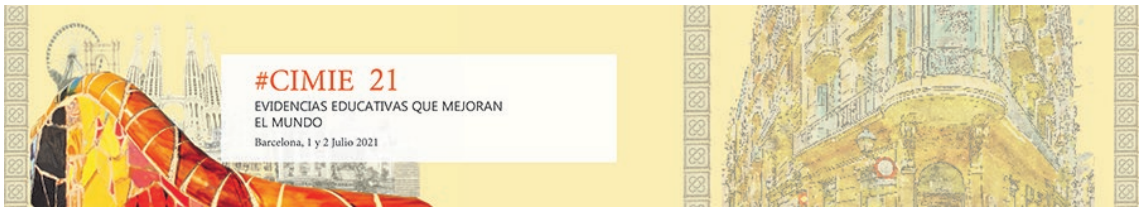

The cognitive reflection test: the intuitive and analytical thinker

Blanca Rosa Olalde Lopez de Arechavaleta, UPV/EHU

Resumen

The Cognitive Reflection Test or CRT is an instrument designed to distinguish the intuitive thinking processing from the analytical one. Gender has a significant impact on CRT scores. On average, low CRT scores are common. Although the CRT score typically employed is the sum of correct responses, some recent researchers have proposed new ways of scoring CRT responses by differentiating intuitive responses from intuitive incorrect responses. These results provide evidence that this way of scoring gives more details on the intuitive and reflective thinkers and the type of errors they make.

Palabras clave:

Keywords: *analytical thinking, biases, Cognitive Reflection Test, intuitive thinking.*

Objetivos o propósitos:

This work aims to study the cognitive reflection by means of CRT or Cognitive Reflection test (Frederick, 2005) and observe the gender difference in CRT performance.

Although this test has been widely used in the English speaking countries, fewer studies exit outside this context.

This study seeks to understand the basis of performance on the CRT and the understanding of sex differences and the level of difficulty of the items.

Marco teórico:

Cognitive reflection is defined as the tendency for individuals to suppress an intuitive and spontaneous response that turns out to be incorrect; in favor of a more reflective,

deliberative and analytical response that is the correct one when making a decision or judgment (Toplak et al., 2014).

Dual process theories distinguish between two fundamentally different types of cognitive processes: type-1 processes—fast, intuitive, effortless, automatic, and autonomous—and type-2 processes—effortful, deliberative, cognitively demanding and governed by rules. In accordance with the framework of the dual process theories, different measuring instruments for the cognitive reflection have emerged.

The CRT is a three-item test to measure the tendency to take an intuitive versus analytic approach to problems. CRT questions tend to elicit a rapid, intuitive but incorrect response. In order to find the correct answer, participants need reflection and deliberation on each question. Incorrect responses need correction through deliberative processes. Giving the correct answer requires type 2 processes to override an intuitive but erroneous judgment produced by type 1 processes.

According to Kahneman (2011), most of our day-to-day decisions are made through type 1 thinking, with type 2 thinking intervening to override these intuitive assessments when they go wrong. As type 2 thinking is often too effortful, many of these wrong answers go unnoticed, especially when they seem to be correct. Type 1 thinking quickly proposes intuitive answers to judgement problems as they arise, while type 2 thinking monitors the quality of these proposals, which it may endorse, correct or override (Kahneman and Frederick, 2005).

As Stanovich (2011) also indicates, a heuristic response is the default and it can occur in three different ways: (a) when a person lacks the 'mindware' to solve a problem; that is, they do not possess the necessary knowledge for rule-based reasoning, (b) when they fail to see the need to engage Type 2 reasoning, they will also give this response, (c) when they lack the cognitive capacity to solve a problem reasoning. People can only solve a problem by reasoning when they intervene with Type 2 reasoning and have the necessary mindware and the necessary cognitive capacity.

Evidence indicates correlation between CRT response times and accuracy (Jimenez et al., 2018). In general, response times indicate that correct answers take longer than incorrect answers. In this case, longer response times are generally considered evidence of longer deliberation and use of type 2 processes.

The CRT has mainly been characterized as a measure indicating miserly information processing (Toplak et al., 2014), but basic numerical competence has also been shown to be an important predictor of task performance (Liberali et al., 2012). Although they have concluded that the CRT is not just another numeracy scale.

