HHS Public Access
Author manuscript
Exp Aging Res. Author manuscript; available in PMC 2024 June 05.

Published in final edited form as:
Exp Aging Res. 2018 ; 44(1): 18–34. doi:10.1080/0361073X.2017.1398508.

Does the Cognitive Reflection Test Actually Capture Heuristic Versus Analytic Reasoning Styles in Older Adults?

Christopher Hertzog,
R. Marit Smith,
Robert Ariel
Georgia Institute of Technology

Abstract

Background: We evaluated adult age differences in the original three-item Cognitive Reflection Test (CRT; Frederick, 2005) and an expanded seven-item version of that test (Toplak et al., 2013). The CRT is a numerical problem solving test thought to capture a disposition towards either rapid, intuition-based problem solving (Type I reasoning) or a more thoughtful, analytical problem solving approach (Type II reasoning). Test items are designed to induce heuristically-guided errors that can be avoided if using an appropriate numerical representation of the test problems.

Methods: We evaluated differences between young adults and old adults in CRT performance and correlates of CRT performance. Older adults (ages 60 to 80) were paid volunteers who participated in experiments assessing age differences in self-regulated learning. Young adults (ages 17 to 35) were students participating for pay as part of a project assessing measures of critical thinking skills or as a young comparison group in the self-regulated learning study.

Results: There were age differences in the number of CRT correct responses in two independent samples. Results with the original 3-item CRT found older adults to have a greater relative proportion of errors based on providing the intuitive lure. However, younger adults actually had a greater proportion of intuitive errors on the long version of the CRT, relative to older adults. Item analysis indicated a much lower internal consistency of CRT items for older adults.

Conclusion: These outcomes do not offer full support for the argument that older adults are higher in the use of a “type I” cognitive style. The evidence was also consistent with an alternative hypothesis that age differences were due to lower levels of numeracy in the older samples. Alternative process-oriented evaluations of how older adults solve CRT items will probably be needed to determine conditions under which older adults manifest an increase in the Type I dispositional tendency to opt for superficial, heuristically-guided problem representations in numerical problem solving tasks.

Keywords

Cognitive Reflection Test; Type I reasoning; heuristics; quantitative reasoning; numeracy

---

Address correspondence to: Christopher Hertzog, School of Psychology, 654 Cherry St. NW, Georgia Institute of Technology, Atlanta, GA 30332-0170.

Traditionally, studies of reasoning and problem solving in older adults have focused on age differences in inductive and deductive reasoning abilities through adulthood into old age (e.g., Salthouse, 1992; Schaie, 2013). Older adults are known to perform more poorly than younger adults on virtually all reasoning tests and tasks, including inductive reasoning (e.g., Raven’s Progressive Matrices test performance; Der et al., 2010; Heron & Chown, 1967) and relational reasoning (e.g., Fisk & Sharpe, 2002; Viskontas et al., 2004). These age differences are often attributed to age-related declines in central nervous system function and/or age-related changes in working memory capacity (e.g., Salthouse et al., 1989; but see Fisk, 2005). It is also the case that tests of inductive reasoning show substantial generational differences in cohorts born during the 20th century (Pietschnig & Voracek, 2015; Schaie, 2013).

The use of routines or well-learned behavior patterns (what Baltes [1997] termed ‘cognitive pragmatics’) supports adults’ everyday cognition (e.g., Park et al., 1999) and can create a discrepancy between laboratory versus everyday tests of reasoning and cognition, especially when the latter are embedded in a natural ecology (e.g., Schnitzspahn et al., 2011). Experience with a class of problems or problem contexts fosters high-level everyday problem solving (Strough & Keener, 2014). Furthermore, older adults may even perform better on wisdom-related thought (e.g.., Baltes & Smith, 2008; Staudinger & Glück, 2011) and nominate superior strategies for dealing with everyday problem solving in social contexts (e.g., Blanchard-Fields, 2009; Strough & Keener, 2014). In many of these situations knowledge-based candidate solutions may fluently come to mind for older adults, without a need for formal problem analysis. Experts or others with higher levels of relevant knowledge of a problem domain can often achieve high-quality decisions in complex everyday tasks like financial planning with less effort by honing in on the information critical for problem solving (e.g., Hershey et al., 1990; see Hess & Queen, 2014).

Gigerenzer’s work on adaptive heuristics and their role in cognition and decision making (e.g., Gigerenzer, Todd, & the ABC Research Group, 1999) argues that experience-based decision-making heuristics are normatively effective and highly attractive because they provide a rapid basis for accurate decision making based on minimal need for information search or analytic evidence evaluation. Mata and colleagues have argued that older adults fare well in situations where this kind of adaptive use of heuristics is possible (e.g., Pachur et al., 2009).

There is some evidence that older adults are more reliant on a cognitive style that leads to an emphasis on quick resolution of decisions, being higher in self-rated need for closure and need for structure, including the disposition to seek quick answers and to minimize information search durations (Blanchard-Fields et al., 2001; Hess, 2001; see Peters et al., 2007). Older adults also score higher on Epstein’s et al. (1996) experiential (versus analytical) decision style measure (Horhota, Mienaltowski, & Blanchard-Fields, 2012), indicating a tendency to respond on the basis of experience and intuition rather than effortful and time-consuming analytic thinking. A potential liability of relying on cognitive pragmatics in everyday life is that older adults may be less motivated or otherwise less likely to engage in formal analytic reasoning in those instances when it is needed to arrive at correct solutions. Overreliance on heuristic thought may be especially likely if recruitment of cognitive resources like working memory for formal analysis is experienced as highly taxing or demanding (Hess, 2014). Older adults manifest heuristic biases considered maladaptive in certain reasoning and decision-making contexts (e.g., Finucane & Gullion, 2010), although Bruine de Bruin, Parker, and Fischoff (2012) argued that such age differences are primarily found in decision making tasks that are strongly influenced by fluid intelligence and working memory (see also del Missier et al., 2013). A recent cluster-analytic study actually found older adults to be less likely to belong to an experiential decision style cluster defined by rapid decisions based on heuristics and more likely to belong to a cluster defined by a controlled, independent decision making style (Delaney, Strough, Parker, & Bruine de Bruin, 2015). One cannot claim based on existing studies that older adults are universally more likely to think experientially instead of analytically in decision making contexts.

