Personality and Individual Differences 101 (2016) 425–427

Contents lists available at ScienceDirect

Personality and Individual Differences

journal homepage: www.elsevier.com/locate/paid

Short Communication

Explaining sex differences on the Cognitive Reflection Test

Don C. Zhang [a,*], Scott Highhouse [b], Thaddeus B. Rada [c]

[a] Department of Psychology, Louisiana State University, Baton Rouge, LA 70803, United States
[b] Bowling Green State University, United States
[c] Edinboro University, United States

ARTICLE INFO

Article history:
Received 2 March 2016
Received in revised form 21 May 2016
Accepted 17 June 2016
Available online xxxx

Keywords:
Cognitive reflection
Intuition
Numeracy
Sex differences
Gender effects

ABSTRACT

The Cognitive Reflection Test (CRT; Frederick, 2005) is a three-item performance-based measure designed to assess one's tendency to over-ride automatic responses in favor of further reflection. Although the test has been widely cited, and predicts varied outcomes, little is known about the sex differences observed in the initial report. This study found a 0.37 standard deviation difference between men and women in a large adult sample of respondents. This difference could be explained entirely by differences in quantitative self-efficacy.

Published by Elsevier Ltd.

1. Introduction

The Cognitive Reflection Test is a three-item performance-based measure designed to assess one's tendency to over-ride automatic responses in favor of further reflection. The measure is based on dual-system theories of decision making (e.g., Kahneman, 2003; Sloman, 1996). Generally, these approaches suggest that a heuristic system (System 1) is automatically engaged and will guide decisions unless it is interrupted by more reflective thought (System 2). The CRT is based on the idea that people differ in the ease with which heuristics may be interrupted by careful reflection. One of the three items is as follows:

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

People presented with this problem often give the first response that comes to mind (i.e., 10 cents). Further reflection, however reveals that a 10-cent ball would result in a total price of $1.20. The correct answer is 5 cents. Although the CRT is correlated with cognitive ability and numeracy, research suggests that it explains incremental variance in outcomes, including performance on various judgment and decision tasks (Baron, Scott, Fincher, & Metz, 2015; Liberali, Reyna, Furlan, Stein, & Pardo, 2011; Toplak, West, & Stanovich, 2011) as well as a (dis)belief in God (Shenhav, Rand, & Greene, 2011). Toplak et al. (2011) concluded that the CRT is "a particularly potent measure of the tendency toward miserly processing because it is a performance measure rather than a self-report measure" (p. 1275). A recent Google Scholar search shows that Frederick's (2005) article on the development of the CRT has been cited over 1500 times.

Frederick's (2005) original report of research on the CRT noted that men scored significantly higher than women on the test. He observed that the difference between men and women on the CRT remains even after controlling for SAT math scores. Frederick concluded that the sex effect observed on the CRT remains a mystery. The gender difference in CRT performance has since been observed in several independent investigations (Campitelli & Gerrans, 2014; Pennycook, Cheyne, Koehler, & Fugelsang, 2015; Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2015). The size of the differences in these studies range between Cohen's *d* of 0.32 to 0.46. Other studies using the CRT do not provide data on sex differences (e.g., Liberali et al., 2011) or the data are too skewed to reliably interpret group differences (e.g., Campitelli & Labollita, 2010). Even though gender differences in CRT appear to be robust across multiple studies, little research has examined the source of the gender gap in performance.

The purpose of the current study was to examine (a) the effect size of sex differences on the CRT in a representative adult sample, (b) the relative importance of sex, compared with individual differences in test self-efficacy, quantitative self-efficacy (a.k.a., subjective numeracy), and decision style (i.e., rational and intuitive) in explaining responses to the CRT, and (c) whether these dispositional variables could explain most or all of the differences between men and women on the CRT. Research has observed reliable sex differences in test self-efficacy (Everson, Millsap, & Rodriguez, 1991) and quantitative self-efficacy (Bandalos, Yates, & Thorndike-Christ, 1995) with women scoring

* Corresponding author at: Department of Psychology, Louisiana State University, Baton Rouge, LA 70803, United States.
E-mail address: zhang1@lsu.edu (D.C. Zhang).

http://dx.doi.org/10.1016/j.paid.2016.06.034
0191-8869/Published by Elsevier Ltd.

lower than men. Although the research on sex differences in decision making styles has been scant and inconsistent, some research has suggested that women score higher than men on measures of intuitive styles, and lower on measures of rational styles (see Allinson & Hayes, 2012). Our study seeks to understand the basis of sex differences on the CRT, and the contribution of sex to the understanding of performance on the CRT.