Although the required mathematical operations on the CRT are not complicated and the correct solution is easily understood when explained, people tend to perform poorly.

Frederick (2005) reported a mean score of 1.24 correct answers among university students, sample means ranged from 0.57 to 2.18. The mean for the University of Toledo was just 0.57 and for Massachusetts Institute of Technology was 2.18.

CRT also varies with level of education, with participants who have never attended university scoring lowest, then current university students, and graduates the highest (West eta al., 2013).

Gender has a significant impact on CRT scores and men usually score higher than women (Barcellos et al., 2017; Frederick, 2005; Hoppe and Kusterer, 2011; Oechssler et al., 2009). Moreover, Frederick (2005) found that women presented more intuitive answers for each question on CRT, whereas men presented more diverse kinds of wrong answers to the questions. The meta-analysis by Brañas-Garza et al. (2015) shows that men perform better in every single question of the CRT; women are more likely to answer none of the questions correctly; and men are more likely to answer all three questions correctly. Some studies (Primi et al., 2017) on the gender differences show that the difference in performance between men and women may be due to differences in math anxiety.

Thinking analytically actually does have a meaningful impact on beliefs and behaviors, increasing the accuracy of beliefs. CRT correlates with a wide range of factors and domains, such as religious disbelief (Pennycook et al., 2016), moral judgments, values, and behavior (Paxton et al., 2012; Royzman et al., 2015), or the ability to discern fake news from real news (Pennycook and Rand, 2018). Several studies have also found correlations between the CRT and a large number of decision biases. For example, people with higher CRT scores are less likely to commit conjunction fallacies and base-rate neglect (Alós-Ferrer and Hügelschäfer, 2016; Hoppe and Kusterer, 2011; Oechssler et al., 2009).

The original three items have become increasingly popular and familiar, especially the *famous* bat-and-ball *item,* and the CRT is in need of supplement and extension. Besides, from the point of view of reliability, three items are obviously too few. Using the three-item version in some contexts would be problematic in terms of floor effects. A major limitation is the difficulty of the items, which can lead to floor effects in populations other than highly educated adults. Even educated reasoners fail to solve the problem correctly as numerous studies have shown.

Toplak et al. (2014) proposes a four-item version that can serve as an alternative or a substitute for the original CRT. Alternatively, a seven-item version could also be used including the three classic items to provide a more comprehensive test.

Further extended versions of the CRT have been lately created (Baron et al., 2014; Primi et al., 2017; Thomson and Oppenheimer, 2016). On the other hand, there is a need for more verbal CRT-type questions (Baron et al., 2014) in an attempt to examine the effects of cognitive reflection in the absence of numerical load.

The CRT measure is the number of correct answers. The CRT provides a range of scores between zero and three correct answers. Participants with an overall CRT-score of zero correct answers are considered as lowest cognitive reflection capabilities and those with all three correct answers as highest cognitive reflection capabilities. Traditionally, a CRT score of 0 or 1 indicates an intuitive thinker while 2 and 3 mean an analytical thinker. Some researchers have recently recognized the importance of scoring and distinguishing incorrect responses that are distinct from the standard intuitive incorrect responses (Pennycook et al., 2016; Stupple et al., 2017).

Metodología:

The sample of 993 people (563 women and 430 men) aged 12 to 59, living in the Basque Country (female: M = 19, SD = 5,913, male: M = 19, SD = 5,141). The participants come from nine educational centers and from different levels; Secondary School (19.2%), Vocational Education and Training (5.1%), High School (10.9%), Higher Education (4.7%), University Degree (51.4%) and Graduate Certificate (5.6%). The sample was selected in a non-probabilistic manner.

The subjects filled out the CRT I (Frederick, 2005) containing 3 items: (1) The bat and ball, (2) The machines and time, and (3) The water lily and the lake.

The traditional methodology that scores the number of correct answers (possible scores 0, 1, 2, 3) is followed in order to study the gender differences.

Thus, two groups are differentiated. Those subjects with two or more correct answers make up the group of reflective or analytical people (10.3 % of the sample) compared to those who have given two or more incorrect answers that are considered more impulsive or intuitive (89.7 % of the sample).