Some cognitive psychologists regard a tendency towards making rapid decisions based on a quick (and perhaps superficial) analysis of the problem space as a cognitive or intellectual style (Kozhevnikov, Evans, & Kosslyn, 2014; Zhang & Sternberg, 2006), perhaps related to the personality construct of impulsivity/reflectivity (e.g., Dickman & Meyer, 1988). This tendency has been characterized in a number of ways (Evans, 2009; Kahneman, 2011; Stanovich, 2011), including Epstein’s experiential versus analytic distinction already mentioned and, more recently, the Type I (heuristic) versus Type II (formal or algorithmic reasoning) distinction, (e.g. Evans & Over, 1996; Evans & Stanovich, 2013; Stanovich, 2009, 2011). In this view, Type II reasoning is thought to be the outcome of a conscious choice to reflect and analyze, thereby inhibiting or otherwise overriding a tendency to accept the first solution that comes to mind (Frederick, 2005; Pennycook et al., 2015; Toplak et al., 2011). Kahneman (2011) argued that intuitive, non-analytic thought is driven by rapid automatic retrieval of experience-based associations of situations and likely explanations for them. This view stands in sharp contrast to the perspective that heuristics are often adaptive and generically useful in everyday life. Stanovich (2011) and Kahneman (2011) emphasized that reliance on heuristics over formal analysis can be problematic in many problem solving contexts. Stanovich (2011) in particular has criticized the claim that using heuristics in decision-making contexts is normatively adaptive. Given that the present study is not focused on rationality in the actual ecology, but rather, focuses on a cognitive test designed to lure people into making heuristic errors, this debate about ecological rationality is probably of little consequence for interpreting the present results. Our focus is instead on a particular measure that has been argued to capture heuristic versus analytic decision making and its validity for assessing this tendency in older adults.

To assess the stylistic dimension of Type I versus Type II reasoning, Frederick (2005) created a short 3-item instrument, the Cognitive Refection Test (CRT). Couched as a quantitative reasoning test, it presents problems that are structured so that a quick and superficial analysis of the problem suggests an incorrect solution. Consider the CRT’s first (and most widely cited) item: “*A bat and a ball cost $1.10 in total. The bat costs a dollar more than that ball. How much does the ball cost?”* If formally represented as an elementary two-unknowns in two-equations problem (i.e., using algebra to solve the simultaneous equations: $x + y = 1.10$; $x - y = 1.0$), the correct answer is easily derived as 5 cents. But respondents are lured by the item format to give a response of 10 cents, the observable difference in the provided amounts. Indeed, well-educated and numerically proficient people often mistakenly provide the 10-cent answer; university students choose the heuristic lure a surprisingly large proportion of the time (Kahneman, 2011). Frederick (2005) argued that doing so indicated prima facie evidence for a Type I approach to the problem, an argument that has been supported by additional studies (e.g., Alter et al., 2007; Campitelli & Gerans, 2013; Pennycook et al., 2015; Toplak, West, & Stanovich, 2011) but one that is not without criticism (see below). Those favoring a Type I – Type II interpretation of CRT test scores also point to patterns of CRT correlations with measures of heuristic decision making and other cognitive style measures, such as the Need for Cognition scale (e.g., Juancich et al., 2016). Toplak et al (2013) later created a longer 7-item version of the CRT to increase the sensitivity of the test (see Primi et al., 2015 for an alternative extension).

On its surface, the CRT is a face-valid measure of mathematical or quantitative reasoning. Carroll (1993), in his taxonomy of human abilities, argued that although quantitative reasoning was distinguishable from other primary abilities like inductive reasoning, it did triangulate on a second-order fluid intelligence factor (see also Horn, 1985; McArdle, Ferrer-Caja, Hamagami, & Woodcock, 2002). Correct answers on the CRT correlate substantially with traditional measures of working memory and fluid intelligence (e.g., Stupple et al. 2013; see the extended treatment of this issue by Stanovich, 2011). Thus poor performance on the test may indicate low quantitative reasoning aptitude rather than a tendency to rely on heuristic or Type I thinking; that is, an individual attempting to reason analytically may lack the algorithmic intelligence (Stanovich, 2011) in the quantitative reasoning domain to correctly represent and solve the problem. This issue is particularly germane to the issue of evaluating age differences in CRT performance. Older adults are known to perform more poorly than younger adults in tests of quantitative reasoning (Jokela et al., 2010; Salthouse, 1998), and quantitative reasoning has been shown to decline longitudinally over the adult life span, in tandem with other indicators of fluid intelligence (McArdle et al., 2002).

The distinction between intuitive Type I responding and low quantitative reasoning also matters because one of the principal reasons that quantitative reasoning probably diverges from other types of fluid intelligence is that representing problems in mathematical form requires a broad set of skills and knowledge subsumed under the construct of numeracy. Numeracy involves knowledge of simple arithmetic operations and also the ability to represent verbally-presented quantitative problems in mathematical form. Men and women tend to score similar in tasks measuring basic numeracy (see Miller & Halpern, 2014) but there are sex differences favoring males in quantitative reasoning tests such as the Graduate Record Examination and some intellectual ability batteries (Bleske-Rechek & Browne, 2014; Keith, Reynolds, Patel, & Ridley, 2008). There is some evidence that men are better than women in using visuospatial representations of quantitative reasoning test items presented in a verbal format (e.g., word algebra problems; Boonen et al., 2014), which may help to explain these sex differences.

Given these distinctions, errors in responding to CRT items do not necessarily reflect the use of a heuristic problem solving style, but can also arise from a failure in numerical representation and computation as well as deficient reasoning (e.g. Låg et al., 2013; Liberali et al., 2011; Sinayev & Peters, 2015). Some recent research supports the argument that both intuitive and numerical errors occur during the CRT. These two variables are ipsatively linked (given the fixed number of total items) and hence must be negatively correlated due to statistical artifact (Pennycook et al., 2015). Pennycook et al. measured intuitive reasoning in the CRT by the proportion of total CRT errors that selected the targeted intuitive option, an approach used in this study as well.

The distinction of different causes for correct and incorrect CRT responses is especially relevant to conceptualizing adult age differences in CRT performance. Older adults perform more poorly on reasoning tests in general, and in quantitative reasoning tasks in specific, as already noted. Therefore, poor CRT performance by older adults without explicit evidence of an increased rate of heuristic CRT errors should not be taken as evidence of a propensity toward intuitive responding. To date there have been only a few studies that have evaluated the CRT in older adults, and they have typically assumed it to be a valid measure of cognitive rationality from a Type I versus Type II perspective, using it as a covariate of rational decision making rather than as a focus in its own right (e.g., Besedes et al., 2012; Cokely & Kelley, 2009; Finucane & Gullion, 2010). To our knowledge no study has evaluated age differences in individual item responses on the CRT or evaluated the question of whether the test behaves similarly in older and younger adults (see Primi et al., 2015, for a recent item analysis of CRT-items in younger adults). The assumption that the CRT measures heuristic biases in similar ways across the life span is risky, because there is ample evidence that some cognitive tests lack age-related measurement equivalence (Labouvie, 1980; e.g., Hertzog, 1989). This is especially likely to be true in a test designed to capture an embedded aspect of cognitive style while simultaneously assessing a cognitive ability like quantitative reasoning that is vulnerable to age-related cognitive decline.

