

Contents lists available at [ScienceDirect](https://www.sciencedirect.com)

## Journal of Economic Behavior and Organization

journal homepage: [www.elsevier.com/locate/jebo](http://www.elsevier.com/locate/jebo)Measures of cognitive ability and choice inconsistency<sup>☆</sup>Helene Willadsen<sup>a</sup>, Sarah Zaccagni<sup>b,c,\*</sup>, Marco Piovesan<sup>d</sup>, Erik Wengström<sup>e</sup><sup>a</sup> Department of Economics and Copenhagen Center for Social Data Science, University of Copenhagen, Øster Farimagsgade 5, DK-1353 Copenhagen, Denmark<sup>b</sup> Aarhus University, Department of Economics and Business Economics, Denmark<sup>c</sup> CEBI (University of Copenhagen) Fuglesangs Allé 4 8210 Aarhus V, Denmark Denmark<sup>d</sup> Department of Economics, University of Verona, Polo Santa Marta, 37129 Verona, Italy<sup>e</sup> Department of Economics, Lund University, Sweden, and Hanken School of Economics, Helsinki, Finland

## ARTICLE INFO

## JEL codes:

C90  
C91  
D81  
D91  
J13

## Keywords:

Intelligence  
Raven Progressive Matrices  
Cognitive Reflection  
Cognitive Skills  
Inconsistent Behavior  
Time Preference  
Risk Preference

## ABSTRACT

Cognitive skills affect individual choices. Researchers commonly use Raven's Progressive Matrices (RPM) tests and the Cognitive Reflection Test (CRT) to assess the relationship between cognitive abilities and economic decision making. In this paper, we study the relationship between these measures, and investigate the extent to which they are correlated and whether they are best described as substitutes or complements. Combining a sample of 686 children and a sample of 2,332 adults, we compare individual performances in the RPM test and CRT test. First, we report a significant positive correlation between the two measures of 0.3. Second, we document that performance in both the RPM test and CRT are significant predictors of behavioral inconsistency observed in incentivized time and risk preference elicitation tasks for children and risk preference elicitation task for adults.

## 1. Introduction

One way to explain the observed heterogeneity in economic outcomes and inequality within and across societies is to control for individual cognitive ability (Meisenberg, 2012; Burks et al., 2009). For instance, cognitive abilities play a role in explaining differences in financial, educational, and health decisions (Burks et al., 2015; Muñoz-Murillo et al., 2020; Capraro & Peltola, 2018; Caviglia-Harris & Maier, 2020; Bijwaard et al., 2015). In addition, heterogeneity in cognitive abilities may affect the likelihood of mistakes and inconsistencies both in surveys and in incentivized choice experiments and, thus, bias the estimation of subjects' preferences (Nielsen & Rehbeck, 2022).

Unfortunately, there is not a unique measure of cognitive abilities, and there is also not a unique way to measure cognitive abilities

<sup>☆</sup> We are extremely grateful to the editors Daniela Puzzello and Eugenio Proto, and to two anonymous reviewers for their feedback. We thank the Carlsberg Foundation for generous financial support (Project: 108579). The activities of CEBI are financed by the Danish National Research Foundation, Grant DNR134. Helene Willadsen gratefully acknowledges funding from the H2020 European Research Council (grant number 834540) as part of the project "DISTRACT". We thank the numerous collaborators and students who have helped us in preparing and running this experiment. We thank the staff at the participating schools, the participating children, and their parents.

\* Corresponding author.

E-mail address: [sarah.zaccagni@econ.au.dk](mailto:sarah.zaccagni@econ.au.dk) (S. Zaccagni).

<https://doi.org/10.1016/j.jebo.2024.02.029>

Received 27 June 2023; Received in revised form 14 December 2023; Accepted 19 February 2024

Available online 28 February 2024

0167-2681/© 2024 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

(Toplak et al., 2011; Sirota et al., 2021). Fluid intelligence measures the ability to solve novel tasks based on reasoning and is often measured using logical thinking tasks. For instance, several researchers across disciplines use Raven's Progressive Matrices as a test for general intelligence or shorter the Intelligence Quotient (IQ), from 5-year-olds to the elderly (Kaplan and Saccuzzo, 2009).<sup>1</sup> Although there are many advantages in using a well-established tool as the Raven's Progressive Matrices test, one limitation is the time required to administer it (Bilker et al., 2012). Sometimes, researchers adopt a simplified version of standard IQ tests (Morton et al., 2011) or use years of education as a proxy for cognitive abilities (Castillo et al., 2018; Andreoni et al., 2020). As a different measure, Frederick (2005) proposes the Cognitive Reflection Test (CRT test), a set of three seemingly easy questions that encourage an intuitive but incorrect answer. The CRT test involves specific dimensions of cognitive ability, such as reflectivity, the tendency to engage in fully disjunctive reasoning, and seeking alternative solutions (Stanovich, 2012; Alós-Ferrer et al., 2016).

In this paper, we first add evidence of the positive and significant correlation between the Raven Progressive Matrices (RPM) test score and the Cognitive Reflection Test (CRT) score. The CRT has been shown to correlate ( $r=0.26$ ) with RPM test among millennials (Corgnet, Espin and Hernán-Gonzales, 2016) and to correlate ( $r=0.36$ ) with RPM among older individuals aged 60–79 (Hertzog, Smith and Ariel, 2018). The CRT is also correlated with other measures of cognitive ability (Toplak et al., 2011), like the Wechsler Abbreviated Scale of Intelligence (WASI) and working memory (Paced Auditory Serial Addition Test).

Second, we compare the performance in the RPM test and in the CRT, to investigate whether they complement or substitute each other in explaining behavioral inconsistencies. Cognitive abilities might influence decisions directly or through the mediation of preferences (i.e., risk and time preferences). In both cases, they affect individual decisions and, therefore, outcomes at a later stage in life. Most of the studies that use administrative or experimental data to investigate the relation between cognitive abilities and risk-preferences find that higher cognitive abilities are associated with more risk-taking (Cokely & Kelley, 2009; Burks et al., 2009; Dohmen et al., 2010; Drouvelis & Lohse, 2020; Campitelli & Labollita, 2010; Branäs-Garza & Rustichini, 2011; Rustichini et al., 2016; Benjamin et al., 2013; Taylor, 2013; Booth et al., 2014; Cueva et al., 2015). However, null results are common and found in several papers (Amador-Hidalgo et al., 2021; Andersson et al., 2016; Tymula et al., 2012; Taylor, 2013; and Pachur et al., 2017). This relationship becomes even less clear if we consider differences between loss-domain and gain-domain (Dohmen et al., 2018).

Evidence on the relationship between cognitive abilities and time preferences suggests that individuals with higher cognitive abilities are more patient, in both the short- and long-run (Burks et al., 2009; Dohmen et al., 2010). Other papers focus on the role of cognitive abilities in strategic sophistication (Fe et al., 2022), pro-social behavior (Bonan et al., 2021), behavior in repeated interactions (Gill & Prowse, 2016), behavior in strategic-form games (Bayer & Renou, 2016), and making complex decisions (Mischkowski et al. 2021, Alós-Ferrer et al., 2016). Specifically on children, Bügelmayr & Spieß (2014) model the relationship between cognitive skills and spiteful behavior. Fewer papers report the effect of more than one measure of cognitive ability comparing the effect of each measure on the outcome of interest. Amador-Hidalgo et al. (2021) use both number-additions, CRT, and a test for measuring convergent thinking and report a negative correlation between the two latter and being inconsistent in a risk elicitation experiment. Andersson et al. (2016) uses both RPM test and CRT and shows that manipulating the price list can indeed alter the estimated relationship between both the RPM test and the CRT and risk preferences. Brañas-Garza et al. (2012) also uses the RPM and CRT and show that RPM test is not associated with behavior in the Beauty Contest Game, but those with higher CRT scores are more likely to play dominant strategies in the Beauty Contest Game.

