What is the CRT? Intelligence, Personality, Decision Style or Attention?

Matthew Brian Welsh (matthew.welsh@adelaide.edu.au)
Australian School of Petroleum, University of Adelaide
North Terrace, Adelaide, SA 5005, Australia

Abstract

The well-known CRT is a test designed to measure a person's 'cognitive reflection' and used as a predictor of decision making ability. Within the literature, however there is a growing consensus that it shares the majority of its variance with numerical and other cognitive abilities and thus the question increasingly asked is whether it has predictive power beyond existing measures. That is, is there something unique captured by the CRT? This study examines the CRT in parallel with a wide range of individual differences measures reflecting aspects of intelligence (8 CHC broad ability factors), personality (the Big 5 and 30 facets), other decision styles (5 measures) and attention (12 measures covering six aspects of attention). Results indicate that the CRT is, primarily, a cognitive measure, strongly linked to fluid, crystallized and quantitative ability but may also be capturing some distinct aspects of attention relating to the ability to ignore distractors.

Keywords: CRT; intelligence; personality; decision style; attention.

Introduction

The cognitive reflection test (Frederick, 2005) was introduced as a measure of a person's tendency to engage in cognitive reflection – thinking on responses before delivering them and thus being more likely to notice and correct errors. As such, it has been linked with lessened susceptibility to biases and better decision making across a wide variety of contexts (see, e.g., Otero, Salgado, & Moscoso, 2022). A key observation made in the original paper and by others using this test is that the CRT seemed to do a better job of predicting people's decision-making ability than their intelligence. This suggests it was measuring a trait distinct from the cognitive abilities assessed in IQ tests (see, e.g., Toplak, West, & Stanovich, 2011, 2014) and more closely aligned with the System 1-System 2 divide argued for by such authors as Stanovich and West (2008), which has resulted in consideration of traits separate from intelligence and personality often termed decision styles – like need for cognition and need for cognitive closure (see, respectively, Cacioppo & Petty, 1982; Webster & Kruglanski, 1994).

Other researchers, however, have pointed out that the measures of intelligence used in many decision making studies are relatively 'blunt' – self-reported college entry scores, for example – and thus that the relationship between intelligence and decision making may have been underestimated (see, e.g., Welsh, Burns, & Delfabbro, 2013). In line with this, a number of researchers have examined the relationships CRT has with cognitive ability. For example, Welsh et al (2013) found it to correlate strongly with quantitative ability, Weller et al (2013) concluded that it is a numeracy test and evidence is divided over whether it provides incremental predictive validity over cognitive ability measures (compare, e.g., Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2015; Sinayev & Peters, 2015).

Given the limitations of the original CRT in terms of consisting of only three items, all of which require arithmetic reasoning, researchers have also worked to expand the CRT with additional items and to lessen the role of mathematics in 'solving' the items (see, e.g., Primi et al., 2015; Thomson & Oppenheimer, 2016). These have not, however, resulted in a clear demarcation from cognitive ability, as shown by a recent meta-analysis (Otero et al., 2022) indicating that the balance of evidence still suggests CRT is a cognitive measure, closely related to quantitative ability as defined in the Cattell-Horn-Carrol hierarchical model of intelligence (McGrew, 2009).

A question remains, however, over the adequacy of some of the measures used in previous studies and also the extent to which other traits might explain any incremental predictive power that CRT possesses in different contexts. For example, while the intersection of CRT and intelligence has attracted more attention, there are aspects of personality that match onto the underlying idea of cognitive reflection – impulsivity, for example – and it has been argued that personality and decision styles account for the CRT's predictive power for real life decision making (Juanchich, Dewberry, Sirota, & Narendran, 2016). Similarly, attentional measures relating to the ability to inhibit incorrect responses (see, e.g., Warm, 1984) seem to share similar attributes to the proposed cognitive reflection trait.

Given this, the current study was designed to incorporate as wide a range of theoretically well-grounded cognitive abilities, personality traits, decision styles and attention measures as was feasible in order to establish how, exactly, the CRT relates to established individual differences.

