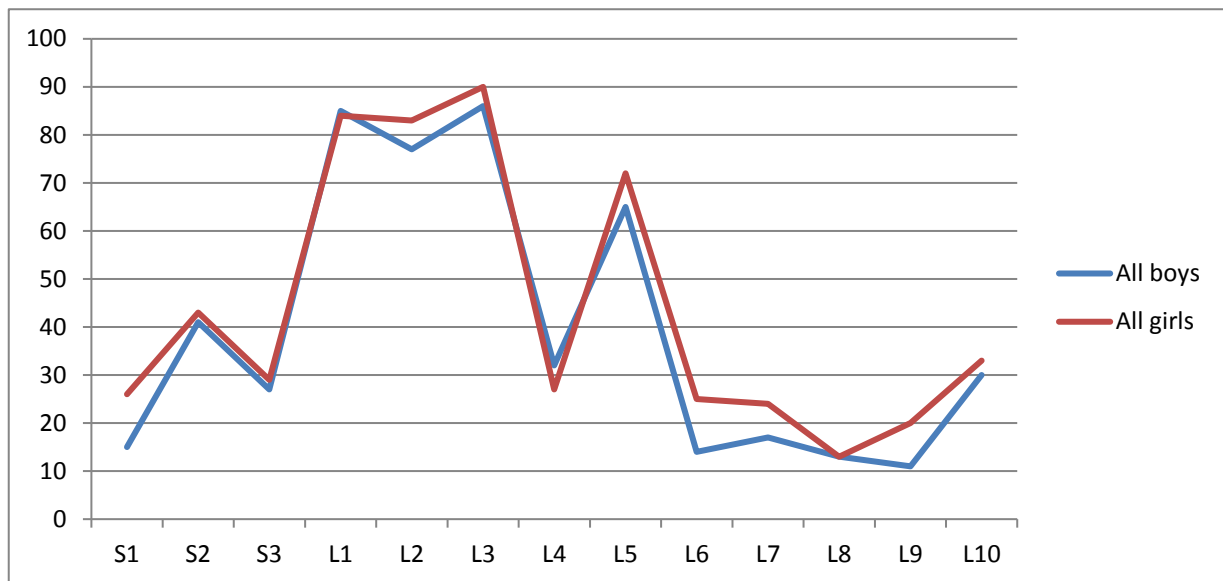


Source: Author's figure based on the respondents' answers

Figure 5: Compared performances of all categories for boys (left) and girls (right) (%)



Source: Author's figure based on the respondents' answers

Figure 6: Compared performances according the gender

Source: Author's figure based on the respondents' answers

The boys from secondary school outperformed all the other boys only on L4 problem and equalled the high school boys in L9 problem and had worse scores on all the other problems. The girls from secondary school outperformed all the other girls only on L4 and L7 and had worse scores on all the other problems. The high school boys outperformed the others in short CRT, as well as on L1, L2, L3 problem, while the high school girls outperformed the others in short CRT and L1, L2, L3, L7 and L10 problem. University boys outperformed the others on L5, L6, L7, L9, L10 problem, while university girls outperformed the others on L5 only. Therefore, high school girls, followed by high school boys and university boys had the highest performances. All in all, girls scored better than boys in all questions, except L4. These results are unlike the results in the previous studies.

Short CRT and long CRT have a positive and powerful correlation, 0.852. EViews was used to find out if there is a Granger causality between short CRT and long CRT and, further on, between the Math problems, the practical problems and the logical problems in the two tests. Overall, long CRT Granger causes short CRT, which is in line with the findings of Toplak et al. (2014), Thomson and Oppenheimer (2016), Tamaş (2018).⁹ Other interesting Granger causalities were found between the short CRT problems and the long CRT ones.

⁹ Toplak, M. E., West, R. F. and Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39, pp. 1275–1289, <https://doi.org/10.3758/s13421-011-0104-1>.

Table 3. Granger causality between CRT problems

Null Hypothesis	Probability
L7 does not Granger cause S1	0.0003
S1 does not Granger cause L7	3.E-06
L8 does not Granger cause S1	0.0201
S1 does not Granger cause L8	0.0065
S1 does not Granger cause L10	0.0071
L5 does not Granger cause S2	3.E-05
L7 does not Granger cause S1	5.E-06
S2 does not Granger cause L6	0.0020
S2 does not Granger cause L9	0.0016
L2 does not Granger cause S3	0.0053
L4 does not Granger cause S3	0.0044

Source: Author's table based on EViews outputs

Conclusions

The CRT score was used both for short CRT and long CRT. The results on all three educational level categories are in range with the original results. For university level, the score is a little bit higher than the lowest results in Frederick (2005)¹⁰, but, surprisingly, the results for secondary school are the same with the ones of the University of Toledo. The low results in Math problems confirm the findings of Szollosi et al. (2017), that people often rely on thinking system 1 without being aware of it and they are not willing for a cognitive effort (De Neys et al., 2013).

As for Granger sensitivity, the results of the previous studies (Campitelli and Gerrans, 2014) (Thomson and Oppenheimer, 2016) are not sustained. Instead, gender difference is the other way round, girls outperformed boys, irrespectively of the educational level. Exposure to short CRT increased long CRT scores, similar with Toplack et al. (2011), Thomson and Oppenheimer (2016), Tamaş (2018).¹¹

The first hypothesis is sustained, the higher results of the high school students could be explained by the fact that the students completed another questionnaire first and, by doing that, the thinking system 2 was used. Once the system 2 was triggered, the results in the second questionnaire were better. The first hypothesis was that once thinking system 2 is triggered, in this case by the first questionnaire, the results of CRT will increase, although the nature of the two questionnaires is completely different. The findings sustained the second hypothesis, but the third one was just partially sustained, only for university boys.