Our contribution is to estimate how similar these two tests are in predicting behavioral inconsistency observed in risk and time preferences elicitation and to show how robust this correlation is across the full lifespan. When studying the role of cognitive abilities on decision making, a critical issue is whether decisions and behaviors adopted by subjects with lower cognitive abilities are more likely driven by errors and a lack of understanding than preferences (Alós-Ferrer & Garagnani, 2021). Andersson et al. (2016; 2020) investigate this issue using the same data on adults as we do. Those papers start by demonstrating that decision noise can lead to overestimation or underestimation of risk aversion depending on the construction of the multiple price list. If IQ is related to decision noise, the inferred relationship between IQ and risk preferences will thus depend on the specifics of the price list. When the switch point for a risk neutral individual is placed early in the list, noise-prone subjects with low IQ are likely to switch later in the list and hence appear more risk averse than less noisy high-IQ subjects. In contrast, if the switch point is placed further down the list, the relationship between IQ and estimated risk preferences is reversed. Andersson et al. (2016) show that manipulating the price list can indeed alter the estimated relationship between IQ and risk preferences.<sup>2</sup> To establish that the findings are driven by a relation between IQ and decision noise, Andersson et al. (2016; Table 8), demonstrate that consistent subjects have higher average cognitive ability than inconsistent subjects. In the present paper, we analyze the relationship between IQ and inconsistencies further by studying it separately by age groups and by using regression analyses. Most importantly, in line with the overall theme of the present paper, we also perform all analyses using the cognitive reflection scores to provide a comparison between the two measures.

We combine two separate and independent large datasets of 686 children and 2,332 adults living in Denmark to investigate 1) whether a 3-item CRT test is associated with individual performance in the Raven Progressive Matrices (RPM) test and, 2) whether cognitive abilities (measured by either CRT test or RPM test) significantly predict inconsistent behavior when using Multiple Price Lists for preferences elicitation. For children, we elicit cognitive abilities using both a RPM test (14-point scale) and a CRT test (4-point scale) and risk and time preferences using both incentivized choice experiments and hypothetical tasks. For adults, we elicit cognitive

<sup>1</sup> For additional evidence on advantages and limitations of using the Raven's Progressive Matrices test to measure general intelligence, refer to Gignac, 2015.

<sup>2</sup> Amador-Hidalgo et al., 2021; Bonan et al., 2021 have subsequently corroborated that noise is related to measures of cognitive ability and that it affects the estimated relationship between cognitive ability and preferences.

abilities using both a RPM test (20-point scale) and a CRT test (4-point scale) and risk preferences using the Multiple Price Lists. Both datasets are crucial for our analysis: if the children dataset report the most novel results, the adult dataset adds robustness (and consistency) to our results.

Our results show that the positive correlations between RPM and CRT tests are significant in both samples and similar in terms of magnitude (0.32 in children, and 0.30 in adults). The correlations are significant for males and females, and in almost all the age groups, ranging from a minimum of 0.23<sup>3</sup> to a maximum of 0.34. In addition, both these measures of cognitive ability predict behavioral inconsistency in time and risk preferences in children and in risk preferences in adults. Interestingly, this prediction holds across all walks of life, from children of nine years old to adults of 80 years old.

## 1.2. Measurement of cognitive abilities

Cognitive ability encompasses the diverse functions of the human brain, including memory storage, information processing, and key aspects such as attention, memory, logical reasoning, and thinking transformation. In the debate on the investigation of cognitive ability, general intelligence signifies a broad mental capacity influencing performance on cognitive ability assessments. This overarching concept is further categorized into fluid intelligence and crystallized intelligence. Fluid intelligence involves fundamental processes of reasoning and mental activities that rely minimally on prior learning, such as formal and informal education, and cultural influences. On the other hand, crystallized intelligence encompasses learned procedures and knowledge, reflecting the impact of experience and acculturation.

Psychology researchers employ various methodologies to assess levels of fluid and crystallized intelligence. Prominent tests for fluid intelligence include the Raven's Progressive Matrices Test (RPM), the Woodcock-Johnson Test of Cognitive Abilities, and the Wechsler Intelligence Scale for Children. Assessments for crystallized intelligence often involve Vocabulary and general knowledge tests, along with the Wechsler Adult Intelligence Scale (WAIS). In a more recent development, [Frederick \(2005\)](#) identified a specific cognitive ability - the inclination or capacity to reflect on a question and resist immediately reporting the initial response that comes to mind. We referred to that as cognitive reflection.

In this paper, we elicit cognitive ability focusing on fluid intelligence and cognitive reflection.

For fluid intelligence, adult subjects answer 20 questions from the "IST2000R" in 10 min; for cognitive reflection, they are administered the classic CRT test by [Frederick \(2005\)](#). For fluid intelligence, children subjects solve 14 Raven-type matrices in 5 min; for cognitive reflection, they are administered three questions from [Young et al \(2018\)](#) (i.e. questions 1, 2 and 7).<sup>4</sup> We used measurements with different features mostly due to the age differences of the two subgroups.

For Raven's Progressive Matrices, the test-retest reliability has been investigated and ranges between 0.89 for children, where the test is repeated after a week ([Raven et al., 1995](#)) down to 0.50 when the test is repeated after two years ([Kazlauskaite & Lynn, 2002](#)). For adults the test-retest reliability for a 14-item version retaken within 7–10 days is reported to be 0.76; and 0.83 for a 36-item version ([Arthur et al., 1999](#)). For CRT, a study reports a test-retest reliability among adults of 0.806 ([Stagnaro et al., 2018](#)).

## 2. Data: children and adults

### 2.1. Children

Our children dataset comprises 686 subjects (52 % boys) aged 9–16<sup>5</sup> in 42 different classes from 10 primary schools in Denmark: 168 in 3<sup>rd</sup> grade, 241 in 5<sup>th</sup> grade, 157 in 7<sup>th</sup> grade, and 120 in 8–9<sup>th</sup> grade. These subjects participated in a series of experiments and questionnaires as part of the KIDS project ([Piovesan & Willadsen, 2021](#)).<sup>6</sup>

For participating children, we have two measures of cognitive abilities: Intelligence Quotient and Cognitive Reflection. We measure children's intelligence using progressive matrices (similar in spirit to Raven's colored progressive matrices) specifically designed to assess children's and adolescents' ability to deduce relationships rather than reproduce learned material. In our test, children see 14 different matrices that increase in difficulty. There is a large picture consisting of 1–9 smaller geometric patterns that follow some rules in the matrices. In all matrices, the bottom-right part is missing. The subjects need to infer the rule that relates the collection in the big matrix and use this rule to select the missing element among 6–8 possible candidates. We measure fluid intelligence as the number of correctly solved matrices in the RPM test. The literature has used this type of test extensively, especially with children and adolescents, since the test does not require language processing skills. Finally, there are no gender and cultural differences in performance in this test ([Court, 1983](#); [Urbina, 2011](#)). Our subjects had 5 min to answer as many of the 14 matrices as possible. With instructions, this task took approximately 6 min.<sup>7</sup>

We measure cognitive reflection using the Cognitive Reflection Test (CRT test; [Frederick, 2005](#)). This test measures the tendency to override an intuitive but incorrect response and engages in further reflection to find the correct answer. We used the questions 1, 2, and

<sup>3</sup> Excluding age-groups of 13-14 years old where the correlation is insignificant.