Method

Participants

Participants were 301 university students, graduates and members of the general public recruited as part of a larger study; 172 identifying as female, 120 male and 9 non-binary. Participants listed 34 countries-of-origin but the majority (207) indicated English was their first language. Most were university educated, including 26 with post-graduate degrees, 101 with bachelor degrees (including 38 current post-grads) and 107 current university students. Ages ranged from 18 to 79 (M = 28.8, SD = 12.8).

Materials

Demographics
Demographics were gathered via an introductory online survey prior to online and in-person individual differences and experimental testing. These included: age; gender; native vs non-native English speaker; and level of educational attainment - recorded via options ranging from 'Did not complete High School' to 'Doctorate'.

Cognitive Reflection
Cognitive reflection was tested using a seven-item scale composed of the original three questions from Frederick's (2005) CRT and the four additional questions designed by Thomson and Oppenheimer's (2016) to have intuitively obvious, wrong answers but to not require any numerical calculation to find the correct answer in the way that the original questions do.

Decision Styles
Five decision style measures designed around concepts with the potential to overlap or influence cognitive reflection were included. These included the Rationality and Intuition scales from the Decision Styles Scale (Hamilton, Shih, & Mohammed, 2016) which are similar in conception to the Need for Cognition and Faith in Intuition scales (Cacioppo & Petty, 1982; Epstein, Pacini, Denes-Raj, & Heier, 1996), reflecting a person's desire to engage in deep thought and faith in their intuitive responses, respectively.

Need for Cognitive Closure (Webster & Kruglanski, 1994) and Actively Open-Minded Thinking (Haran, Ritov, & Mellers, 2013) were also included as a means of capturing people's tolerance for ambiguity and willingness to devote effort to maintaining an open mind.

Finally, the Brief Maximization Scale (Nenkov, Morrin, Schwartz, Ward, & Hulland, 2008) was included, which distinguishes between preferences for satisficing versus more cognitively taxing optimization.

Cognitive Abilities
Tasks corresponding to 8 broad abilities from the Cattell-Horn-Carroll model of intelligence (McGrew, 2009) were included. Six of these abilities were represented by 3 tasks each, allowing a factor score to be extracted using PAF with direct oblimin rotation in SPSS. In each case a single factor was returned using Kaiser's rule of eigenvalues>1 (1.74 being the lowest observed) and confirmed by examination of the Scree plot. The remaining two were represented only by single tasks due to experimental constraints. The specific tasks and key results of the factor analyses are shown in Table 1.

Table 1. Cognitive ability measures and factor analyses

Broad Ability Factors, Measures and Factor Loadings (FL)

Gf – Fluid Ability (KMO = 0.666), Bartlett's Test of Sphericity $\chi^2(2) = 218$, p<.001.
12-item Ravens APM (Arthur Jr & Day, 1994), FL = .853
CAB-I (Hakstian & Bennet, 1977), FL = .662

WJIV Number Series (this and all subsequent measures labelled WJIV are from the Woodcock-Johnson IV Tests of Cognitive Abilities, Schrank & Wendling, 2018), FL = .618.

Gc – Crystallized Ability (KMO = 0.668) , Bartlett's Test of Sphericity $\chi^2(2) = 243$, p<.001.
Mill-Hill Vocabulary Scale (Raven & Court, 1998), FL = .867
Spot-the-Word (Baddeley, Emslie, & Nimmo-Smith, 1993), FL = .699
WJIV Oral Vocabulary, FL = .612.

Gsm – Short Term Memory Ability (KMO = 0.669) , Bartlett's Test of Sphericity $\chi^2(2) = 173$, p<.001.
WJ-IV Numbers Reversed, FL = .610
Memory Span Forward, FL = .743
Memory Span Backwards. These tasks were written for this project in Matlab, displaying numbers of increasing length – from 1 to 10 – presented one digit at a time at 1 second intervals. After presentation, participants were asked to enter the digits in either the order presented or reversed order. Scores were the number of digits correctly recalled out of the total of 55 (10+9+8+... +1) from each task, FL = .780.