The relationship between cognitive reflection as measured by the CRT and judgment behavior has been found to be, for the most part, weakly related to or even independent of age (Cokely, Parpart, & Schooler, 2012). Age has also been reported to be weakly but negatively related to CRT scores (e.g., Finucane & Gullion, 2010). A modest negative correlation of age and CRT performance seems surprising given (a) widely replicated findings of robust age differences in formal reasoning tests and (b) the relationship between general fluid intelligence and the quality of reflective (Type II) reasoning (Stanovich, 2011). Small age-related effect sizes for the CRT could support the inference that performance on the test is predominantly a question of cognitive style, not reasoning ability. For instance, strong Type I tendencies in younger adults could reduce the magnitude of observed age differences relative to standardized tests of cognitive reasoning. However, age differences in numeracy favoring older adults could in principle limit the magnitude of age differences in the CRT, aside from any bias to engage in intuitive problem solving skills. Older adults are reported to perform well on some tests of numerical ability, often considered an aspect of crystallized intelligence (e.g., Schaie, 2013). However, the evidence for this finding is based on tests that assess low-level arithmetic skills (e.g., multiple-column addition) not higher-level aspects of numeracy such as the ability to solve algebraic equations. There is even evidence of cohort differences favoring old adults on elementary arithmetic problems, possibly reflecting negative transfer from more recently born cohorts' higher reliance on technology over mental arithmetic (Schaie, 2013). Cohort differences could also reduce age differences in CRT performance. On the other hand, Bruine de Bruin et al. (2015) reported age differences in performance on numeracy tests (including word algebra problems) that correlated with self-reported need for cognition, suggesting that at least some older adults are not motivated to maintain or apply numeracy skills in everyday life. However, their observed correlation of age and numeracy was rather small ($r = -.15$).

This study compared older and younger adults on the 3-item (Study 1) and 7-item (Study 2) versions of the CRT. We used an extreme age groups design, comparing university students with reasonably well educated older adults. Our younger adults were students at a technical university, so they were all highly capable of solving verbal mathematics problems of the difficulty level captured by the CRT. However, they were likely, given the literature, to be susceptible to Type I reasoning biases. We evaluated how their CRT scores compared to that of older adults.

Although the 7-item CRT should increase the sensitivity of the test because the larger number of items creates more opportunities to vary. However, it is an open question whether the new items are as useful as the originals in capturing intuitive errors, and it is currently unknown how older adults respond to the new items. We considered the 7-item version particularly important as a vehicle for evaluating the proportion of different types of errors in older adults, so as to test the argument that older adults are more likely to be lured by the intuitive response options. Furthermore, the expanded 7-item CRT provides a better basis for evaluating whether the CRT has equivalent measurement properties in the older and younger adults. We evaluated the internal consistency of the seven items and the consistency of age differences in correct responses and different error types.

We also measured the Raven's Progressive Matrices test to capture inductive reasoning (fluid intelligence), and to evaluate the magnitude of age differences in CRT performance relative to the expected substantial age differences in the Raven's test. In addition, the test was useful as a correlational benchmark in evaluating both age and gender differences in CRT performance. To the extent that quantitative reasoning ability influences correct responses on the CRT, we would expect gender differences in CRT performance, especially in older adults, but not in the Raven's test. We used these data to evaluate whether this was reducible to previously well-known age differences in induction.

Method

Sample

The variables evaluated in this study were taken from available data from larger projects focusing on different research questions. Older adults were paid volunteers recruited from the larger Atlanta metropolitan area who participated in experiments assessing age differences in self-regulated learning (see Ariel, Price, & Hertzog, 2015). Young adults were students enrolled at the Georgia Institute of Technology either participating for pay as part of a project assessing measures of critical thinking skills (Hertzog, Hale, & Krepps, 2015) or as a young comparison group in the self-regulated learning study. These students' SAT mathematics scores rank above the 90th percentile in the U.S. and over 90% have taken advanced placement mathematics classes (http://factbook.gatech.edu/admissions-andenrollment/sat-scores). Study 1 consisted of 50 younger adults (ages 18 to 23, M = 19.48, 38% female) and 40 older adults (ages 60 to 79, M = 69.7). Of these older adults, 63% were women and 85% had attained at least a Bachelor's degree. Study 2 consisted of 170 younger adults (ages 17 to 35, M = 20.32, 60% female) and 74 older adults (ages 61 to 80, M = 69.7; 58% women; 65% held a Bachelor's degree).

Measures

Cognitive Reflection Test.—The original Cognitive Reflection Test (CRT; Frederick, 2005) uses three questions to elicit a Type I or Type II response. All of the questions have an intuitive, easily activated, yet incorrect answer and a correct response that can be reached by analytic processing. Higher scores on the test indicate successful engagement in quantitative reasoning. Toplak, West, and Stanovich (2013) added four additional items to a revised CRT, keeping the original 3 items but appending the new items to the end of the test. Older adults in Study 1 were given the three-item version while younger adults were given the seven-item version. We compared the two groups on the three-item versions, which were identical in order of item administration. In Study 2, all older and younger adults were given the seven-item CRT. All CRTs were administered on a personal computer using LiveCode software (version 7.06; Livecode LTD., 2015). Individuals entered a numeric answer to each question (except the 7th item, which uses a 3-option multiple choice format). Individuals were allowed to skip items rather than being required to enter an answer.

The CRT is scored as the number of correct (Type II) responses. One can also compute the number of intuitive (Type I) CRT errors, numerical errors, and omitted responses (see also Liberali et al., 2011; Sinayev & Peters, 2015). Numeric errors are defined as explicitly provided numbers that are not either correct or the heuristic (Type I) alternative. Omitted are skipped items of unknown origin, although it is reasonable to assume they are often if not always implicit failures to convert the word problem to a numerical representation or some kind of low-confidence calculation error that leads to a decision to skip the item. These four CRT scoring categories are ipsative, in the sense that there is a linear dependency among them (Pennycook et al., 2015). Good performance on the CRT inevitably reduces the raw number of errors of any type that are committed. To address this problem we computed three proportional error scores: proportion of total errors that are: (a) intuitive errors, (b) numeric calculation errors, and (c) omitted responses (i.e., with each type of error divided by the total number of errors). These measures of relative error rates are in theory less dependent on the quality of problem solving, as indexed by the number of correct CRT item responses. Given concerns about the stability of the estimated error proportions for the short original 3-item CRT, in Study 1 we focused only on the number of correct CRT responses and the proportion of intuitive errors. For Study 2 using the 7-item CRT we report all three relative error measures, but of course these three error scores are also not independent of one another, so our focus is on the proportion of intuitive and numeracy errors.