<sup>4</sup> More details about the tests are available in the data Section and Appendix B.

<sup>5</sup> Three children are reported as younger than 9 years old, and one child is reported to be older than 16 years old. This happens as there is some flexibility regarding school starting age. However, the vast majority of children fall into the age-groups we report in the tables.

<sup>6</sup> See Appendix B.1 for General Information on the KIDS Project. Further details are also available on the website: <http://kidslab.dk/>.

<sup>7</sup> See Appendix B.2 – RPM elicitation.

7 from Young et al., (2018), translated to Danish.<sup>8</sup> Young et al. (2018) proposed a CRT test for children comprising nine questions, and this test has been used to predict performance in science in school (Young & Shtulman, 2020). We read each question aloud twice and asked children to write their answers immediately and in private. This task took approximately 2 min.<sup>9</sup> For the sample of children, we also have measures of time and risk preferences elicited using Multiple Price Lists (MPL) as both hypothetical choices and incentivized choices, with four weeks between the two measures. For time preferences, subjects choose between an amount today and six equal or larger amounts in four weeks; for risk preferences, children choose between a lottery and an increasingly safer option (11 choices). In the incentivized choice experiment, we randomly pick one of their decisions in the time and risk experiment to determine the payoff (tokens later converted into small toys and school materials). MPL allows time and risk preferences to be inferred by looking at the switching point. If a child has multiple switching points, we categorize this as preference *inconsistency*. We assess the role of cognitive abilities in predicting inconsistency in both time and risk preferences.

## 2.2. Adults

Our adult data comes from an experiment run in 2008 in Denmark and comprises 2,232 subjects aged 18–80. These participants were randomly selected from the Danish population and recruited via hard-copy invitation letters in collaboration with Statistics Denmark (the Danish national bureau of statistics). The invitation letters invited the participants to log on to an online platform (iLEE internet Laboratory for Experimental Economics developed at the University of Copenhagen). Subjects were paid for their participation using a bank transfer.

For participating adults, we have two measures of cognitive abilities: Intelligence Quotient and Cognitive Reflection. We measure fluid intelligence with a module of a standard intelligence test called “IST 2000 R”. The module we use is a variation of Raven’s Progressive Matrices (RPM) test (Raven, 1938). It provides a measure of fluid intelligence and does not depend much on verbal skills or other kinds of knowledge taught during formal education. The test consists of 20 tasks in which a matrix of symbols must be completed by picking the symbol that fits best from a selection presented to the subjects (see Appendix B.3 for a screenshot). Subjects had 10 min to work on the tasks. The cognitive ability score used in the analysis is simply the number of tasks a subject managed to solve correctly. As in the children dataset, we measure cognitive reflection using the Cognitive Reflection Test (CRT test) proposed by Frederick (2005). The test is composed of three questions designed to capture the ability or disposition to reflect on a question and to resist reporting the first response that springs to mind. The CRT test’s score refers to the number of correct answers to the three questions.<sup>1011</sup>

For this sample of adults, we have a measure of risk preferences (but not time preferences) and whether a subject was inconsistent in the task. We elicited risk preferences using multiple price lists (MPL) as incentivized choice experiments. Subjects were asked to make a series of 10 choices between binary outcome gambles. The results of Dave et al. (2010) show that participants with a low level of numeracy have difficulties in understanding MPL formats with varying probabilities. Hence, to make the tasks easy to comprehend, which seemed important given the broad sample we targeted, we used fixed 50–50 probabilities and instead varied the prizes (like Binswanger, 1980 or Tanaka et al., 2010). One gamble was selected for payment after the experiment.<sup>12</sup> Screenshots and translations of the instructions are available in Appendix B.3. Similar to the children’s data, we categorize a subject that has multiple switching points as being inconsistent.

The experiment also contained other modules in addition to the ones we use in this paper (e.g., a public good game and survey questions). The description of the data tracks the one in Andersson et al. (2016) and Andersson et al. (2020) very closely, since we use the same experiment.

## 3. Results

### 3.1. Correlation between RPM test and CRT test

Table 1 summarizes our measures of fluid intelligence, cognitive reflection, and the correlation between the RPM test and CRT test by age and gender. Our results show that the score in the RPM test significantly increases with age among children: from 7.70 correct

<sup>8</sup> Question 1) If you’re running a race and you pass the person in second place, what place are you in? [Intuitive answer: First; Correct answer: Second]; question 2) Emily’s father has three daughters. The first two are named Monday and Tuesday. What is the third daughter’s name? [Intuitive answer: Wednesday; Correct answer: Emily]; question 3) What hatches from a butterfly egg? [Intuitive answer: baby butterfly; Correct answer: caterpillar]

<sup>9</sup> See Appendix B.2 – CRT.

<sup>10</sup> Question 1) A bat and a ball cost \$1.10 in total. The bat costs \$1.00 more than the ball. How much does the ball cost? [Intuitive answer: 10 cents; Correct answer: 5 cents]; question 2) If it takes 5 machines 5 min to make 5 widgets, how long would it take 100 machines to make 100 widgets? [Intuitive answer: 100 min; Correct answer: 5 min]; question 3) In a lake, there is a patch of lily pads. Every day, the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of the lake? [Intuitive answer: 24 days; Correct answer: 47 days]

<sup>11</sup> Due to the age difference between the two subgroups (children and adults), we measure intelligence and cognitive ability with slightly different metrics. As this might have an impact on the analysis, in the following sections of the manuscript we present results obtained from analyses ran on the separate samples, rather than pooling them.

<sup>12</sup> The experiment contained a total of 17 choices between gambles, divided into one screen with 10 choices capturing risk aversion and one screen with 7 choices designed to capture loss aversion. We only use the data from the first screen.

questions for 9–10 years old, to 9.91 for 14–16 years old (Cuzick's Wilcoxon-type test for trends,  $p < 0.001$ ). Similarly, the number of correct answers in the CRT test significantly increases with age: 1.57 correct answers for 9–10 years old to 2.30 for 14–16 years old (Cuzick's Wilcoxon-type tests for trends,  $p < 0.001$ ).

For adults, RPM test score significantly decreases with age: from 10.20 correct among 18–34 years old to 7.01 for 55–80 years old (Cuzick's Wilcoxon-type test for trends,  $p < 0.001$ ). The number of correct answers in the CRT test is constant across ages (Cuzick's Wilcoxon-type tests for trends,  $p = 0.41$ ).