Glr – Long Term Retrieval Ability, Bartlett's Test of Sphericity $\chi^2(2) = 239$, p<.001.
WJ-IV Rapid Picture Naming, FL = .829
WJ-IV Retrieval Fluency, FL = .827
Comprehension task. The comprehension task was written in Matlab for this project and required participants to read two ~500 word passages about historical events and answer four multiple (4) choice questions after each. Scores were the number of questions correctly answered out of eight. FL = .437.

Gq – Quantitaive (Numerical) Ability (KMO = .642) , Bartlett's Test of Sphericity $\chi^2(2) = 113$, p<.001.
12-Item Numerical Aptitude Test (Welsh et al., 2013), FL = .684
Berlin Numeracy Test (Cokely, Galesic, Schulz, Ghazal, & Garcia-Retamero, 2012), FL = .519
Subjective Numeracy Test (Fagerlin et al., 2007), FL = .623.

Gt – Decision Speed Ability (KMO = .613) , Bartlett's Test of Sphericity $\chi^2(2) = 136$, p<.001.
Inspection Time (Preiss & Burns, 2012), FL = .448
Simple Reaction Time, FL = .830
Go-No Go Reaction Time. The RT tasks were written for this project in Matlab. SRT asked participants to press a key as soon as a red 'R' appeared onscreen for 10 trials. Go-No Go required responses only to the letter 'E', which occurred on 10 out of 100 trials (ten each of each letter from A to J, presented to participants in the same, randomized order. FL = .614.

Gs – Processing Speed Ability
WJIV Letter-Pattern

Gv – Visuo-Spatial Ability
WJIV Visualization

Note: The KMO statistics all exceed the standard 0.6 cut-off for adequacy of the data for factor analysis (0.5 using Kaiser's original threshold). This and Bartlett's Test of Sphericity being significant at <.001 suggests the data are suitable for factor analysis in all cases. That said, some explanation of the relatively low KMO statistics may be required with 0.6 characterized as 'mediocre' by Kaiser. These lower scores likely reflect that some of the measures being analysed, while belonging to the same broad ability, are drawn from different narrow abilities in the CHC model and would thus be expected to be somewhat less closely linked.

Personality Traits
The Big 5 personality traits - Neuroticism, Extraversion, Openness-to-Experience, Agreeableness and Conscient-iousness - were measured using the full version of the NEO-PI3 personality test (McCrae, Costa, & Martin, 2005). In addition to the five factor scores, this includes six facets per trait, yielding another 30 sub-measures. In the interest of space, these facets are discussed only where they show significant relationships or help to explain other results but were all included in the linear regressions described later.

Attention Measures
Six tasks were included in the in-person tasks measuring the five aspects of attention outlined by Sohlberg and Mateer (1989) - focused attention, sustained attention, selective attention, alternating attention and divided attention. These differ in terms of whether a person is simply waiting for a stimulus (focused), maintaining vigilance and responding repeatedly across an extended task (sustained), having to ignore distractors (selective), switching back and forth between two tasks (alternating) or require two tasks to be done simultaneously (divided). In addition to these, people's ability to inhibit automated responses was measured using a variant of the Sustained Attention to Response Task (SART; Warm, 1984).

These tasks were coded in Matlab and response times and accuracy were recorded under these different conditions. As a result of the reliance on reaction/response times these measures overlap the decision speed measures described above to some extent, but additionally incorporate errors of omission and commission and changes in RT within and across conditions and tasks. In all, twelve measures reflecting speed and accuracy across these six tasks were included.

The Focused tasks yielded a simple reaction time (RT) and the number of false starts (Errors) while the Sustained task yielded a go-no-go RT and Errors (of omission or commission). Selective attention was assessed using a flanker task with responses on trials with congruent and incongruent flanking stimuli compared in terms of RT and Errors. Alternating attention was measured comparing a numerical

with an alpha-numerical trail-making test (Reitan, 1958) and recording the difference in time (ΔT) taken to complete. The divided attention task measured RT and Errors (omission and commission) from a task requiring responses to only Odd Blue or Even Red numbers from a series of numbers presented in four colours. Finally, Inhibition was measured using a variant of the SART, requiring participants to respond to all numbers presented except for 3s, yielding separate RTs for correct and incorrect responses and an Error count.

