Cognitive Reflection Test & Alternatives

Matthew R. Arp

Research Associate, LISPOP, Wilfrid Laurier University

Abstract

The Cognitive Reflection Test (CRT) is a three-item test that is widely used in survey research and by academic researchers to measure cognition, specifically reflection. In recent years, researchers have grown increasingly concerned about the utility of the CRT, developing several alternatives tests that strongly warrant consideration.

Keywords: Survey Research, Cognitive Reflection Test, CRT

Introduction

The Cognitive Reflection Test (CRT) is an extremely influential tool used by survey methodologists and practitioners in experimental social sciences such as psychology, economics and political science. It is a simple three-item test that can be used to measure a specific type of cognition, specifically reflection. The CRT is a rubric for measuring respondents on an intuitive-analytic scale. Questions are crafted in such a way that the incorrect response is the obvious and most intuitive, while the correct response requires deeper analysis.[1] There is evidence that demonstrates correlation between CRT scores and other tests of risk/reward behaviour and delayed gratification. However, there are some concerns regarding the limitations of the CRT due to overexposure and significant gender differences. As a result, researchers have developed a myriad of alternative tests that maintain the structural components of the CRT, while overcoming some of the limitations.

Features of the CRT

The CRT is a three-question test developed by Shane Frederick in 2005 as a simple measurement of reflective cognitive ability.[2] The CRT is not an intelligence test, such as an IQ exam, but rather a measure of Type 2 cognition. The dual process model of human cognition distinguishes between Type 1 and Type 2 forms of cognition. Type 1 can be characterized as quick and spontaneous thinking, while Type 2 is more deliberate and reasoned thinking.[3] In other words, the CRT measures the respondents' ability to rationally approach a problem and resist the urge to answer spontaneously.

The three-items included on the CRT are the following:

1. A bat and a ball cost $110 in total. The bat costs $100 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 lake, how long would it take for the patch to cover half the lake?

Each of these three questions has both an intuitive and quick incorrect answer and a more deliberate and well-reasoned correct answer. For the first question, the spontaneous answer may seem to be $10, but upon further examination and thought the correct answer is actually $5. Each of these questions are designed to "trick" respondents into believing the answer is simple when it is more complicated. Frederick described the CRT as "easy in the sense that their solution is easily understood when explained, yet reaching the correct answer often requires the suppression of an erroneous answer that springs "impulsively" to mind".[4]

The use of CRT questions in survey research is a proxy measurement for this Type 2 cognition, but has also proven to be strongly correlated with other respondent behaviours. Research has found a strong relationship between CRT performance and instant gratification. Respondents that performed higher on the CRT were also more likely to accept a larger sum of money in the future than less money right now ($100 now vs. $105 next week). CRT performance is consistent with the performance in other forms of standardized testing, such as the ACT or SAT. Additionally, CRT scoring is strongly correlated with high risk/high reward behaviour (e.g. choosing $500 guaranteed or a 15%

chance of $1,000,000).[5] Additionally, the CRT can serve as a predictor of performance on heuristics and bias.[6]

Criticisms of the CRT

While the CRT has been widely used in academic research to measure cognitive ability, it is not without its shortcomings. One of the major concerns surrounding the CRT is overexposure and oversaturation. Researchers have grown concerned that since the introduction and widespread adoption of the CRT, respondents are more prepared for the specific questions. Rather than accurately gauging the reflective cognition of respondents, they are merely regurgitating the correct answers.[7] In one study, over half of respondents surveyed have been exposed to at least one of the CRT questions, from previous research studies, social media, books, websites or school. Respondents that were exposed to the CRT, on average, answered one additional question correctly.[8] Other researchers, including Frederick, downplay the effect exposure has on performance. While he found that scores do improve over time for respondents that are continually exposed to CRT questions, CRT performance increases a measly 0.024 increase per exposure.[9] Additionally, these scholars argue that bias is a feature, not a bug of overexposure. In their opinion, respondents that take the time to assess their past performance and dedicating effort to remembering solutions in the future actually means they are becoming more reflective.

Another major concern of the CRT is the gap between correct answers and gender. On the three question CRT panel, male respondents tend to score higher (1.47) compared to females (1.03). However, gender was not a statistically significant factor when comparing SAT, ACT, Wonderlic or NFC scores. The creator of the CRT notes that "It appears, instead, that these items measure something that men have more of. That something may be mathematical ability or interest, since the CRT items have mathematical content, and men generally score higher than women on math tests".[10] However, even when controlling for SAT math scores, male respondents scored higher at a statistically significant level (p < 0.0001) relative to females. Scholarship does not have a definite explanation to these relationships, while other scholars have noted the connection between math ability and the CRT.