The last column of Table 1 reports the correlation between the two measures of cognitive abilities. Overall, the correlations for children are positive and significant ( $p < 0.001$ ), across all ages (0.29 for boys and 0.39 for girls). However, the score in the RPM test and in the CRT test are uncorrelated for girls aged 9–10 years old, boys of 11–12 years old and all children aged 13–14 years old. In line with the previous literature, for adults the correlation between the RPM test and the CRT test are positive and significant across all ages, also within genders (0.30 for males and 0.32 for females). The magnitudes of the correlations are strikingly similar across the two samples: 0.32 for children and 0.30 for adults.

To understand the similarity of our RPM test and CRT test, we investigate the distribution of the two tests, and ask whether being above average in one of the tests also implies being above average in the other test. Fig. 1 shows the distribution of performance on the CRT test and on the RPM test separately for each age group. We divide the graphs into four quadrants by using the means of the RPM test and the CRT test. We report the percentage of students in each quadrant. Fig. 1 shows that more subjects are either above or below on both tests as opposed to being above on one and below on the other. For example, for children aged 9–10 years old, 59.52 % have either high RPM test's score and high CRT test's score or low RPM test's score and low CRT test's score. This leaves a large share, 40.48 %, as above the mean on the RPM test and below the mean on the CRT test, or vice versa. For adults of 55–80 years old, 64.79 % have either high RPM test's score and high CRT test's score or low RPM test's score and low CRT test's score. Again, a large part of subjects, 35.21 %, are above the mean on the RPM test and below the mean on the CRT test, or vice versa. Across all ages, we see the pattern that the larger share of subjects, between 52.86 %–64.79 %, have either high RPM test's score and high CRT test's score or low RPM test's score and low CRT test's score. This leaves a good share of subjects, 35.21 %–47.14 % as having either high RPM test's score and high CRT test's score or low RPM test's score and low CRT test's score. With this result, we find it important to investigate whether the two measures have separate explanatory power as we proceed to use the RPM test and the CRT test to explain behavioral inconsistency.

### 3.2. Cognitive abilities and behavioral inconsistency

Table 1 shows that RPM test's score and CRT test's score are positively correlated, but how do these measures relate to behavioral inconsistency? We test whether these two measures of cognitive abilities similarly predict time and risk inconsistency in children, and risk inconsistency in adults.

Analyzing inconsistency is relevant as inconsistency is widespread in our sample: even for adults who are more consistent than children, the percentage of inconsistent choices is 14.62 %. In children, this percentage increases up to 19.97 % for inconsistency in time preferences and 23.62 % for inconsistency in risk preferences.

To investigate and compare the impact of the different measures of cognitive abilities on inconsistency, first, we estimate the impact of each of them separately. As we are interested in the size of the association, we perform the analysis using the incentivized preferences and using the standardized measures for both the RPM test score and the CRT score. Since for the children we also have non-incentivized preferences, we also perform the analysis on those. Second, we estimate the joint association of RPM test and CRT test with behavioral inconsistency and test whether including both measures increases predictive power.

The top and middle panels of Fig. 2 show point estimates and confidence intervals from a regression of risk- and time-inconsistency on RPM test, CRT test, gender, and grade using school fixed effects and standard errors clustered at the class level for the children's data. We use incentivized preferences as outcomes. The bottom panel of Fig. 2 shows point estimates and confidence intervals from a regression of risk-inconsistency on RPM test, CRT test, gender, and age for the adults' data. Although we use standardized measures of cognitive abilities to be able to compare the impact of the RPM test with that of CRT test, the results are robust if we use the original scales and, for children, if we use non-incentivized preferences. The figure shows that RPM test and CRT test have a similar contribution to explaining behavioral inconsistency (RPM test's coefficient on time inconsistency for children:  $-0.047$ ;  $p = 0.014$ ; CRT test's coefficient on time inconsistency for children:  $-0.049$ ;  $p = 0.018$ ; RPM test's coefficient on risk inconsistency for children:  $-0.061$ ;  $p$ -value:  $0.025$ ; CRT test's coefficient on risk inconsistency for children:  $-0.036$ ;  $p = 0.020$ ; RPM test's coefficient on risk inconsistency for adults:  $-0.031$ ;  $p = 0.008$ ; CRT test's coefficient on risk inconsistency for adults:  $-0.047$ ;  $p = 0.007$ ).<sup>13</sup>

Regression estimates of the main results are reported in Tables 2–4. In Table 2, we report the results of the impact of cognitive ability on time inconsistency in children; in Table 3 we report the results of the impact of cognitive ability on risk inconsistency in children; and in Table 4 we report the results of the impact of cognitive ability on risk inconsistency in adults. In Appendix A, we test the robustness of the results using the original scale for the RPM test score (Tables A1–A3), and using risk and time inconsistency measured through non-incentivized preferences collected in the children dataset (Tables A4 and A5). The conclusions do not change.

Inconsistencies are quite prevalent in MPL tasks, but it is conceivable that inconsistent behavior in MPL are related to inconsistent and inattentive behavior in other tasks. To test this, we correlate MPL inconsistency with three other measures of inconsistency and inattention available in the adults' data. First, we consider a misperception test (see Fosgaard, Hansen and Wengström, 2017, for

<sup>13</sup> Cognitive ability is one of the factors that contributes to explain behavioral inconsistency (see Section 3.3 and, specifically, the  $R^2$  reported in Table 2–4).

**Table 1**  
Correlation between RPM test and CRT test.

		Average RPM score	Average CRT score	Correlation
9–10 years old (grade 3)	Male (n=84)	7.45	1.80	0.48***
	Female (n=84)	7.95	1.35	0.14
	Total (n=168)	7.70	1.57	0.24**
11–12 years old (grade 5)	Male (n=132)	8.78	2.06	0.13
	Female (n=109)	9.05	1.80	0.37***
	Total (n=241)	8.90	1.94	0.23***
13–14 years old (grade 7)	Male (n=76)	9.42	2.41	-0.03
	Female (n=81)	9.94	2.23	0.15
	Total (n=157)	9.69	2.32	0.05
14–16 years old (grade 8–9)	Male (n=66)	9.74	2.44	0.28*
	Female (n=54)	10.11	2.13	0.39**
	Total (n=120)	9.91	2.30	0.32***
Kids total	Male (n=358)	8.78	2.14	0.29***
	Female (n=328)	9.16	1.84	0.39***
	Total (n=686)	8.96	2.00	0.32***
18–34 years old	Male (n=238)	10.21	1.65	0.33***
	Female (n=242)	10.19	1.26	0.35***
	Total (n=480)	10.20	1.45	0.34***
35–44 years old	Male (n=261)	9.53	1.75	0.32***
	Female (n=267)	9.44	1.28	0.35***
	Total (n=528)	9.48	1.51	0.34***
45–54 years old	Male (n=298)	8.51	1.69	0.26***
	Female (n=316)	8.32	1.29	0.28***
	Total (n=614)	8.41	1.49	0.27***
55–80 years old	Male (n=405)	6.99	1.58	0.31***
	Female (n=305)	7.04	1.19	0.4***
	Total (n=710)	7.01	1.41	0.34***
Adults total	Male (n=1,202)	8.55	1.66	0.3***
	Female (n=1,130)	8.64	1.25	0.32***
	Total (n=2,332)	8.60	1.46	0.3***

Notes: Correlations are Spearman correlation coefficients. Stars denote whether the correlation between the CRT test and the RPM test is significant. Significance levels:

- \*  $p < 0.05$ ,
- \*\*  $p < 0.01$ ,
- \*\*\*  $p < 0.001$ .

details) that was collected after a public goods game. In short, this captures whether subjects understood how to maximize own monetary payoffs. Second, we consider cooperation profiles from a strategy-method version of the public good game, in which 80 % of subjects are classified as either conditional cooperators or free riders. We proxy inconsistency/inattention with being neither classified as a conditional cooperator nor a free rider. Third, we compare MPL inconsistencies across the two waves of the adults’ study. We observe that all three measures are significantly correlated with MPL inconsistencies in the first wave (Spearman rank correlation tests,  $p < 0.001$ ). That is, inconsistent decision making in the MPL is highly correlated with other measures of inconsistent and inattentive behaviors.