Procedure
The measures described above were part of a larger study on biases. Participants were recruited online and on campus but needed to be available to come to campus for testing sessions. The testing was conducted in four parts and participants were paid $100 on completion. The first was an online survey that combined participant information, consent and demographic data collection. Four-hundred and four participants complete this initial survey.

Following this, participants were sent invitations to a second, ~1hr survey including the personality and decision style questions; a subjective attention scale; the subjective numeracy scale; and the spot-the-word test.

On completion, they were invited to a third survey, including: the CRT; the cognitive abilities measures not conducted in person (see below); and a variety of decision bias tasks not discussed herein, taking around 2 hours.

Finally, participants attended a 2 hour, in-person session, to complete the remaining cognitive ability measures including: the WJ-IV tasks; the NAT; the comprehension task; the memory span tasks; and a series of computerised tasks yielding the inspection time, reaction time and attention measures. Additional bias tasks (again, not discussed herein) were included at the end of this session. Overall, 301 of the 404 participants who signed up online completed all testing.

Results

CRT

Cognitive reflection tasks can be scored in distinct ways. The original being to simply count the number of correct questions and a more recent suggestion (see, e.g., Pennycook, Cheyne, Koehler, & Fugelsang, 2016) being to distinguish between responses that are, simply, incorrect and those that are incorrect and also match the expected, intuitive response. This can be maintained as a separate variable or combined with the normal CRT score by scoring incorrect answers as 0 but intuitive incorrect answers as -1. Given the dependencies between these, all three possible scores - CRT, CRT intuitive (CRTi) and CRT combined (CRTc) - will tend to correlate highly together but are maintained separately herein to allow for potential distinctions in terms of their other relationships.

Table 2 summarises the descriptive statistics and intercorrelations between these three variant CRT scores. Note that, here and throughout, while both Pearson and Spearman correlations were calculated, the Pearson correlations are reported due to their greater ease of

interpretation as linear relationships – consistent with assumptions in the factor analysis and linear regression analyses described later. This is done even where the variables do not strictly meet the criteria for a Pearson correlation - such as being ordinal rather than continuous or skewed rather than normally distributed - as Pearson correlations are robust to these violations (Havlicek & Peterson, 1976) and the pattern of results remained consistent regardless of which type of correlations were examined.

Table 2. Alternative CRT score comparisons

	M (SD)	Range	1	2	3
1. CRT	4.1 (2.0)	0-7	1	<.001	<.001
2. CRTi	1.9 (3.7)	0-7	.98	1	<.001
3. CRTc	2.2 (1.7)	-7-7	.90	.97	1

Note: N=300. CRTi – intuitive; CRTc – combined. (2-tailed.)

Demographic Variables

Pearson correlations conducted on CRT scores and demographic variables indicated that these did not differ significantly with participant's age or level of educational attainment. Similarly, while gender differences were observed in the direction expected from the literature (i.e., males scoring somewhat better on CRT than females) the differences were small (mean differences of 0.23, .07 and 0.31) and not statistically significant, $t(289) = .958, .360$ and $.702, p = .339, .719$ and $.483$ for the CRT, CRTi and CRTc respectively.

Decision Styles

Pearson correlations between the CRT scores and the five decision styles variables were calculated and are shown in Table 3.

Table 3. CRT scores' correlations with decision styles

	CRT	CRTi	CRTc
DS_Rational	*.121*	-.111	*.119*
DS_Intuitive	*-.122*	.098	*-.114*
NFCC	-.051	.005	-.031
AOT	**.267***	*-.217*	**.251**
Brief_Max	-.090	.106	-.100

Note: N=300. *Italic*, **Bold** & **Bold*** = sig. at .05, .01 & .001 level (2-tailed), respectively.

Looking at Table 3, one sees CRT scores are related to some of the decision styles (DS Rational, DS Intuitive and Actively Open-minded Thinking) in directions one might expect - with greater reflection positively relating to rationality and actively open-minded thinking and negatively with intuition. By contrast, CRT seems distinct from the Brief Maximization Scale and Need for Cognitive Closure. In all cases, the strongest significant relationships are seen with the basic CRT measure and the weakest with the CRT intuitive measure.

Attention

Pearson correlations between the CRT scores and the twelve attention measures are shown in Table 4.