Discusión de los datos, evidencias, objetos o materiales:

Indeed, the results confirm the level of difficulty of the test, where almost the 90% of the participants occupy the lowest scores. The CRT has shown a very high level of difficulty and, therefore, little sensitivity to detect the reflective thinking of many subjects.

When following the criterion of the number of correct answers, the general results indicate the existence of sex differences (X_2 =.000). When analyzing the number of correct answers according to the sex (See Table 1), it is observed that there are more women than men who do not give any correct in the CRT. In addition, the number of men who give 1, 2 or 3 correct answers is greater than that of women. On average, the men resolved .55 of the three items in the CRT (SD = .81) compared to .30 (SD =.62) among the women.

Table 1. Percentage of correct answers

	CRT		
Number of correct answers	M	F	Total
0	62.7	77.5	71.1
1	21.9	16.1	18.6
2	12.7	5.0	8.3
3	2.8	1.4	2.0

Considering the grouping of the subjects into these two groups (reflective vs impulsive) the differences are significant (X_2 = 22.042, p = .005).

Table 2. Average scores obtained by men and women

	Sex	*N*	Mean	Standar Desviation	Mean standard error
CRT	Men	434	0.5553	0.81721	.03923
	Women	565	0.3027	0.62915	.02647

When each item is addressed (See Table 3), the results indicate that while there are no differences in the first item (The bat and the ball), differences can be seen in the second (Machines and time) and the third (The water lily and the lake) items. Thus, despite the difference in the percentages of the correct responses by men (11.5%) and women (9.5) in item 1, the difference is not significant (Pearson's X2 = 300).

The measure of reflective thinking obtained through the number of correct answers in the CRT I appears to be significantly related to gender, when applying both parametric and non-parametric techniques (u = 95279.500, p <.001), which indicates that both groups have a different behavior with respect to this measure of cognitive reflection.

According to the previous literature (Frederick, 2005), the most difficult problem to be solved has also been the first item (in this sample 10.4%) and it is likely that this is the origin of some difficulties to detect the differences, which are statistically significant in items 2 and 3 (Pearson's X2 = .000, in both cases).

Table 3. Percentage of subjects with correct answers in each item.

	M	F	Total	X^2 Pearson
CRT I				
Item 1	11.5	9.5	10.4	0.300
Item 2	21.2	9.5	14.6	0.000
Item 3	24.0	12.8	17.8	0.000

This data reveals that sex was correlated with performance on the CRT. Specifically, men score higher than women and these differences are significant.

Resultados y/o conclusiones:

The results of this investigation show the clear sexist bias in the Cognition Reflection Test (CRT) by Frederick (2005) as a measure of the cognitive reflection in people, which is consistent with previous research. Frederick's CRT scores still exhibited considerable difference between men and women. The gender difference in CRT performance has since been observed in a large number of investigations (Pennycook et al., 2016; Primi et al., 2016; Zhang et al. 2016).

Even though these gender differences in the Cognitive Reflection test (CRT) appear to be robust across multiple studies, little research has examined the source of the gender gap in performance. Findings suggest that cognitive reflection may be affected by numerical skills and related math anxiety feelings.

On the other hand, the bat and ball question is the most difficult on average. These results coincide with those obtained in the meta-analysis by Brañas-Garza et al. (2015) where only 32% of all participants get it right, compared with 40% for the widgets and 48% for the lilypads. As observed, means in this work are lower than average means in previous studies due to the young age of the sample and the educational level. Most of the previous studies on CRT usually present an adult sample.

Contribuciones y significación científica de este trabajo:

These results offer compelling evidence for the need to create of new items and tests that may be less affected by the sex bias. Some authors have suggested that the gender gap in CRT performance may be attributed to anxiety toward solving mathematical problems. Feelings of anxiety, and inadequacy toward math and lack of self-efficacy may affect cognitive reflection negatively (Morsanyi et al., 2014) Gender differences can be significantly reduced when the effects of numeracy are controlled (Primi et al., 2016). Reducing the mathematical nature of the items and introducing new ways of analyzing the answers could be useful.

As many authors have pointed out, it is advisable to expand the measurement of the cognitive reflection. It is strongly recommending to include new intuitive vs. reflection items (Primi et al., 2017; Toplak et al., 2014).