### 3.3. Substitutes or complements?

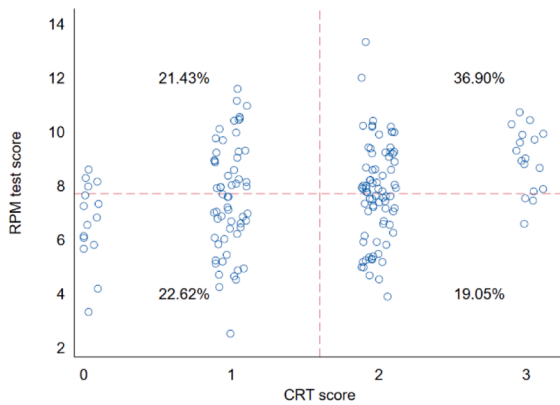
We then ask to what degree cognitive reflection and fluid intelligence are substitutes – in which case having one of the predictors is as good as having both – or they are complements – in which case having both predictors is better than having only one.

In Tables 2–4, we aim at testing how each measure contributes to explaining inconsistency using an incremental approach. Column (1) of each table reports the estimates and R-squared value of the regression with control variables (‘female’ and ‘grade’ for children and ‘female’ and ‘age’ for adults) but without any ability measures. In columns (2) and (3), we add respectively the RPM test and the CRT standardized measures, and in column (4) we add both, and in column (5) we report the marginal effects of a logit regression.<sup>14</sup>

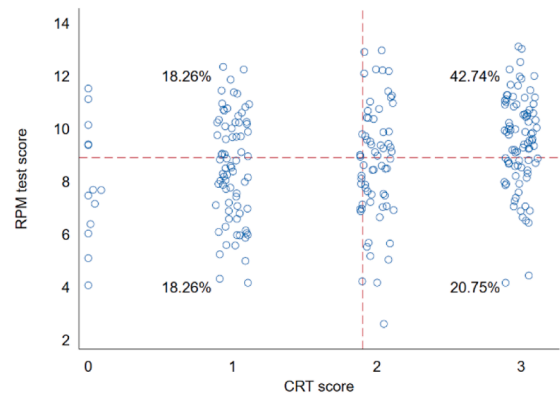
First, we observe that magnitude of the coefficients associated to RPM test and CRT test is very similar for time inconsistency (-0,059 and -0,060), while for risk inconsistency the RPM test has a higher coefficient for children (-0,069 vs -0,050) and a lower for adults (-0,046 vs -0,056).

Looking at the increase in R<sup>2</sup> value, we see that for time inconsistency in children it increases from 0.047 (column 1) to 0.065 (0.066) when adding RPM test (CRT test) to 0.077 when adding both. For risk inconsistency in children R-square increases from 0.065 (no measures of cognitive ability) to 0.087 (0.077) when adding RPM test (CRT test) to 0.092 when adding both. For risk inconsistency in adults R<sup>2</sup> increases from 0.028 (no measures of cognitive ability) to 0.042 (0.052) when adding RPM test (CRT test) to 0.058 when

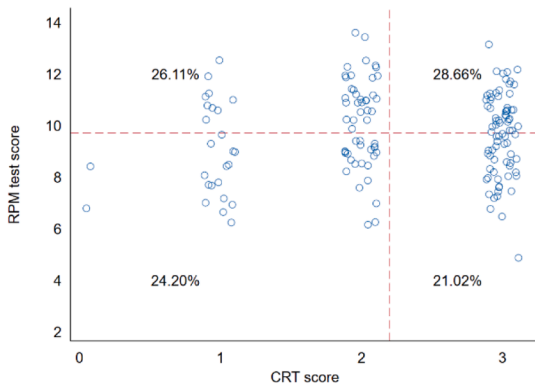
<sup>14</sup> The coefficients reported in columns (4)-(5) are those displayed in Fig. 2.



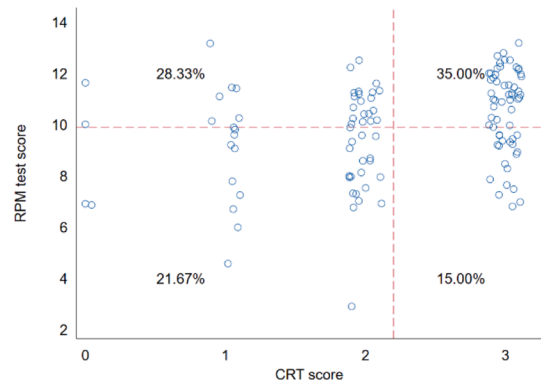
(a) 9-10 years old



(b) 11-12 years old



(c) 13-14 years old



(d) Grades 14-16 years old

**Fig. 1.** Distribution of CRT test and RPM test across ages, children (a-d) and adults (e-h). Notes: The red lines mark the average of the CRT test and RPM test, respectively. The blue dots represent the individual students and their performance on the RPM test and CRT test. Percentages are calculated as share of individuals in each quadrant, compared to all in that age range. Due to rounding errors, some graphs do not sum to 100.00%. N and mean values are shown in Table 1.

adding both. Comparing the increase in  $R^2$  for adding the first measure to both the results are ambiguous. Although we observe that including the two measures is better than having only one (and this is intuitive given that they capture different underlying cognitive processes), we still provide evidence that in presence of resources constraints researchers might want to collect and include in the model only CRT.

Since using only the  $R^2$  does not facilitate the choice between the models, we also investigate the best model for explaining inconsistency by using the Akaike Information Criteria (AIC) and the Bayesian Information Criteria (BIC) to penalize the inclusion of regressors<sup>15</sup>. While the results are ambiguous when looking at the BIC, the AIC favors the models with both RPM test and CRT test for both children and adults for both time inconsistency and risk inconsistency. Still, the marginal improvement is relatively small.

#### 4. Conclusions

This paper investigates the relationship between children’s and adults’ performances in a Raven’s Progressive Matrices test

<sup>15</sup> rs that contributes to explain behavioral inconsistency (see Section 3.3 and, specifically, the  $R^2$  reported in Table 2–4). The coefficients reported in columns (4)–(5) are those displayed in Fig. 2.