2. Method

2.1. Sample

The sample consisted of 205 American adults recruited via Amazon's Mechanical Turk (MTurk, see Buhrmester, Kwang, & Gosling, 2011). Participants received 50 cents for completing on-line the CRT, along with other measures. 67% of the participants were between the ages of 18–34. 77% were white, and 50% were male.

2.2. Measures

2.2.1. Cognitive Reflection Test (CRT)

The CRT measures "the ability or disposition to resist reporting the response that first comes to mind" (p. 35, Frederick, 2005). It is a three-item, open-ended performance test, requiring respondents to fill in the correct numerical response. Responses were coded as either incorrect = 1 or correct = 2.

2.2.2. Quantitative self-efficacy (QSE)

Respondents' perceived fluency with numerical information was assessed using the subjective numeracy scale (Fagerlin et al., 2007). This is an eight-item measure of perceived competence with quantitative information (e.g., How good are you at working with fractions?). Responses are provided on a 6-point scale, ranging from 1 (not at all good) to 6 (extremely good). Although the scale was originally developed as an alternative measure of numeracy, Liberali et al. (2011) found that subjective and objective numeracy load on separate factors. We believe it is best to view the subjective numeracy measure as a measure of self-efficacy (Bandura, 1986) in the quantitative domain.

2.2.3. Test self-efficacy (TSE)

Test self-efficacy was measured with three items adapted from Arvey, Strickland, Drauden, and Martin (1990). The items were: "I am confident in my test-taking abilities," "I know, when it comes to taking tests, that I do well," "I am afraid of taking standardized tests." Responses were provided on a 5-point scale (1 = strongly disagree; 5 = strongly agree).

2.2.4. Decision making style

The *rational* and *intuitive* sub-dimensions of the general decision making style inventory (Scott & Bruce, 1995) were administered. Each scale contains five items. A higher score on any of the five scales indicates a higher presence of that particular decision-making style. Someone high on the rational scale uses a logical process of considering alternatives to arrive at the best possible decision (e.g., "My decision making requires careful thought."). Someone high on the intuitive scale relies on feelings or hunches to make decisions ("When making a decision, I rely upon my instincts."). Responses were provided on a 5-point scale (1 = strongly disagree; 5 = strongly agree).

2.3. Analytic procedures

Cohen's *d* was calculated to examine the difference between males and females on the CRT in standard deviation units. Relative weight analysis (Tonidandel & LeBreton, 2011) was used to examine the extent to which sex drives the prediction of CRT responses, relative to the dispositional variables. This analysis solves the problem of predictor intercorrelation by using a variable transformation approach to create a new set of orthogonal predictors. The resulting metrics sum to the model R^2 allowing for the interpretation as measures of relative effect size. Regression analyses were used to examine whether sex accounts for incremental variance in the CRT, after controlling for the other predictors. Finally, a mediation analysis was conducted to examine if individual differences in quantitative self-efficacy mediated the gender difference in CRT performance.

3. Results and discussion

The intercorrelations of the study variables are presented in Table 1. Inspection of the table reveals that sex was correlated with performance on the CRT. Specifically, men (M = 1.57; SD = 0.39) score higher than women (M = 1.42; SD = 0.42) to a significant degree, $t(203) = 2.67$; $p = 0.008$; Cohen's $d = 0.37$. Inspection of Table 1 also reveals that there were significant sex differences on QSE, TSE, and intuitive decision style. Men and women did not differ on the rational decision style. Moreover, as reviewer one pointed out, the high correlation between QSE and TSE suggests an overlap between the two constructs. Given both measures are indicators of one's self-efficacy in two domains: quantitative ability, and testing ability, we expect that they are domain-specific facets of one's core self-evaluation (Judge & Bono, 2001).

Table 2 presents the results of the relative weight analysis. This table shows that sex is a relatively minor player in the prediction of CRT performance. Of the variance explained by all of the variables in our study, sex accounts for 8% of that explained variance. This can be compared with QSE, which accounts for nearly 54% of the explained variance.