The LiveCode program saved all keystrokes. We counted as correct instances where individuals entered an error but then attempted to correct it (e.g., by re-entering an answer or using the backspace key), and also counted as correct responses the relatively rare instances where individuals made an obvious scaling error, such as responding '.05' instead of 5 (cents) on the first problem. All other explicit numeric responses other than the correct answer or the heuristic foil were classified as numeracy errors.

Raven's Progressive Matrices.—A computerized version of Raven's Progressive Matrices (RPM; Raven & Court, 1998) was administered to assess nonverbal fluid intelligence. The task consisted of 18 items ordered ascending in their normative difficulty, adapted from Stanovich and Cunningham (1993). On each trial, a 3 × 3 array was presented. It consisted of eight geometric figures with a missing ninth figure presented in the bottom right-hand position of the array. Participants could choose from eight potential figures positioned below the 3 × 3 array to complete the pattern. Participants had 12 minutes to complete this task. Performance on Raven's Progressive Matrices was computed as the sum of correctly answered items.

Results

Study 1

Internal consistency.—Table 1 reports the correlations of item responses on the 3-item CRT for younger and older adults. Estimates of internal consistency (Cronbach's alpha) were .56 for younger adults and .68 for older adults. The moderate level of internal consistency could reflect a lack of homogeneity among items, but it could be affected by the small number of items. There was no indication here of particular problems with the CRT for older adults.

Age differences in CRT performance.—Table 2 reports the means and standard deviations for the 3-item CRT test and RPM for both age groups. There were reliable age differences in both measures, with younger adults achieving higher scores on both the CRT, $t(81) = 2.03$, $p = .05$, Cohen's [1988] $d = 0.45$, and the RPM, $t(67.431) = 8.38$, $p<.01$, $d = 1.79$. Clearly the effect size was substantially larger on the test of inductive reasoning. There were no gender differences in RPM scores, $F < 1$, $d = 0.04$. There were significant gender differences in CRT scores, $F(1,79) = 5.12$, $p < .02$, $d = 0.50$, with males attaining higher mean CRT scores ($M_{Male} = 1.22$, SE = 0.16; $M_{Female} =. 70$, SE = 0.16). The interaction between age and gender was not significant, $F < 1$.

The CRT and RPM correlated similarly in younger adults ($r = .32$, $p < .04$) and older adults ($r = .39$, $p < .01$), and the group difference was not reliable when tested after a Fisher's r-to-z transformation ($p < .73$). Age differences in CRT correct responses were not significant when controlling on the RPM, $F(1,75) = 1.77$, $p < .19$.

The mean proportion of intuitive errors was reliably higher in older adults (see Table 2), $t(64.096) = 2.99$, $p < .04$, $d = -0.70$. The correlation between proportion of intuitive errors and RPM was not significant in either group ($r_{Young} = -.33$, $p = .06$; $r_{Old} = -.06$, $p =.71$). There was no effect of gender on proportion of intuitive errors, $F < 1$. Age differences in the proportion of intuitive errors when controlling on RPM were also not significant, $F < 1$. Given the limited opportunity for scores to vary in the 3-item test, we did not focus on the other error scores in Study 1.

Table 2 shows the proportion of correct responses on all 3 CRT items, as well as the proportion of different error types for both older adult and younger adult groups. Given the limited number of items we offer these data for descriptive purposes only. Older adults were relatively more prone to the intuitive errors on the first and third items; indeed, all errors for older adults on the first item were intuitive errors.

Study 2

Internal consistency.—Given the availability of 7 items for the expanded CRT, we computed a scale analysis separately for both groups using the binary outcome (correct versus incorrect). The results were surprising. Estimated internal consistency (Cronbach's alpha) was .66 for younger adults (comparable to Study 1). However, estimated alpha was only .35 for the older adults. Table 3 reports the inter-item correlations correlations that help explain the age differences in internal consistency. The older adults actually generated near zero correlations among several of the items. Checking the cross-tabulation of item responses revealed very limited consistency; for instance, literally no older adult in Study 2 solved both Item 1 and Item 2 correctly. Older adults' errors were also heterogeneous in nature. Table 4 provides the proportion of types of errors made in each of the seven CRT questions for both older adult and younger adult groups. Age differences in intuitive (Type I) response errors varied substantially across the items. As in Study 1, older adults were more likely to provide the designated intuitive error on Item 1, but were actually less likely than younger adults to make intuitive errors on several other items.

Age difference in CRT scale scores.—Although the internal consistency measures indicate potential issues with the aggregate scale scores for older adults, we proceeded with their analysis. Table 2 reports the means and standard deviations of the correct scores and proportion error scores for the 7-item version of the CRT. Age differences in CRT correct responses favoring younger adults were considerably larger in Study 2, $t(201) = -8.400$, $p < .001$, $d = -1.30$. There was also a gender difference in correct CRT scores favoring males, $F(1,192) = 11.99$, $p < .001$ $d = 0.50$ (M_{male}=1.95, SE=.16: M_{female}=1.15, SE=.16). The interaction between age and gender was not significant, $F(1,192) = 1.10$, $p = .30$. Older adults had slightly higher raw numbers of intuitive errors, but younger adults actually had a higher proportion of their errors based on the intuitive response that did not achieve statistical significance, $F(1,188) = 3.14$, $p = .08$, $d = -0.31$. As in Study 1, there were no reliable gender differences in proportion of intuitive errors, $F < 1$.

Given the ipsative nature of the proportional error scores, the complementary proportion indicates that older adults were reliably higher in the combined proportion of numerical errors and omitted responses (.54 vs .46). As can be seen in Table 2, both groups had similar proportions of their errors associated with explicit wrong numerical answers. Hence older adults showed a higher proportion of omitted responses. There was no reliable gender difference in the proportion of calculation errors, $F(1,188) = 2.17$, $p = .14$.