Measures of Cognitive Ability and Choice Inconsistency

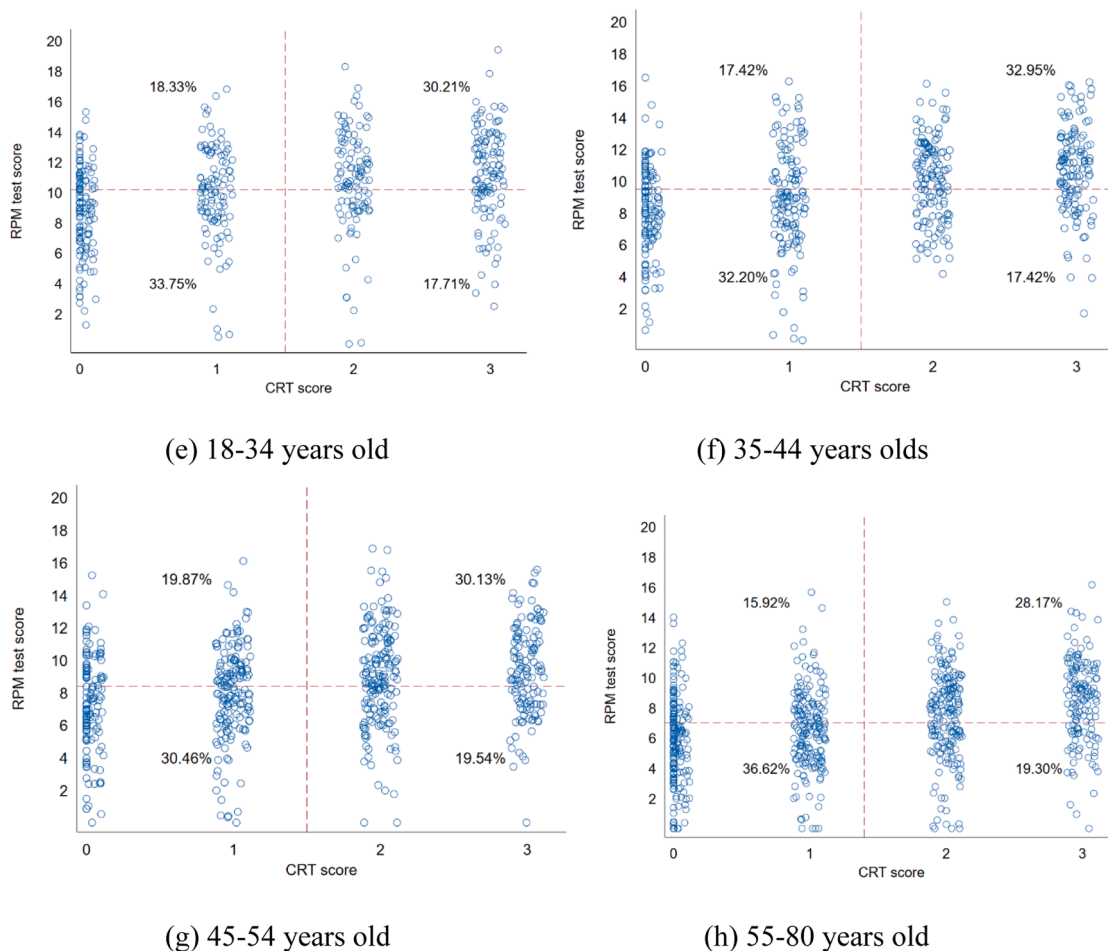


Fig. 1. (continued).

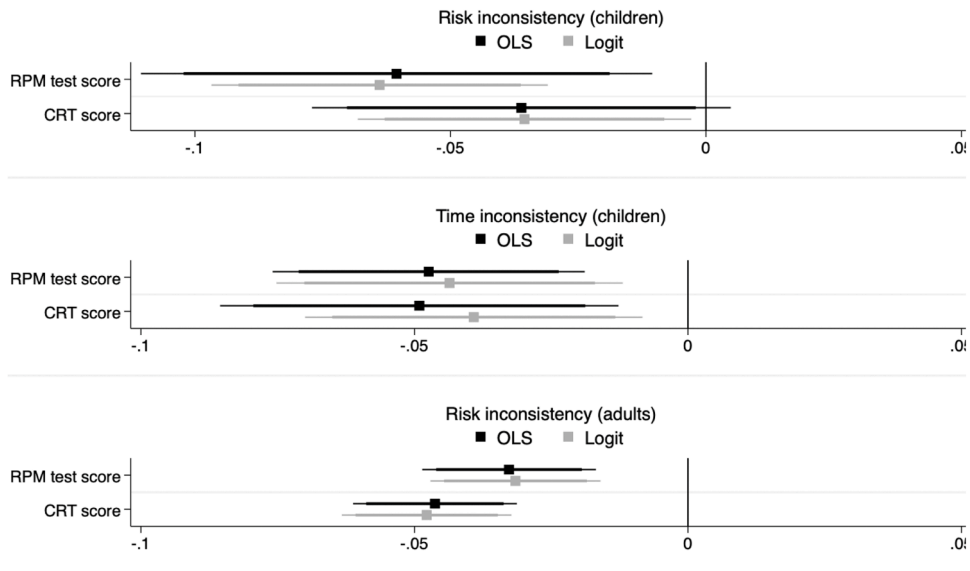
consisting of progressive matrices, and a Cognitive Reflection test. We use two samples covering the age-span from 9 to 80 years old. Even though the RPM test and CRT test tasks measure distinct traits of intelligence, logical thinking, and usage of system 2 thinking, we find that the correlation between the two measures is positive and significant (0.3–0.32) among the total sample – with exceptions in some age groups among children (most notably for children in grade 7). In addition, we test their interchangeability by studying how effective they are in predicting behavioral inconsistency in time and risk preferences. This investigation confirms the results obtained in the analysis of the correlation between the two measures, again this result is robust across the entire age span and both independent samples.

A clear advantage of the CRT is the shortness of the task. However, a disadvantage of the task is that these questions are sometimes posted and shared on the internet or in other experiments (Otero, Salgado and Moscoso, 2022). Hence, some children and adults might know the answers and do not have to engage in system 2 thinking to answer. When using the CRT test, the researchers must, therefore, continually update the questions. For adults, we consider this to be even more critical. RPM tests are rarely shared, so there are few problems of knowing the answers beforehand.

Considering our findings, although we acknowledge some limitations and we do not consider the two measures to be perfect substitutes, we conclude that when researchers have limited resources, they can elicit cognitive abilities using only one measure if cognitive abilities are not the center of focus.

**Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.



**Fig. 2.** The impact of cognitive abilities on behavioral inconsistency.

Notes: The figure shows parameter estimates and 95 %-confidence intervals from a regression of time and risk-inconsistency (risk only for adults) on the standardized measures from RPM test and CRT test using OLS (black line) and marginal effects estimated by a Logit model (grey line). In the OLS regression with children’s data we control for gender, grade school fixed effects and standard errors clustered at the class-level; for the OLS regression with adult data we control for gender, and age. To compare the magnitudes of the estimates of the RPM test and the CRT test, we standardized the two measures to obtain mean=0 and SD=1.

**Table 2**

The relationship between time inconsistency and cognitive ability, children’s data.