A regression analysis was run in which QSE, TSE, Rational, and Intuitive were entered in the first step. Respondent sex was entered in the second step. The analysis revealed a non-significant change in R^2, $\Delta R^2 = 0.01$; $F = 1.86$; $p = 0.174$. This suggests that sex does not provide unique prediction in CRT performance, over and above the study variables. Together, the variables accounted for 21% of the variability in performance on the CRT.

Because QSE was found to be the dominant predictor in the relative weight analysis, we ran regression with only QSE entered in the first step, and with sex entered in the second step. The analysis revealed a non-significant change in R^2, $\Delta R^2 = 0.01$; $F = 2.21$; $p = 0.139$. It appears that QSE alone explains the sex difference observed on the CRT in our sample.

A mediation analysis was conducted to explore the role of QSE for the gender gap in CRT performance. To estimate the indirect effect of gender on CRT performance through QSE, mediation analysis was conducted with the bootstrapping method with bias-corrected confidence estimates (Preacher & Hayes, 2004). The 95% confidence interval of the indirect effect was obtained with 10,000 bootstrap resamples ($\beta = -0.21$, 95% $CI = -0.37$ to -0.09, $p = 0.005$). The 95% confidence interval for the indirect effect estimate did not contain 0, which indicates that the indirect effect was significant. The direct effect of gender on CRT performance was no longer statistically significant after accounting for QSE ($\beta = -0.23$, 95% $CI = -0.56$ to 0.10, $p = 0.162$). The results showed that QSE fully mediated the relationship between gender and CRT performance.

Table 1
Means, standard deviations, and correlations between study variables.

Scale	M	SD	CRT	QSE	TSE	Rational	Intuitive	Sex
CRT	4.49	1.24	0.76	0.386**	0.253**	0.014	−0.215**	−0.184**
QSE	4.51	0.95		0.85	0.517**	0.353**	−0.089	−0.231**
TSE	3.79	0.91			0.87	0.249**	−0.041	−0.150*
Rational	4.14	0.53				0.78	−0.114	−0.101

Note: Reliabilities in the main diagonal. CRT = Cognitive Reflection Test; QSE = quantitative self-efficacy; TSE = test self-efficacy.
* $p < 0.05$.
** $p < 0.01$.

Table 2
Results of relative weight analysis.

Variable	β	Raw relative weight	Relative weight as % of *R*-square
QSE	0.36**	0.12	53.7%
TSE	0.07	0.03	15.9%
Rational	−0.18*	0.01	4.7%
Intuitive	−0.20**	0.04	17.9%
Sex	−0.10	0.02	7.9%
Model R^2 = 0.214			

Note: β = standardized regression weights.
* $p < 0.05$.
** $p < 0.01$.

4. Conclusion

Our study suggested that men perform 0.37 of a standard deviation better than women on the CRT, which was consistent with previous research. Some have posited that the gender gap in CRT performance may be attributed to heightened anxiety toward solving mathematical problems (Primi et al., 2015). Indeed, we believe both QSE and TSE contributes to test takers' anxiety in math tests, which results in worse performance on the CRT. However, the degree to which the two variables uniquely contributes to test anxiety is beyond the scope of this paper. As shown in this study, sex difference disappears when one's self-assessment of numerical aptitude is statistically controlled. Furthermore, the relationship between sex and CRT performance appears to be mediated by QSE. Men perform better on the CRT because they are more confident in their quantitative abilities. In an education setting, feelings of anxiety and inadequacy toward math negatively affect cognitive reflection (Morsanyi, Busdraghi, & Primi, 2014). Young adults, in particular, are susceptible to the effect of self-confidence on mathematical problem solving (Beilock, 2008). Increasing quantitative self-efficacy could substantially close the gap between men and women on the CRT, as well as improve quantitative reasoning and decision-making. However, the degree to which QSE and TSE uniquely contributes to test anxiety in math and other domains is unclear. More research is needed to identify the relative importance of these two variables in order to better identify targeted interventions and training efforts to reduce anxiety on tests.

Recent research has shown that much of the explanatory power of the original CRT on decision outcomes is explained by mathematical and numerical ability (Sinayev & Peters, 2015; Welsh, Burns, & Delfabbro, 2013). As shown in this study, performance on the CRT is also predicted by QSE, which is different from objective numerical ability (Liberali et al., 2011). Therefore, to better assess cognitive reflection, it might be useful to examine problems that are not quantitative in nature. For instance, Shepard (1990) reports a number of illusions that may be remedied by reflection. The cover of Plous' (1993) *The Psychology of Judgment and Decision Making* shows a black three of hearts, which people rarely notice when first glancing at the book. The Watson-Glaser Critical Thinking Appraisal (Watson & Glaser, 1980) contains a number of inference problems that may reflect one's tendency to respond without reflection. Measuring cognitive reflection in a way that is not influenced as much by quantitative or testing self-efficacy may require a larger and more varied set of items than the current CRT does. However, the resulting measure would be one that better captures cognitive reflection and is not contaminated with quantitative ability or self-efficacy.