Table 4. CRT scores' correlations with attention measures

	CRT	CRTi	CRTc
Focussed RT	**-.193***	**-.202***	**.202***
Focussed Errors	-.077	.053	-.067
Sustained RT	**-.265***	**.256***	**-.268***
Sustained Errors	*-.116*	.062	-.093
Selective ΔRT	.022	-.006	.016
Selective ΔCorrect	*-.127*	.112	*-.123*
Alternating ΔT	*-.130*	*.138*	*-.137*
Divided RT	*-.125*	.109	*-.121*
Divided Errors	**-.258***	**.207***	**-.241***
Inhibition RT	.021	.034	-.004
Inhibition RT Errors	*-.143*	.097	*-.125*
Inhibition Errors	-.063	.013	-.041

Note: N=258-300. Note: *Italic*, **Bold** & **Bold*** = sig. at .05, .01 & .001 level (2-tailed), respectively.

The data in Table 4 show a number of relatively weak relationships between attention and the CRT and have a similar pattern to Table 3, with the basic CRT score tending to have higher correlations with the other measures than the CRTi. The strongest relationships with CRT are observed for the response time measures from the focused and sustained attention tasks and the number of errors observed in the Divided attention task. Overall, the pattern of results, while weak, suggests that better CRT scores tend to predict faster response times and fewer errors, which only partly aligns with the idea of greater reflection underlying performance.

Personality

Pearson correlations between the Big 5 traits and the CRT scores were calculated and are shown in Table 5.

Table 5. CRT scores' correlations with Big 5 measures

	CRT	CRTi	CRTc
Neuroticism	-.008	-.026	.008
Extraversion	-.058	.047	-.055
Openness	**.204***	**-.172**	**.194***
Agreeableness	.081	-.072	.079
Conscientiousness	-.034	-.019	-.010

Note: N=300. **Bold** & **Bold*** = sig. at .01 & .001 level (2-tailed), respectively.

Table 5 reveals only one personality factor linked to the CRT scores – openness-to-experience. An examination at the facet level revealed that three of the six facets of openness – Fantasy, Ideas and Values – correlated significantly with the three CRT measures. Of these, Values showed the strongest relationships at .297, -.249 and .283 with CRT, CRTi and CRTc, respectively (p<.001 in all cases) with Ideas lower but still above .2 (p<.001) while Fantasy's highest correlation was .134 with CRT (p<.05).

This examination also indicated that the Dutifulness facet of Conscientiousness correlated significantly (p<.05) with the three CRT scores at .116, -.130 and .125, respectively. However, while this seems a reasonable result – more dutiful people showing greater reflection on their answers - given the large number of comparisons made, any correction of the family-wise alpha would leave this being non-significant.

Intelligence

Pearson correlations were calculated between the CRT scores and the eight measures (six factors and two lone measures) reflecting broad abilities from the Cattell-Horn-Carroll model of intelligence and are shown Table 6.

Table 5. CRT scores' correlations with CHC broad abilities

	CRT	CRTi	CRTc
Gf - Fluid	.626	-.521	.593
Gc – Crystallized	.405	-.337	.384
Gsm – STM	.251	-.233	.249
Glr – LT Retrieval	.280	-.265	.280
Gq - Quantitative	.473	-.387	.445
Gt – Decision Speed	.258	-.260	.266
Gs – Processing Speed	.304	-.288	.304
Gv – Visuo-Spatial	.434	-.349	.406

Note: N=296-300. ***All*** correlations significant at the .001 level (2-tailed)

The strongest correlation in Table is above 0.6 and all are significant at the .001 level, suggesting the broad cognitive abilities are more closely related to the CRT than the previously considered measures. In fact, these results likely explain some of the previous findings. Specifically, the response time measures from the attention tasks are closely related to the components of the decision time ability while the predictive power of openness to experience is likely explained by the well-established relationship between openness and intelligence.

Looking in finer detail, the best predictors of CRT scores are a person's fluid ability and their quantitative ability – that is, their ability to understand and solve novel problems and to work with numbers. This holds true regardless of which CRT score is considered but, once again, higher correlations are seen using the original scoring rather than the variants.