	Time inconsistency (=1)				
	(1) OLS	(2) OLS	(3) OLS	(4) OLS	(5) Logit
RPM Test		-0.059*** (0.016)		-0.047*** (0.014)	-0.044** (0.016)
Cognitive Reflection Test (CRT)			-0.060*** (0.018)	-0.049*** (0.018)	-0.039* (0.016)
Female	0.134*** (0.030)	0.146*** (0.030)	0.115*** (0.028)	0.128*** (0.029)	0.124*** (0.030)
Grade	-0.028** (0.010)	-0.015 (0.012)	-0.020* (0.011)	-0.011 (0.012)	-0.004 (0.008)
Constant	0.292*** (0.068)	0.216*** (0.077)	0.258*** (0.071)	0.203** (0.079)	-1.745*** (0.340)
Observations	686	686	686	686	686
R-squared	0.047	0.065	0.066	0.077	0.063
Sample	Children	Children	Children	Children	Children
School FE	Yes	Yes	Yes	Yes	
BIC	675.2	669	667.9	666.6	675.5
AIC	661.6	650.9	649.8	644	652.9

Notes: Linear Probability Model with school fixed effects. Robust standard errors clustered at the class-level (42 clusters) in parentheses. Using paired bootstrapped standard errors does not affect the conclusions. See Table A6 in the Appendix. Significance levels:

\*\*\*  $p < 0.001$ ,

\*\*  $p < 0.01$ ,

\*  $p < 0.05$

**Table 3**  
The relationship between risk inconsistency and cognitive ability, children’s data.

	Risk inconsistency (=1)				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	OLS	Logit
RPM Test		-0.069*** (0.023)		-0.061** (0.025)	-0.064*** (0.017)
Cognitive Reflection Test (CRT)			-0.050** (0.019)	-0.036* (0.020)	-0.036* (0.017)
Female	0.080* (0.033)	0.094*** (0.034)	0.064* (0.035)	0.081** (0.037)	0.082** (0.032)
Grade	-0.041*** (0.011)	-0.026** (0.012)	-0.035*** (0.011)	-0.024* (0.012)	-0.020* (0.008)
Constant	0.430*** (0.068)	0.341*** (0.074)	0.401*** (0.066)	0.331*** (0.072)	-0.848** (0.313)
Observations	686	686	686	686	686
R-squared	0.065	0.087	0.077	0.092	0.073
Sample	Children	Children	Children	Children	Children
School FE	Yes	Yes	Yes	Yes	
BIC	745.1	735.9	743	738.1	727.9
AIC	731.5	717.8	724.8	715.4	705.2

Notes: Linear Probability Model with school fixed effects. Robust standard errors clustered at the class-level (42 clusters) in parentheses. Using paired bootstrapped standard errors does not affect the conclusions. See Table A6 in the Appendix. Significance levels:

- \*\*\*  $p < 0.001$ ,
- \*\*  $p < 0.01$ ,
- \*  $p < 0.05$

**Table 4**  
The relationship between risk inconsistency and cognitive ability, adults’ data.

	Risk inconsistency (=1)				
	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	OLS	Logit
RPM Test		-0.046*** (0.008)		-0.031*** (0.008)	-0.031*** (0.008)
Cognitive Reflection Test (CRT)			-0.056*** (0.007)	-0.047*** (0.007)	-0.048*** (0.008)
Female	0.056*** (0.015)	0.055*** (0.014)	0.035** (0.015)	0.038*** (0.014)	0.037** (0.015)
Age	0.004*** (0.001)	0.002*** (0.001)	0.004*** (0.001)	0.003*** (0.001)	0.036*** (0.007)
Constant	-0.056* (0.026)	0.004 (0.028)	-0.040 (0.025)	-0.003 (0.027)	-2.597*** (0.152)
Observations	2,333	2,333	2,333	2,333	2,333
R-squared	0.028	0.042	0.052	0.058	0.070
Sample	Adults	Adults	Adults	Adults	Adults
BIC	1723	1696	1672	1666	1842
AIC	1706	1673	1649	1637	1814

Notes: Linear Probability Model. Robust standard errors in parentheses. Significance levels:

- \*\*\*  $p < 0.001$ ,
- \*\*  $p < 0.01$ ,
- \*  $p < 0.05$ .

**Data availability**

The authors do not have permission to share data.

**PsycINFO classification categories and codes**

2340, 2346

**Supplementary materials**

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.jebo.2024.02.029](https://doi.org/10.1016/j.jebo.2024.02.029).