Differentiating the intuitive responses from other types of incorrect answers (Erceg and Bubić, 2017) is recommended. Some protocol analysis to unfold the steps of the reasoning process in solving the CRT is recommendable (Szaszi et al., 2017), because there are several ways people solve or fail the test.

Bibliografía:

Alós-Ferrer, C., Garagnani, M., & Hügelschäfer, S. (2016). Cognitive reflection, decision biases, and response times. *Frontiers in Psychology, 7*, 1402. doi:10.3389/fpsyg.2016.01402

Barcellos, L. P. (2017). *Cognitive reflection abilities and accounting practice: A two-way road of influences.* Road of Influences.

Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2014). Why does the cognitive reflection test (sometimes) predict utilitarian moral judgment (and other things)? *Journal of Applied Research in Memory and Cognition, 4*, 265–284.

Brañas-Garza, P., Kujal, P., & Lenkei, B. (2015). Cognitive reflection test: Whom, how, when. *Journal of Behavioral and Experimental Economics*, *82,* 101455. doi:10.1016/j.socec.2019.101455

Erceg, N., & Bubić, A. (2017). One test, five scoring procedures: Different ways of approaching the cognitive reflection test. *Journal of Cognitive Psychology, 29*(3), 381-392. doi:10.1080/20445911.2016.1278004

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

Hoppe, E. I., & Kusterer, D. J. (2011). Behavioral biases and cognitive reflection. *Economics Letters, 110*(2), 97-100. doi:10.2139/ssrn.1488752

Jimenez, N., Rodriguez-Lara, I., Tyran, J., & Wengström, E. (2018). Thinking fast, thinking badly. *Economics Letters, 162*, 41-44. doi: 10.1016/j.econlet.2017.10.018

Kahneman, D. (2011). *Thinking, fast and slow.* Macmillan.

Kahneman, D., & Frederick, S. (2005). A model of intuitive judgment. *The Cambridge Handbook of Thinking and Reasoning*, 267-293.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making, 25*(4), 361-381. doi:10.1002/bdm.752

Oechssler, J., Roider, A., & Schmitz, P. W. (2009). Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization, 72*(1), 147-152. doi: 10.1016/j.jebo.2009.04.018

Paxton, J. M., Ungar, L., & Greene, J. D. (2012). Reflection and reasoning in moral judgment. *Cognitive Science, 36*(1), 163-177. doi:10.1111/j.1551-6709.2011.01210.x

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2016). Is the cognitive reflection test a measure of both reflection and intuition? *Behavior Research Methods*, *48*(1), 341-348.

Pennycook, G., Ross, R. M., Koehler, D. J., & Fugelsang, J. A. (2016). Atheists and agnostics are more reflective than religious believers: Four empirical studies and a meta-analysis. *PloS One, 11*(4), e0153039. doi: 10.1371/journal.pone.0153039

Pennycook, G., & Rand, D. G. (2020). Who falls for fake news? The roles of bullshit receptivity, overclaiming, familiarity, and analytic thinking. *Journal of Personality, 88*(2), 185-200. doi:10.1111/jopy.12476

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & 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*(5), 453-469.

Primi, C., Morsanyi, K., Donati, M. A., Galli, S., & Chiesi, F. (2017). Measuring probabilistic reasoning: The construction of a new scale applying item response theory. *Journal of Behavioral Decision Making, 30*(4), 933-950. doi:10.1002/bdm.1883

Stanovich, K. (2011). *Rationality and the reflective mind.* Oxford University Press.

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. doi:10.1080/13546783.2017.1292954

Stupple, E. J., Pitchford, M., Ball, L. J., Hunt, T. E., & Steel, R. (2017). Slower is not always better: Response-time evidence clarifies the limited role of miserly information processing in the cognitive reflection test. *PloS One, 12*(11), e0186404. doi:10.1371/journal.pone.0186404

Toplak, M. E., West, R. F., & Stanovich, K. E. (2014). Rational thinking and cognitive sophistication: Development, cognitive abilities, and thinking dispositions. *Developmental Psychology, 50*(4), 1037-1048.

Welsh, M., Burns, N., & Delfabbro, P. (2013). The cognitive reflection test: How much more than numerical ability? In *Proceedings of the Annual Meeting of the Cognitive Science society* (Vol. 35, No. 35).

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