Linear Regression

Given the large number of relationships the CRT scores have with the variables above and the relationships those variables have with one another, linear regressions using the Forward method were conducted in SPSS ,using all of the decision style, attention, ability measures and the 30 personality facets in order to identify variables with distinct explanatory power in terms of predicting a person's CRT score and, conversely, those that are redundant.

For the three CRT scores, similar results were obtained, with Gf, Gc and Gq being the first three variables added to the model in all cases, followed by two attention measures -

the Response Time from the Inhibition task's Error trials and the difference in Response Time between congruent and non-congruent trials in the Selective task. The regression for the CRTc score added the Modesty facet of Agreeableness to the model, while the original CRT regression model also included the Aesthetics facet from Openness – neither of which were strongly related to CRT in the initial correlations. Table 7 summarises the final models for each of the CRT scores.

Tables 7. Linear regression models for CRT, CRTi and CRTc

	Predictors	Adj. R^2	ANOVA
CRTi	Gf Gc Gq Inhibition RT Err Selective ΔCorr	.309	$F_{(5,220)}$ = 21.2, p <.001
CRTc	+Modesty	.422	$F_{(6,219)}$ = 28.4 p<.001
CRT	+Aesthetics	.485	$F_{(7,218)}$ = 31.3 p<.001

Note: using just the five predictors from the CRTi model the Adjusted R^2 values for CRTc and CRT would be .414 and .468 respectively.

The regression models indicate that the original CRT score is better explained by the variables included herein than are the alternate scoring options. Neither are the alternatives predicted by any variables not found in the regression model for the original CRT.

Discussion

The above results clearly place CRT within the constellation of human individual differences, using modern, gold-standard psychometric theories and a wider variety of traits than can commonly be addressed within a single study, enabling its relationships with these existing measures to be examined in detail. The results are discussed below, expanding on the relationships seen with particular types of individual differences and the implications of these for understanding the CRT.

CRT Scoring

Before continuing to the main discussion of results, however, it is worth taking a moment to note the lack of results resulting from the alternative scoring methods for the CRT. Overall, it seems that these methods do not alter CRT results in any distinct way that that is captured by any of the 27 primary variables (i.e., excluding the 30 facets) included here. That is, while scoring the intuitive responses separately might seem like it will capture aspects of behaviour relating to intuition more strongly, there is no evidence in the data that this is the case.

The alternative scoring methods also tended to reduce the strength of relationships between the CRT and other measures, which implies that they are worse at capturing the essence of the CRT. Given this, retaining the original scoring method seems sensible.

Intelligence

The results of this study largely support the growing body of evidence (see, e.g., Otero et al., 2022) that CRT is, primarily, a measure of intelligence or cognitive ability. Its highest correlations were with Gf (.626), Gq (.473), Gv (.434) and Gc (.405) and, despite their inter-correlations, three of these emerged from the regression as the largest individual contributors to the variance in CRT.

Specifically: fluid ability (Gf), which measures a person's ability to reason and solve novel problems; crystallized ability (Gc), which measures a person's acquired knowledge; and quantitative ability (Gq), which measures their ability to undertake numerical tasks. (NB: the reversal of the relative importance of Gc and Gq form the correlations and the absence of Gv, visuo-spatial, in the regression is the result of Gc being less closely related to Gf than these two abilities and thus retaining more predictive power.)

Attention

The above conclusion is complicated slightly, however, by the additional measures emerging from the regression analyses. Specifically, two of the attention measures remain significant predictors of CRT after accounting for the effect of the cognitive abilities noted above. The first is the Response Time measure from incorrect trials on the Inhibition task, which is only weakly related to more direct measures of reaction time and decision speed (Gt) as it measures the speed at which a person responds when making a mistake. Interestingly, the direction of the relationship suggests people who are faster when making errors in the attention task do better on the CRT. This could be interpreted as indicating people who only tend to make mistakes when rushing – as becomes common during the extended Inhibition task.

The second attention measure indicated by the regression is the difference in the number of correct responses made between congruent and incongruent trials on the Selective attention task. People with smaller differences between these values are those who are better able to ignore distractors within the task and tended to do better on the CRT.