The last criticism of the CRT is that it does not accurately measure Type 2 or cognitive ability, instead measuring math ability or numeracy. Scholars have found that performance on the CRT strongly correlates with performance on other standardized tests, including the SAT, Wonderlic and WASI Vocabulary and Matrix Reasoning scale. In fact, the strongest correlation between tests with the CRT was the Numerical Ability Test (NAT).[11] The NAT is a test that assesses the ability to interpret numerical data and drawing conclusions from data presented. It requires a mastery of basic math knowledge including but not limited to: arithmetic, percentages, ratios and averages.[12]

Alternatives

In order to overcome the aforementioned obstacles, researchers have begun to develop alternative tests that also measure cognitive ability. Even minor changes to the questions, such as changing simple wording (e.g. bat and ball) may be novel enough to remove any bias.[13] The consensus is to develop alternative questions that are different from the standard 3-item test. This includes an extended four and seven-item version developed by Toplak, et al (2014)[14], the four-item CRT-2 developed by Thomson & Oppenheimer (2016)[15] and the six-item CRT-Long developed by Primi, et al (2015).[16][17]

Each of these alternatives maintains the same style as the original CRT with an intuitive/incorrect and analytic/correct question format. They have two distinct advantages: (1) exposure and (2) less dependent on math knowledge. Because these questions have not been in circulation as long as the original CRT, respondents are less prepared to answer these questions when encountering them. This provides a more unbiased measurement of cognition, rather than regurgitation. Additionally, these questions do not require mathematical sophistication to determine the correct answer, which also achieves the goal of more accurately measuring Type 2 cognition. Because of this, performance on these alternatives is less dependent on gender and math ability.

Conclusion

The CRT is a novel and useful tool that can be distinguished from other "intelligence" or "knowledge" tests. Rather than measuring computational ability, the CRT is based on the dual-system theory of cognition. After being in circulation for over 15 years, there is little consensus regarding the future of the CRT. However, for researchers concerned about its longevity, there are alternatives that capture the same pheonmena as the original CRT with less drawbacks.

Biographical Information

Matthew R. Arp earned his Ph.D. in Political Science from West Virginia University. His research interests include great power conflict and international norm creation. He joined LISPOP as a research associate with a background in large-N data collection, including the During-Conflict Justice (DCJ) and Northern Ireland Research Initiative (NIRI).

References

[1] 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.

[2] Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25-42.

[3] Evans, J. S. B. (2008). Dual-processing accounts of reasoning, judgment, and social cognition. *Annu. Rev. Psychol.*, *59*, 255-278.

[4] Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 26.

[5] Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 25-42.

[6] 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.

[7] Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, *11*(1), 99; Baron, J., Scott, S., Fincher, K., & Metz, S. E. (2015). Why does the Cognitive Reflection Test (sometimes) predict utilitarian moral judgment (and other things)?. *Journal of Applied Research in Memory and Cognition*, 4(3), 265-284..

[8] Haigh, M. (2016). Has the standard cognitive reflection test become a victim of its own success?. *Advances in cognitive psychology*, *12*(3), 145.

[9] Meyer, A., Zhou, E., & Shane, F. (2018). The non-effects of repeated exposure to the Cognitive Reflection Test. *Judgment and Decision making*, *13*(3), 246.

[10] Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic perspectives*, *19*(4), 37.

[11] 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);

[12] Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, *11*(1)

[13] Chandler, J., Mueller, P., & Paolacci, G. (2014). Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. *Behavior research methods*, *46*(1), 112-130.

[14] The four additional questions include the following: (1) 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? (2) Jerry received both the 15th highest and the 15th lowest mark in theclass. How many students are in the class? (3) A man buys a pig for $60, sells it for $70, buys it back for $80, andsells it finally for $90. (4) Simon 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 pur-chased 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. broken even in the stock market, b. is ahead of wherehe began, c. has lost money.

[15] The CRT-2 includes the following questions: (1) If you're running a race and you pass the person in sec-ond place, what place are you in? (2) A farmer had 15 sheep and all but 8 died. How manyare left? (3) Emily's father has three daughters. The first two arenamed April and May. What is the third daughter's name? (4) How many cubic feet of dirt are there in a hole that is 3'deep x 3' wide x 3' long?

[16] The CRT-Long includes the original three CRT questions as well as the following three additional questions: (1) If three elves can wrap three toys in hour, how many elves are needed to wrap six toys in 2 hours? (2) Jerry received both the 15th highest and the 15th lowest mark in the class. How many students are there in the class? (3) In an athletics team, tall members are three times more likely to win a medal than short members. This year the team has won 60 medals so far. How many of these have been won by short athletes?

[17] 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; Thomson, K. S., & Oppenheimer, D. M. (2016). Investigating an alternate form of the cognitive reflection test. *Judgment and Decision making*, *11*(1), 99; 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.