References

Allinson, C., & Hayes, J. (2012). *The cognitive style index: Technical manual and user guide.* Pearson Education Ltd.: United Kingdom.

Arvey, R. D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivational components of test taking. *Personnel Psychology*, *43*, 695–716.

Bandalos, D. L., Yates, K., & Thorndike-Christ, T. (1995). Effects of math self-concept, perceived self-efficacy, and attributions for failure and success on test anxiety. *Journal of Educational Psychology*, *87*, 611–623.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice-Hall.

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.

Beilock, S. L. (2008). Math performance in stressful situations. *Current Directions in Psychological Science*, *17*(5), 339–343.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk a new source of inexpensive. *yet high-quality, data?, Perspectives on psychological science*, *6*(1), 3–5.

Campitelli, G., & Gerrans, P. (2014). Does the Cognitive Reflection Test measure cognitive reflection? A mathematical & modeling approach. *Memory & Cognition*, *42*(3), 434–447.

Campitelli, G., & Labollita, M. (2010). Correlations of cognitive reflection with judgments and choices. *Judgment and Decision making*, *5*, 182–191.

Everson, H. T., Millsap, R. E., & Rodriguez, C. M. (1991). Isolating gender differences in test anxiety: A confirmatory factor analysis of the test anxiety inventory. *Educational and Psychological Measurement*, *51*, 243–251.

Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: Development of the subjective numeracy scale. *Medical Decision Making*, *27*, 672–680.

Frederick, S. (2005). Cognitive reflection and decision making. *Journal of Economic Perspectives*, *19*, 25–42.

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—Self-esteem, generalized self-efficacy, locus of control, and emotional stability—With job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology*, *86*(1), 80.

Kahneman, D. (2003). A perspective on judgment and choice: Mapping bounded rationality. *American Psychologist*, *58*, 697–720.

Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2011). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. *Journal of Behavioral Decision Making*. http://dx.doi.org/10.1002/bdm.752.

Morsanyi, K., Busdraghi, C., & Primi, C. (2014). Mathematical anxiety is linked to reduced cognitive reflection: A potential road from discomfort in the mathematics classroom to susceptibility to biases. *Behavioral and Brain Functions*, *10*(1), 1.

Pennycook, G., Cheyne, J. A., Koehler, D. J., & Fugelsang, J. A. (2015). Is the Cognitive Reflection Test a measure of both reflection and intuition? *Behavior Research Methods*, 1–8.

Preacher, K. J., & Hayes, A. F. (2004). SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, *36*(4), 717–731.

Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2015). The development and testing of a new version of the Cognitive Reflection Test applying item response theory (IRT). *Journal of Behavioral Decision Making*.

Plous, S. (1993). The psychology of judgment and decision making. *Mcgraw-Hill Book Company*.

Scott, S. G., & Bruce, R. A. (1995). Decision-making style: The development and assessment of a new measure. *Educational and Psychological Measurement*, *55*, 818–831.

Shenhav, A., Rand, D. G., & Greene, J. D. (2011). Divine intuition: Cognitive style influences belief in God. *Journal of Experimental Psychology: General*, *141*, 423–428.

Shepard, R. N. (1990). Mind sights: Original visual illusions, ambiguities, and other anomalies. *With a commentary on the play of mind in perception and art*. New York: WH Freeman & Co.

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in Psychology*, 6.

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin*, *119*, 3–22.

Tonidandel, S., & LeBreton, J. M. (2011). Relative importance analysis: A useful supplement to regression analysis. *Journal of Business and Psychology*, *26*, 1–9.

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*, 1275–1289.

Watson, G., & Glaser, E. M. (1980). Critical thinking appraisal: Manual. *Psychological Corporation*.

Welsh, M., Burns, N., & Delfabbro, P. (2013). The Cognitive Reflection Test: How much more than numerical ability. *Proceedings of the 35th annual conference of the cognitive science society* (pp. 1587–1592).