Both of these aspects of attention do seem relevant to the concept of cognitive reflection as this is meant to entail an increased likelihood to notice and inhibit incorrect responses. Their effect on the variance explained in the regression model is, however, modest – increasing the Adjusted R^2 by .017 and .016, respectively.

Decision Styles and Personality

The decision styles and personality measures seem to have the least relationship with CRT and, more importantly, their unique contributions to its variance are very limited. This seems to be because these measures are largely accounted for by combinations of other variables.

For example, the relationship between openness-to-experience and the CRT can be accounted for entirely by their shared relationships with Gc and disappears if Gc is controlled for. Similarly, the relationships observed between CRT and the various decision styles measures can be accounted for by the relationships these measures have with various of the cognitive ability measures and/or openness-to-experience. For example, actively open-minded thinking, which had one of the stronger relationships with CRT in the data, does not emerge in the regression because AOT is quite strongly related to openness-to-experience, a number of the intelligence measures and agreeableness.

One result in need of explanation, however, is the regression suggesting that the NEO-PI3 facets of Aesthetics (openness-to-experience) and Modesty (agreeableness) are significant (if small) contributors to CRT variance, increasing Adjusted R^2 by .017 together. An examination of these measures showed them to correlate weakly with various of the decision styles measures and to retain some of these significant relationships even if Gf and Gc were controlled for. Specifically, Modesty – which includes the willingness to accept that you might be wrong – is negatively related to Need for Cognition Closure, the DSS intuitive scale and the Brief Maximization Scale, suggesting that a willingness to *not* quickly jump to a strongly held conclusion may be small a part of the CRT.

The Aesthetics facet, by contrast, correlates positively with actively open-minded thinking and negatively with NFCC but further interpretation is difficult as the underlying concept of Aesthetic or artistic appreciation bears no obvious relationship to CRT. In both cases, it is possible these relationships reflect an artefact of regression – with noise in the less impactful variables determining which covariates are retained and which excluded. That is, these measures could be reflecting the left-over effects of the various decision style measures once intelligence and personality are removed.

Caveats

As noted above, interpreting a regression run on so many inter-correlated variables, while useful for determining which overlap one another and which retain separate predictive power, does allow for noise in the variables to affect which, specific variables emerge as significant. While the effect of the intelligence measure, Gf, Gc and Gq is quite clear, interpretations regarding the variables with smaller impacts must be regarded cautiously.

There are also questions of range truncation and how this affects correlations to be kept in mind. This is often a concern with intelligence measures using university samples but an analysis of the seven WJ-IV measures used herein suggests that the current sample is quite representative of the general population, with IQ conversion scores (on the individual tests) ranging from 55 to 150, with means between 99 and 109 and SDs between 12 and 16 (cf – the expected means of 100 and SDs of 15).

Other measures, however, do display range truncation that may be affecting results. For example, scores on the 11-item AOT scale used herein theoretically range from 11 to 55. The data, however, have a mean AOT score of 41.8 and the lowest score was 29. This suggests that lower scorers on AOT may not have been sampled – perhaps reflecting bias due to more open-minded participants being more likely to engage in a scientific study. Regardless of the reason, this will tend to reduce the strength of any correlations calculated for AOT.

Conclusions

The overall conclusion, given the results and discussion presented above, is that the CRT is mainly a measure of intelligence – specifically, it seems most closely related to fluid intelligence in this data but also, separately, to crystallized and quantitative ability.

The regression results do, however, suggest that there are other contributors to CRT. For example, it seems there are aspects of people's attention, specifically their ability to notice and inhibit incorrect responses, that contribute a small amount of variance. Additionally, there may be a little variance relating to the decision styles measures being captured by personality facets – including a person's Modesty or willingness to accept the possibility of being wrong. However, the contributions of these non-cognitive variables, while statistically significant, are very small.

References

Arthur Jr, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological measurement, 54*(2), 394-403.

Baddeley, A., Emslie, H., & Nimmo-Smith, I. (1993). The Spot-the-Word test: A robust estimate of verbal intelligence based on lexical decision. *British Journal of Clinical Psychology, 32*(1), 55-65.

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*(1), 116-131.