¹⁰ Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), pp. 25–42, DOI: 10.1257/089533005775196732.

¹¹ Thomson, K. S., and Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11, pp. 99–113, <http://journal.sjdm.org/15/151029/jdm151029.pdf>.

The short CRT and the long CRT are highly correlated at 0.852 level and the long CRT Granger causes the short CRT. S1 Granger causes the Math problems in long CRT: L7, L8, L10 and the causality is bilateral for L7 and L8 as well. S2 Granger causes two out of three practical problems in long CRT: L6 and L9. There is no Granger causality between S3 and the logical problems in long CRT.

Due to these in depth correlations and Granger causalities between short CRT and long CRT, the long CRT might be used instead of the short CRT, being less known compared to the short one.

References

- Barr, N., Pennycook, G., Stolz, J. A. and Fugelsang, J. A. (2015, July). The brain in your pocket: Evidence that Smartphones are used to supplant thinking. *Computers in Human Behavior*, 48, pp. 473–480, <https://doi.org/10.1016/j.chb.2015.02.029>.
- Bialek, M. and Domurat, A. (2017, August). Cognitive abilities, analytic cognitive style and overconfidence: A commentary on Duttie (2016). *Bulletin of Economic Research*, 70(1), pp. 119–125, <https://doi.org/10.1111/boer.12117>.
- Campitelli, G. and Gerrans, P. (2014). Does the cognitive reflection test measure cognitive reflection? A mathematical modeling approach. *Memory & Cognition*, 42, pp. 434–447, <https://doi.org/10.3758/s13421-013-0367-9>.
- Campitelli, G. and Labollita, M. (2010, June). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision Making*, 5(3), pp. 182–191, <https://www.sas.upenn.edu/~baron/journal/10/91230/jdm91230.pdf>.
- Cokely, E. T. and Kelley, C. M. (2009, February). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgement and Decision Making*, 4(1), pp. 20–33, <http://journal.sjdm.org/81125/jdm81125.pdf>.
- De Neys, W., Rossi, S., and Houdé, O. (2013). Bats, balls, and substitution sensitivity: Cognitive misers are no happy fools. *Psychonomic Bulletin & Review*, 20, pp. 269–273, <https://doi.org/10.3758/s13423-013-0384-5>.
- Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, 19(4), pp. 25–42, DOI: 10.1257/089533005775196732.
- Haigh, M. (2016, September). Has the standard cognitive reflection test become a victim of its own success?. *Advances in Cognitive Psychology*, 12, pp. 145–149, doi: [10.5709/acp-0193-5](https://doi.org/10.5709/acp-0193-5).
- Liberali, J., Reyna, V., Furlan, S., Stein, L. and Pardo, S. (2012). Individual Differences in Numeracy and Cognitive Reflection, with Implications for Biases and Fallacies in Probability Judgment. *Journal of Behavioral Decision Making*, 25, pp. 361–381, doi: 10.1002/bdm.752.
- Mata, A., Ferreira, M. B. and Sherman, S. J. (2013). The metacognitive advantage of deliberative thinkers: A dual-process perspective on overconfidence. *Journal of Personality and Social Psychology*, 105, pp. 353–373, <https://doi.org/10.1037/a0033640>.
- Mastrogiorgio, A. and Petracca, E. (2014, June). Numerals as triggers of System 1 and System 2 in the 'bat and ball' problem. *Mind & Society*, 13(1), pp. 135–148, doi: 10.1007/s11299-014-0138-8.

- Pennycook, G., Cheyne, J. A., Koehler, D. J. and Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition?. *Behavior Research Methods*, 48, pp. 341–348, <https://doi.org/10.3758/s13428-015-0576-1>.
- Pennycook, G., Fugelsang, J. A. and Koehler, D. J. (2015a). Everyday consequences of analytic thinking. *Current Directions in Psychological Science*, 24, pp. 425–432, <https://doi.org/10.1177/0963721415604610>.
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A. and Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making*, 29, pp. 453–469, <https://doi.org/10.1002/bdm.1883>.
- Stanovich, K., West, R. and Toplak, M. (2016). *The rationality quotient: Toward a test of rational thinking*. Cambridge: MIT Press, <https://mitpress.mit.edu/books/rationality-quotient>.
<https://doi.org/10.7551/mitpress/9780262034845.001.0001>
- Stieger, S. and Reips, U.-D. (2016, September). A limitation of the cognitive reflection test: Familiarity. *PeerJ*, 4(2):e2395, DOI: 10.7717/peerj.2395.
- Szollosi, A., Bago, B., Szászi, B. and Aczel, B. (2017, August). Exploring the determinants of confidence in the bat-and-ball problem. *Acta Psychologica*, 180, pp. 1-7, DOI: [10.1016/j.actpsy.2017.08.003](https://doi.org/10.1016/j.actpsy.2017.08.003).
- Tamaş, A. (2019). Correlation Investigation: The Cognitive Reflection Test and the Math National Evaluation scores . *International Journal of Teaching and Education*, VII(1), pp. 92-103, DOI: 10.20472/TE.2019.7.1.007.
- Thomson, K. S. and Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision Making*, 11, pp. 99–113, <http://journal.sjdm.org/15/151029/jdm151029.pdf>.
- Toplak, M. E., West, R. F. and Stanovich, K. E. (2011). The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks. *Memory & Cognition*, 39, pp. 1275–1289, <https://doi.org/10.3758/s13421-011-0104-1>.
- Van Der Walt, M. and Potgieter, E. (2019). Design-based research: Mathematics teachers' professional development and low cost mathematics learning resources. *Book of Abstracts of the 8th Teaching & Education Conference, Vienna, Austria*, ISBN 978-80-87927-90-8, IISES, pp. 54.
- 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.