Younger adults also had reliably higher RPM scores, $t(197) = -15.61$, $p < .001$, $d = -2.33$. As in Study 1, there were no gender differences in RPM, $F(1, 189) = 1.24$, $p = .27$, $d = 0.11$. As in Study 1, older adults had somewhat higher correlations of CRT correct scores with RPM scores (r_{Old} = .36 versus r_{Young} = .30) but the difference was not statistically significant, $p < .60$. Table 5 shows the correlations between CRT scores, proportion of CRT errors, the RPM, and scores on the Advanced Vocabulary test (older adults) or the Shipley Vocabulary test (younger adults). Controlling on RPM did not eliminate age differences in CRT correct scores, $F(1, 185) = 8.39$, $p < .004$, in contrast to Study 1. As noted earlier, there were no reliable age differences in the proportion of intuitive errors, and this did not change when covarying for RPM, $F(1, 181) = 1.37$, $p = .25$.

Discussion

The present results provide limited support for the hypothesis that older adults have a greater disposition or cognitive style of engaging in Type I reasoning compared to younger adults. Certainly there were robust age differences in CRT correct responses, probably indicating age differences in effectiveness of Type II reasoning in numerical problem solving. Both age groups showed reliable sex differences favoring males on the CRT but no sex differences in RPM performance. Both of these patterns is consistent with the literature on gender differences in quantitative reasoning and induction, respectively. Likewise, the positive correlations of CRT with RPM in both age groups is consistent with the claim that the test measure fluid intelligence at both ages. Given this evidence, we suggest there are age differences in analytical reasoning as measured by the CRT, consistent with the larger literature on aging and intellectual abilities (e.g., Salthouse, 1992).

Based on the emerging literature on the CRT and cognitive style in younger populations, age differences in CRT performance might generically be attributed to greater tendency towards Type I reasoning in older adults. Whether this inference would be valid is still an open question.

The CRT generated poor internal consistency in older adults in Study 2, and the items varied substantially in terms of age differences in generating the Type I intuitive errors. Consider the relative proportion of intuitive errors as an indicator of a Type I (heuristic) reasoning style. Study 1, with the smaller sample and the original 3-item version of the CRT, produced greater relative intuitive error rates in older adults, consistent with the hypothesis. However, Study 2, with the larger sample and the extended 7-item CRT, did not. Some of the new Toplak et al. (2013) items appear to produce more numeracy errors than intuitive errors when we inspected individual item distributions. However, the divergent Study 2 outcome was not merely due to the addition of four items in the extended CRT; when we restricted the Study 2 analysis to the original three CRT items we obtained similar disparities regarding CRT performance by older adults. The disparity in patterns of age differences in proportion of intuitive errors across the 7-item CRT in Study 2 also argues against a simple account of older adults showing greater Type I reasoning biases on the CRT. We would not rule out the possibility that there are other intuitively-grounded response options besides the answer anticipated by the CRT designers, which might suggest that some of the numeric answers derive from heuristics, not failed attempts at numerical representation. Nevertheless, the available data do not compel us to conclude that older adults are more likely to make the expected Type I reasoning error on CRT items.

It is interesting, then, that older adults were recently found to be less likely to self-report using intuitive or experientially-based decision making styles in everyday life (Delaney et al., 2015) -- consistent with the Study 2 outcomes, but in contrast to other self-report measures on experiential problem solving styles (e.g., Horhota et al., 2012).

Study 2 results also did not strongly support the argument that older adults' lower performance can be attributed to a greater tendency for numeracy-based errors, which represented a rival potential explanation for age differences in the CRT (e.g., Bruine de Bruin et al., 2015; Sinayev & Peters, 2015). That is, older and younger adults did not differ in the proportion of errors based on providing explicitly incorrect numerical responses. The largest age difference in relative error rates was found for omitted responses (skipped items). Certainly skipped items could be due to some form of numeracy problem, such as a failure to achieve an initial numerical problem representation. If so, the age difference in omitted responses could indicate an underlying issue with numeracy. But there are other potential explanations for older adults skipping more items. That behavior could indicate a form of conservative metacognitive control (Hertzog, 2016) operating in the CRT test environment. Older adults could be more loathe to volunteer responses to items when they are unsure about the accuracy of a candidate answer (but for an opposite tendency for episodic memory tests, see Pansky et al., 2009). We regard the numeracy hypothesis as the more likely account of item skipping in the CRT. However, the case is far from closed.

We argue that the CRT does measure Type I reasoning biases in the younger adults in our sample, because in principle all younger adults who are enrolled at a technical institute have adequate numerical problem solving skills to solve the relatively easy CRT items if motivated to engage Type II reasoning skills. Clearly they did not – young adults' mean performance on the CRT in Study 1 was 1.19 out of 3 items correct and in Study 2 was just 2.5 items correct out of 7 total items, which is truly mediocre numerical problem solving performance. Performance on the more widely-used 3-item version of the CRT in Study 1 is comparable to scores seen from other selective universities including Massachusetts Institute of Technology, Princeton University, and University of Michigan: Ann Arbor with mean scores of 2.18, 1.63, and 1.18, respectively (Frederick, 2005). Students assessed at Purdue University on the longer 7-item version achieved M = 2.83 correct (Ariel, personal communication, December 15, 2015). Younger adults also showed reliably higher relative rates of intuitive errors in Study 2 than generated by our older sample. It is highly plausible, then, that the pattern of behavior in our younger samples reflects Type I reasoning bias in the absence of a strong motivation for analytic thinking.

Is this true of our older adults as well? Our older samples were also well educated, on average, although there could well be group differences in numeracy due to the use of Georgia Tech students as a young comparison sample. Certainly older adults are less likely to be solving verbal mathematics problems in everyday life, relative to a student at a technical university. Given the fact that older adults in Study 2 had somewhat lower RPM scores and were somewhat less likely to have a college degree, we also compared high and lower education older adults in Sample 2 to each other (and younger adults) and found no major differences in outcomes. For instance, the education difference in explicit calculation errors in the two older groups was only .02, and both groups of older adults were equivalent in the proportion of errors that were omitted responses. Differences in education may play a role, but a viable account of the pattern of results is that older adults' lower performance is more likely about cognitive decline affecting Type II reasoning than it is a dispositional bias toward Type I reasoning per se.

Taken together, then, these results are fully consistent with age differences in the quality of quantitative reasoning, as has been reported in other studies assessing this construct. This type of analytic (Type II) reasoning is needed to perform well on the CRT. However, the data are ambiguous about whether age differences are also due to a disposition to a heuristically guided, superficial problem analysis (Type I reasoning), greater issues with numeracy in older adults, or other factors. Future research should more carefully evaluate the process by which older adults initially represent and solve CRT items using think-aloud or other strategy assessment techniques (e.g., Cokely & Kelley, 2009; Sinayev & Peters, 2015) to determine whether age differences in cognitive style are truly implicated in the observed age differences in CRT test performance. Such research would also benefit from administering an independent measure of numeracy skills relevant for the CRT (e.g., Cokely et al., 2012; Sinayev & Peters, 2015) so that numeracy contributions to older adults' CRT can be statistically evaluated. For now, we caution gerontologists using the CRT not to axiomatically assume that they are measuring older adults' dispositional bias towards heuristic over analytic thinking when doing so.