Cokely, E. T., Galesic, M., Schulz, E., Ghazal, S., & Garcia-Retamero, R. (2012). Measuring risk literacy: The Berlin numeracy test. *Judgment and Decision Making, 7*(1), 25.

Epstein, S., Pacini, R., Denes-Raj, V., & Heier, H. (1996). Individual differences in intuitive–experiential and analytical–rational thinking styles. *Journal of Personality and Social Psychology, 71*(2), 390.

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*(5), 672-680.

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

Hakstian, A. R., & Bennet, R. W. (1977). Validity Studies Using the Comprehensive Ability Battery (CAB): 1. Academic Achievement Criteria. *Educational and Psychological measurement, 37*(2), 425-437.

Hamilton, K., Shih, S.-I., & Mohammed, S. (2016). The development and validation of the rational and intuitive decision styles scale. *Journal of Personality Assessment, 98*(5), 523-535.

Haran, U., Ritov, I., & Mellers, B. A. (2013). *The role of actively open-minded thinking in information acquisition, accuracy, and calibration*. ScholarlyCommons, (5-2013).

Havlicek, L. L. and N. L. Peterson (1976). Robustness of the Pearson correlation against violations of assumptions. *Perceptual and Motor Skills, 43* (3_suppl): 1319-1334.

Juanchich, M., Dewberry, C., Sirota, M., & Narendran, S. (2016). Cognitive reflection predicts real-life decision outcomes, but not over and above personality and decision-making styles. *Journal of Behavioral Decision Making, 29*(1), 52-59.

McCrae, R. R., Costa, J., Paul T, & Martin, T. A. (2005). The NEO–PI–3: A more readable revised NEO personality inventory. *Journal of Personality Assessment, 84*(3), 261-270.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*(1), 1-10. doi:https://doi.org/10.1016/j.intell.2008.08.004

Nenkov, G. Y., Morrin, M., Schwartz, B., Ward, A., & Hulland, J. (2008). A short form of the Maximization Scale: Factor structure, reliability and validity studies. *Judgment and Decision Making, 3*(5), 371-388.

Otero, I., Salgado, J. F., & Moscoso, S. (2022). Cognitive reflection, cognitive intelligence, and cognitive abilities: A meta-analysis. *Intelligence, 90*, 101614.

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.

Preiss, A. K., & Burns, N. R. (2012). Accurately measuring inspection time with computers. *International Journal of Intelligence Science, 2*(4). doi:DOI:10.4236/ijis.2012.24013

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*.

Raven, J. C., & Court, J. H. (1998). *Raven's progressive matrices and vocabulary scales* (Vol. 759): Oxford pyschologists Press Oxford, England.

Reitan, R. M. (1958). Validity of the Trail Making Test as an indicator of organic brain damage. *Perceptual and Motor Skills, 8*(3), 271-276.

Schrank, F. A., & Wendling, B. J. (2018). *Woodcock–Johnson IV Tests of Cognitive Abilities Examiner's Manual, Standard and Extended Batteries.* Itasca, IL.: Riverside Publishing.

Sinayev, A., & Peters, E. (2015). Cognitive reflection vs. calculation in decision making. *Frontiers in psychology, 6*, 532.

Sohlberg, M. M., & Mateer, C. A. (1989). *Introduction to cognitive rehabilitation: Theory and practice*. New York, NY: Guilford Press.

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive ability. *Journal of Personality and Social Psychology, 94*(4), 672-695.

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

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*(7), 1275-1289.

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.

Warm, J. S. (1984). *Sustained attention in human performance*. New York, NY: Wiley.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*(6), 1049.

Weller, J. A., Dieckmann, N. F., Tusler, M., Mertz, C. K., Burns, W. J., & Peters, E. (2013). Development and Testing of an Abbreviated Numeracy Scale: A Rasch Analysis Approach. *Journal of Behavioral Decision Making, 26*, 198-212. doi:doi: 10.1002/bdm.1751

Welsh, M. B., Burns, N., & Delfabbro, P. (2013). *The Cognitive Reflection Test: how much more than numerical ability.* Paper presented at the Proceedings of the 35th Annual Conference of the Cognitive Science Society.