Acknowledgments

Robert Ariel is now located at the Department of Psychology, Purdue University. The research was partially supported by the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), via contract number 2014-13121700006. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the ODNI, IARPA, or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. We thank Aimee Kudela, Ryan Krepps, and Alyssa Candelmo for assistance with data collection on that portion of the project.

This research was also supported in part by a Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Research Training Grant from the National Institutes of Health (National Institute on Aging), Grant #5T32AG000175. We thank Antonio Lee, Aishwarya Murthy, Katie Lively, and Natalie Lembeck for help with older adult data collection.

We also thank Michael Bixter for comments on this manuscript.

References

Alter AL, Oppenheimer DM, Epley N, & Eyre RN (2007). Overcoming intuition: Metacognitive difficulty activates analytic reasoning. Journal of Experimental Psychology: General 146(4) 569–576.

Allaire JC (2012). Everyday cognition. In Whitbourne SK & Sliwinski MJ (Eds.), in The Wiley-Blackwell handbook of adulthood and aging. (pp.190–207). Hoboken, NJ: Wiley-Blackwell.

Ariel R, Price J & Hertzog C (2015). Age-related associative memory deficits in value-based remembering: The contribution of agenda-based regulation and strategy use. Psychology and Aging, 30, 795–808. doi: 10.1037/a0039818 [PubMed: 26523692]

Baltes PB (1997). On the incomplete architecture of human ontogeny: Selection, optimization, and compensation as foundation of developmental theory. American Psychology, 52, 366–368.

Baltes PB, & Smith J (2008). The fascination of wisdom: Its nature, ontogeny, and function. Perspectives on Psychological Science, 3, 56–64. [PubMed: 26158670]

Besedes T, Deck C, Sarangi S, Shor M (2012). Decision making strategies and performance among seniors. Journal of Economic Behavior and Organization, 81(2). 524–533. doi:10.1016/j.jebo.2011.07.016. [PubMed: 22408282]

Blanchard-Fields F (2009). Flexible and adaptive socio-emotional problem solving in adult development and aging. Restorative Neurology and Neuroscience, 27, 539–550. [PubMed: 19847075]

Blanchard-Fields F, Hertzog C, Stein R, & Pak R (2001). Beyond a stereotyped view of older adults' traditional family values. Psychology and Aging, 16, 483–496. doi: 10.1037/0882-7974.16.3.483 [PubMed: 11554525]

Bruine de Bruin W, McNair SJ, Taylor AL, Summers B, & Strough J (2015). "Thinking about numbers is not my kind of fun": Need for cognition mediates age differences in numeracy performance. Medical Decision Making, 35, 22–26. doi: 10.1177/0272989X14542485 [PubMed: 25035261]

Bruine de Bruin W, Parker AM, & Fischoff B (2012). Explaining adult age differences in decision making competence. Journal of Behavioral Decision Making, 25, 352–360. doi: 10.1002/bdm712

Bruine de Bruin W, Strough J, & Parker AM (2014). Getting older isn't all that bad: Better decisions and coping when facing "sunk costs." Psychology and Aging, 29, 642–647. doi: 10.1037/a0036308 [PubMed: 25244483]

Campitelli GJ & Gerrans P (2013). Does the Cognitive Reflection Test measure cognitive reflection? A mathematical modeling approach. Memory and Cognition, 42(3), 434–447. doi: 10.3758/s13421-013-0367-9

Cohen J (1988). Statistical power analysis for the behavioral sciences (rev. ed.). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Cokely ET, Galesic M, Schulz E, Ghazal S, & Garcia-Retamero R (2012). Measuring risk literacy: The Berlin Numeracy Test. Judgment and Decision Making, 7, 25–47.

Cokely ET, & Kelley CM (2009). Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. Judgment and Decision Making, 4, 20–33.

Cokely ET, Parpart P, & Schooler LJ (2009). On the link between cognitive control and heuristic processes. In Taatgen N, & Van Rijn H (Eds.), Proceedings of the 31st Annual Conference of the Cognitive Science Society (pp.2926–2931). Austin, TX: Cognitive Science Society.

Delaney R, Strough J, Parker AM, & Bruine de Bruin W (2015). Variations in decision making profiles by age and gender: A cluster-analytic approach. Personality and Individual Differences, 85, 19–24. doi: 10.1016/j.paid.2015.04.034 [PubMed: 26005238]

Del Missier F, Mäntylä T, Hansson P, Bruine de Bruin W, Parker AM, & Nilsson L-G. (2013). The multifold relationship between memory and decision making: An individual differences study. Journal of Experimental Psychology: Learning, Memory, and Cognition, 39(5). 1344–1364. doi:10.1037/a0032379 [PubMed: 23565790]

Dickman SJ & Meyer DE (1988). Impulsivity and speed-accuracy tradeoffs in information processing. Journal of Personality and Social Psychology, 54, 274–290. [PubMed: 3346814]

Diehl M, Willis SL, & Schaie KW (1995). Everyday problem solving in older adults – Observational assessment and cognitive correlates. Psychology and Aging, 10, 478–491. doi: 10.1037/0882-7974.10.3.478 [PubMed: 8527068]

Der G, Allerhand M, Starr JM, Hofer SM, & Deary I (2010). Age-related changes in memory and fluid intelligence in a sample of healthy old people. Aging, Neuropsychology, and Cognition, 17, 55–70. doi: 10.1080/13825580903009071

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, 390–405. [PubMed: 8765488]

Evans J. St. B.T. (2009). How many dual-process theories do we need: One, two or many? In Evans J. St. B. T. & Frankis K (Eds.), In two minds: Dual process and beyond (pp. 31–54). Oxford, England: Oxford University Press.

Evans J.St.B.T., & Over DE (1996). Rationality in the section task: Epistemic utility versus uncertainty reduction. Psychological Review, 103(2). 356–363. doi: 10.1037/0033-295X.103.2.356

Evans J. St. B. T. & Stanovich KE (2013). Dual-process theories of higher cognition: Advancing the debate. Perspectives on Psychological Science, 8(3), 223–241. doi: 10.1177/1745691612460685 [PubMed: 26172965]

Finucane ML & Gullion CM (2010). Developing a tool for measuring the decision-making competence of older adults. Psychology and Aging 25(2). 271–288. doi:10.1037/a0019106. [PubMed: 20545413]

Fisk JE (2005). Age and probabilistic reasoning: Biases in conjunctive, disjunctive and Bayesian judgments in early and late adulthood. Journal of Behavioral Decision Making, 18(1), 55–82.

Fisk JE, & Sharp C (2002). Syllogistic reasoning and cognitive aging. The quarterly journal of experimental psychology Section A: Experimental psychology, 55(4), 1273–1293. doi:10.1080/02724980244000107

Frederick S (2005). Cognitive reflection and decision making. The Journal of Economic Perspectives, 19(4), 25–42. http://www.jstor.org/stable/4134953

Fox MC & Mitchum AL (2013). A knowledge based theory of rising scores on “culture-free” tests. Journal of Experimental Psychology: General, 142, 979–1000. [PubMed: 23025559]

Heron A, & Chown SM (1967). Age and function. London: UK: Churchill.

Hershey DA, Walsh DA, Read SJ, & Chulef AS (1990). The effects of expertise on financial problem solving: Evidence for goal-directed, problem-solving scripts. Organizational Behavior and Human Decision Processes, 46, 77–101.

Hertzog C (2016). Aging and metacognitive control. In Dunlosky J & Tauber SK (Eds), Oxford Handbook of Metacognition. Oxford, UK: Oxford University Press.

Hertzog C, Hale C, & Krepps RD (2015). Development of a Quantitative Measure of Abductive Reasoning Quality. Georgia Institute of Technology Final Technical Report, Intelligence Advanced Research Projects Activity (IARPA) Strengthening Human Adaptive Reasoning and Problem-solving (SHARP) Program, August, 2015.

Hess TM (2001). Ageing-related influences on personal need for structure. International Journal of Behavioral Development, 25, 482–490.

Hess TM (2014) Selective engagement of cognitive resources: Motivational influences on older adults’ cognitive functioning. Perspective on Psychological Science 9(4), 388–407. doi: 10.1177/1745691614527465

Hess TM, & Queen TL (2014). Aging influences on judgments and decision processes: Interactions between ability and experience. In Verhaeghen P & Hertzog C (Eds.), The Oxford handbook of emotion, social cognition, and problem solving in adulthood (pp.238–255). New York: Oxford University Press.

Horhota M, Mienaltowski A, & Blanchard-Fields F (2012). If only I had taken my usual route…: Age-related differences in counterfactual thinking. Aging, Neuropsychology, and Cognition, 19, 339–361.

Kozhevnikov M, Evans C, & Kosslyn SM (2014). Cognitive style as environmentally sensitive individual differences in cognition: A modern synthesis and applications in education, business, and management. Psychological Science in the Public Interest, 15(1), 3–33. [PubMed: 26171827]

Låg T, Bauger L, Lindberg M, & Friborg O (2013). The role of numeracy and intelligence in health-risk estimation and medical data interpretation. Journal of Behavioral Decision Making, 27(2), 95–108. 10.1002/bdm.1788.

Liberali JM, Reyna VF, Furlan S, Stein LM & Pardo ST (2011). Individual differences in numeracy and cognitive reflection, with implications for biases and fallacies in probability judgment. Journal of Behavioral Decision Making, 25, 361–381. doi: 10.1002/bdm.752

Livecode LTD. (2015). Livecode [Computer Software]. Edinburgh Scotland, UK: Livecode LTD.

Pachur T, Mata R, & Schooler LJ (2009). Cognitive aging and the adaptive use of recognition in decision making. Psychology and Aging, 24(4), 901–915. [PubMed: 20025405]

Pansky A, Goldsmith M, Koriat A, & Pearlman-Avnion S (2009). Memory accuracy in old age: Cognitive, metacognitive, and neurocognitive determinants. European Journal of Cognitive Psychology, 21(2-3), 303–329. doi: 10.1080/09541440802281183

Park DC, Hertzog C, Leventhal H, Morrell RW, Leventhal E, Birchmore D, Martin M, & Bennett J (1999). Medication adherence in rheumatoid arthritis patients: Older is wiser. Journal of the American Geriatric Society, 47, 172–183. doi; 10.1111/j.1532-5415.1999.tb04575

Pennycook G, Cheyne JA, Koehler DJ, & Fugelsang JA (2015). IS the cognitive reflection test a measure of both reflection and intuition? Behavioral Research Methods. doi: 10.3758/s13428-015-0576-1

Peters E, Hess TM, Auman C, & Västfjäll D (2007). Adult age differences in dual information processes and their influence on judgments and decisions: A review. Perspectives on Psychological Science, 2, 1–23. [PubMed: 26151915]

Pietschnig J, & Voracek M (2015). One century of global IQ gains: A formal meta-analysis of the Flynn effect. Perspectives on Psychological Science, 10, 282–306. doi: 10.1177/1745691615577701 [PubMed: 25987509]

Primi C,Morsanyi K, Chiesi F, Donati MA, & Hamilton J (2015). The development and testing of a new version of the Cognitive Reflection Test applying item response theory. Journal of Behavioral Decision Making. Prepublication dated 6/14/15. doi: 10.1002/bdm.1883.

Raven JC, & Court JH (1998). Raven's progressive matrices and vocabulary scales. Oxford Psychologists Press.

Salthouse TA (1992). Influence of processing speed on adult age differences in working memory. Acta Psychologica 79, 155–170. [PubMed: 1598844]

Salthouse TA, Mitchell DRD, Skorvronek E & Babcock RL (1989). Effects of adult age and working memory on reasoning and spatial abilities. Journal of Experimental Psychology: Learning, Memory, and Cognition 15(3). 507–516. [PubMed: 2524548]

Schaie KW (2005). Developmental influences on adult intelligence: The Seattle Longitudinal Study. New York: Oxford University Press.

Schnitzspahn KM, Ihle A, Henry JD, Rendell PD, & Kliegel M (2011). The age-prospective memory-paradox. International Psychogeriatrics, 23(4). 583–592. doi:10.1017/S1041610210001651 [PubMed: 20843394]

Sinayev A, & Peters E (2015). Cognitive reflection vs. calculation in decision making. Frontiers in Psychology, 6(532). doi: 10.1002/bdm.752

Stanovich KE (2009). What intelligence tests miss: The psychology of rational thought. New Haven, CT: Yale University Press.

Stanovich KE (2011). Rationality and the Reflective Mind. New York, NY: Oxford University Press.

Stanovich KE, & Cunningham AE (1993). Where does knowledge come from? Specific associations between print exposure and information acquisition. Journal of Educational Psychology, 85, 211–229.

Staudinger UM, & Glück J (2011). Psychological wisdom research: Commonalities and differences in a growing field. Annual Review of Psychology, 62, 215–241. doi: 10.1146/annurev.psych.121208.131659

Strough J & Keener EJ (2014). Goals and strategies for solving interpersonal everyday problems across the lifespan. In Verhaeghen P & Hertzog C, Eds.), The Oxford Handbook of Emotion, Social Cognition, and Problem Solving in Adulthood (pp. 190–205). Oxford, U.K., Oxford University Press.

Toplak ME, West RF, & Stanovich KE (2011). The cognitive reflection test as a predictor of performance on heuristics-and-biases tasks. Memory & Cognition, 39, 1275–1289. [PubMed: 21541821]

Toplak ME, West RF, & Stanovich KE (2013). Assessing miserly information processing: An expansion of the Cognitive Reflection Test. Thinking and Reasoning 20(2). 147–168. doi:10.1080/13546783.2013.844729

Viskontas IV, Morrison RG, Holyoak KJ, Hummel JE, & Knowlton BJ (2004) Relational integration, inhibition and analogical reasoning in older adults. Psychology and Aging, 19, 581–591. [PubMed: 15584784]

Zhang L-F, & Sternberg RJ (2006). The nature of intellectual styles. Mahwah, NJ: Lawrence Erlbaum Associates.

Table 1.

Mean Proportion Correct for each CRT Item and Item Correlations for Young Adults and Old Adults, Study 1.

	M_Y	M_o	CRT 1	CRT 2	CRT 3
CRT 1	.40	.20	-	.22	.48
CRT 2	.26	.30	.29	-	.56
CRT 3	.53	.23	.18	.44	-

Note. Young adult item correlations below diagonal; older adult item correlations above diagonal

Table 2.

Mean and Standard Deviations of Cognitive Reflection Test Scores and Raven's Progressive Matrices for Young and Old Adults in Study 1 and Study 2.

	Young		Old	
	M	**SD**	**M**	**SD**
Study 1 (3 items)				
CRT Correct	1.19	1.05	0.68	0.92
Intuitive Errors	1.14	1.10	1.90	1.03
Proportion Intuitive	.59	.38	0.78	0.31
RPM	8.74	3.50	3.75	1.84
Study 2 (7 items)				
CRT Correct	2.53	1.93	0.59	.870
Intuitive Errors	2.40	1.60	2.96	1.47
Calculation Errors	.98	1.03	1.56	1.15
Skipped Items	1.09	1.06	1.89	1.22
Proportion Intuitive	.54	.29	.46	.22
Proportion Calculation	.22	.23	.25	.18
RPM	9.24	2.97	3.25	2.10

Note. Mean = mean, *SD* = standard deviation, *CRT correct* = raw number of CRT items answered correctly, *RPM* = Raven's Progressive Matrices, *Intuitive Errors* = raw number of CRT items answered with the "intuitive" response, *Skipped Items* = raw number of omitted answers, *Proportion Intuitive* = proportion of total errors with designated intuitive response, and *Proportion Calculation* = proportion of total errors with explicit incorrect numeric responses.

Table 3.

Proportion of Correct Item Responses and Corrected Proportion of Error Scores (Relative to Total Number of Errors) for Each CRT Item in Study 1 for Young and Old Adults

	Item 1		Item 2		Item 3	
Error Type	**Young**	**Old**	**Young**	**Old**	**Young**	**Old**
Correct	.40	.20	.26	.30	.53	.23
Intuitive Errors	.65	1.00	.75	.68	.40	.81
Calculation Errors	.27	.00	.16	.32	.15	.16
Skip	.08	.00	.09	.00	.45	.03

Table 4.

Mean Proportion Correct for each CRT Item and Item Correlations for Young Adults and Old Adults, Study 2.

	M_Y	M_O	CRT 1	CRT 2	CRT 3	CRT 4	CRT 5	CRT 6	CRT 7
CRT 1	.34	.03	-	-.03	.17	−.19	.29	−.07	.17
CRT 2	.26	.04	.28	-	.11	−.02	−.05	−.08	.11
CRT 3	.48	.14	.41	.25	-	−.05	.19	.05	.36
CRT 4	.35	.01	.22	.26	.37	-	−.03	−.05	.05
CRT 5	.34	.06	.21	.12	.27	.29	-	−.11	.34
CRT 6	.39	.14	.10	−.05	.07	.17	.13	-	.05
CRT 7	.38	.14	.14	.26	.22	.36	.32	.20	-

Note. Young adult item correlations below diagonal; older adult item correlations above diagonal

Table 5.

Mean Proportion of Different Types of Errors for each CRT item for Young and Old Adults, Study 2

	Question 1		Question 2		Question 3		Question 4		Question 5		Question 6		Question 7	
Error Type	**Young**	**Old**	**Young**	**Old**	**Young**	**Old**	**Young**	**Old**	**Young**	**Old**	**Young**	**Old**	**Young**	**Old**
Intuitive	.74	.84	.60	.49	.57	.53	.31	.20	.59	.43	.51	.46	.43	.28
Calculation	.09	.03	.17	.29	.22	.13	.34	.40	.32	.48	.36	.30	.05	.06
Skip	.17	.13	.23	.22	.22	.34	.35	.40	.10	.11	.13	.24	.25	.66

Table 6.

Correlations between scores on the Cognitive Reflection Test, Raven's Progressive Matrices, and Vocabulary for Young and Old adults.

	Correct	Prop Int	Prop Skip	Prop Calc	RPM	Vocab*
Correct	-	−.33**	−.21	−.12	.36**	.15
Prop Int	−.67**	-	−.50**	−.50*	−.14	−.28
Prop Skip	−.39**	−.18	-	−.27**	−.11	−.17
Prop Calc	−.43**	−.12	−.02	-	.02	−.18
RPM	.29**	−.16	−.09	.21*	-	.49*
Vocab	.15	−.08	−.03	.12	.07	-

Note. Correct = Raw number of CRT items answered correctly, *Prop Int*= Proportion of errors that were designated as the intuitive alternative response by test authors, *Prop Calculation* = Proportion of errors that were explicit incorrect numeric responses, *Prop Skip* = Proportion of errors that were due to an omitted answer, RPM = Raven's Progressive Matrices score, and *Vocab* = Score on the Advanced Vocabulary Test (old adults) or the Shipley Vocabulary test (young adults)