## References

- Alós-Ferrer, C., Garagnani, M., Hügelschäfer, S., 2016. Cognitive reflection, decision biases, and response times. *Front. Psychol.* 7, 1402.
- Alós-Ferrer, C., Garagnani, M., 2021. Choice consistency and strength of preference. *Econ. Lett.* 198, 109672.
- Amador-Hidalgo, L., Brañas-Garza, P., Espín, A.M., García-Muñoz, T., Hernández-Román, A., 2021. Cognitive abilities and risk-taking: Errors, not preferences. *Eur. Econ. Rev.* 134, 103694.
- Andersson, O., Holm, H.J., Tyran, J.R., Wengström, E., 2016. Risk aversion relates to cognitive ability: preferences or noise? *J. Eur. Econ. Assoc.* 14 (5), 1129–1154.
- Andersson, O., Holm, H.J., Tyran, J.R., Wengström, E., 2020. Robust inference in risk elicitation tasks. *J. Risk Uncertain.* 61 (3), 195–209.
- Andreoni, J., Di Girolamo, A., List, J.A., Mackevicius, C., Samek, A., 2020. Risk preferences of children and adolescents in relation to gender, cognitive skills, soft skills, and executive functions. *J. Econ. Behav. Organ.* 179, 729–742.
- Arthur Jr, W., Tubre, T.C., Paul, D.S., Sanchez-Ku, M.L., 1999. College-sample psychometric and normative data on a short form of the Raven advanced progressive matrices test. *J. Psychoeduc. Assessm.* 17 (4), 354–361.
- Bayer, R.C., Renou, L., 2016. Logical abilities and behavior in strategic-form games. *J. Econ. Psychol.* 56, 39–59.
- Benjamin, D.J., Brown, S.A., Shapiro, J.M., 2013. Who is 'behavioral'? Cognitive ability and anomalous preferences. *J. Eur. Econ. Assoc.* 11 (6), 1231–1255.
- Bilker, W.B., Hansen, J.A., Brensinger, C.M., Richard, J., Gur, R.E., Gur, R.C., 2012. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment* 19 (3), 354–369.
- Bijwaard, G.E., van Kippersluis, H., Veenman, J., 2015. Education and health: the role of cognitive ability. *J. Health Econ.* 42, 29–43.
- Binswanger, H.P., 1980. Attitudes toward risk: Experimental measurement in rural India. *Am. J. Agricult. Econ.* 62 (3), 395–407.
- Bonan, J., Burlacu, S., Galliera, A., 2021. Consistency of prosocial behavior and cognitive skills: Evidence from children in El Salvador. Available at SSRN 3975429.
- Booth, R.W., Mackintosh, B., Mobini, S., Oztop, P., Nunn, S., 2014. Cognitive bias modification of attention is less effective under working memory load. *Cognit. Ther. Res.* 38 (6), 634–639.
- Brañas-Garza, P., Rustichini, A., 2011. Organizing effects of testosterone and economic behavior: Not just risk taking. *PLoS one* 6 (12), e29842.
- Brañas-Garza, P., García-Muñoz, T., González, R.H., 2012. Cognitive effort in the beauty contest game. *J. Econ. Behav. Organ.* 83 (2), 254–260.
- Bügelmayer, E., Spieß, C.K., 2014. Spite and cognitive skills in preschoolers. *J. Econ. Psychol.* 45, 154–167.
- Burks, S.V., Carpenter, J.P., Goette, L., Rustichini, A., 2009. Cognitive skills affect economic preferences, strategic behavior, and job attachment. *Proceed. Natl. Acad. Sci.* 106 (19), 7745–7750.
- Burks, S.V., Lewis, C., Kivi, P.A., Wiener, A., Anderson, J.E., Götte, L., Rustichini, A., 2015. Cognitive skills, personality, and economic preferences in collegiate success. *J. Econ. Behav. Organ.* 115, 30–44.
- Campitelli, G., Labollita, M., 2010. Correlations of cognitive reflection with judgments and choices. *Judgm. Decis. Mak.* 5 (3), 182–191.
- Capraro, V., Peltola, N., 2018. Low cognitive reflection predicts honesty for men but not for women. *arXiv preprint arXiv 1805.08316.*
- Castillo, M., Jordan, J.L., Petrie, R., 2018. Discount rates of children and high school graduation. *Econ. J.* 129 (619), 1153–1181.
- Caviglia-Harris, J., Maier, K., 2020. 'It's not all in their heads: The differing role of cognitive factors and non-cognitive traits in undergraduate success. *Educ. Econ.* 28 (3), 245–262.
- Cokely, E.T., Kelley, C.M., 2009. Cognitive abilities and superior decision making under risk: A protocol analysis and process model evaluation. *Judgm. Decis. Mak.* 4 (1), 20–33. February 2009.
- Corgnet, B., Espín, A.M., Hernán-González, R., 2016. Creativity and cognitive skills among millennials: Thinking too much and creating too little. *Front. Psychol.* 7, 1626.
- Court, J.H., 1983. Sex differences in performance on Raven's progressive matrices: a review. *Alberta J. Educ. Res.* 29 (1), 54–74.
- Cueva, C., Roberts, R.E., Spencer, T., Rani, N., Tempest, M., Tobler, P.N., Rustichini, A., 2015. Cortisol and testosterone increase financial risk taking and may destabilize markets. *Sci. Rep.* 5 (1), 1–16.
- Dave, C., Eckel, C.C., Johnson, C.A., Rojas, C., 2010. Eliciting risk preferences: when is simple better? *J. Risk Uncertain.* 41 (3), 219–243.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2010. Are risk aversion and impatience related to cognitive ability? *Am. Econ. Rev.* 100 (3), 1238–1260.
- Dohmen, T., Falk, A., Huffman, D., Sunde, U., 2018. On the relationship between cognitive ability and risk preference. *J. Econ. Perspect.* 32 (2), 115–134.
- Drouvelis, M., Lohse, J., 2020. Cognitive abilities and risk taking: the role of preferences. Department of Economics, University of Birmingham. Discussion Papers 20-02.
- Fe, E., Gill, D., Prowse, V.L., 2022. Cognitive skills, strategic sophistication, and life outcomes. *J. Politi. Econ.* forthcoming.
- Fosgaard, T.R., Hansen, L.G., Wengström, E., 2017. Framing and misperception in public good experiments. *Scand. J. Econ.* 119 (2), 435–456.
- Frederick, S., 2005. Cognitive reflection and decision making. *Journal of Economic Perspectives* 19 (4), 25–42.
- Gignac, G.E., 2015. Raven's is not a pure measure of general intelligence: Implications for g factor theory and the brief measurement of g. *Intelligence* 52, 71–79.
- Gill, D., Prowse, V., 2016. Cognitive ability, character skills, and learning to play equilibrium: a level-k analysis. *J. Politi. Econ.* 124 (6), 1619–1676.
- Hertzog, C., Smith, R.M., Ariel, R., 2018. Does the Cognitive Reflection Test actually capture heuristic versus analytic reasoning styles in older adults? *Exper. Aging Res.* 44 (1), 18–34.
- Kaplan, R.M., Saccuzzo, D.P., 2009. Standardized tests in education, civil service, and the military. *Psycholog. Test.: Principl. Applic. Issue.* 7, 325–327.
- Kazlauskaitė, V., Lynn, R., 2002. Two-year test-retest reliability of the colored progressive matrices. *Perceptual Motor Skill.* 95, 354–354.
- Meisenberg, G., 2012. National IQ and economic outcomes. *Personal. Individ. Differen.* 53 (2), 103–107.
- Mischkowski, D., Glöckner, A., Lewisch, P., 2021. Information search, coherence effects, and their interplay in legal decision making. *J. Econ. Psychol.* 87, 102445.
- Morton, R., Tyran, J.R., Wengström, E., 2011. Income and ideology: how personality traits, cognitive abilities, and education shape political attitudes. Univ. of Copenhagen Dept. of Economics Discussion Paper (11-08).
- Muñoz-Murillo, M., Álvarez-Franco, P.B., Restrepo-Tobón, D.A., 2020. The role of cognitive abilities on financial literacy: New experimental evidence. *J. Behav. Exper. Econ.* 84, 101482.
- Nielsen, K., Rehbeck, J., 2022. When choices are mistakes. *Am. Econ. Rev.* Forthcoming.
- Otero, I., Salgado, J.F., Moscoso, S., 2022. Cognitive reflection, cognitive intelligence, and cognitive abilities: a meta-analysis. *Intelligence* 90, 101614.
- Pachur, T., Mata, R., Hertwig, R., 2017. Who dares, who errs? Disentangling cognitive and motivational roots of age differences in decisions under risk. *Psychology. Sci.* 28 (4), 504–518.
- Piovesan, M., Willadsen, H., 2021. Risk preferences and personality traits in children and adolescents. *J. Econ. Behav. Organ.* 186, 523–532.
- Raven, J.C., 1938. Raven standard progressive matrices. *Journal of cognition and development.*
- Raven, J.C., Court, J.H., Raven, J., 1995. *Coloured Progressive Matrices.* Oxford Psychologists Press, Oxford.
- Rustichini, A., DeYoung, C.G., Anderson, J.E., Burks, S.V., 2016. Toward the integration of personality theory and decision theory in explaining economic behavior: An experimental investigation. *J. Behav. Exper. Econ.* 64, 122–137.
- Sirota, M., Dewberry, C., Juanchich, M., Valuš, L., Marshall, A.C., 2021. Measuring cognitive reflection without maths: Development and validation of the verbal cognitive reflection test. *J. Behav. Decis. Mak.* 34 (3), 322–343.
- Stagnaro, M.N., Pennycook, G., Rand, D.G., 2018. Performance on the cognitive reflection test is stable across time. *Judgm. Decis. Mak.* 13 (3), 260–267.
- Stanovich, K.E., 2012. On the distinction between rationality and intelligence: Implications for understanding individual differences in reasoning. *Oxf. Handb. Think. Reason.* 343–365.
- Tanaka, T., Camerer, C.F., Nguyen, Q., 2010. Risk and time preferences: linking experimental and household survey data from Vietnam. *Am. Econ. Rev.* 100 (1), 557–571.
- Taylor, M.P., 2013. Bias and brains: risk aversion and cognitive ability across real and hypothetical settings. *J. Risk Uncertain.* 46 (3), 299–320.
- Tymula, A., Rosenberg Belmaker, L.A., Roy, A.K., Ruderman, L., Manson, K., Glimcher, P.W., Levy, I., 2012. Adolescents' risk-taking behavior is driven by tolerance to ambiguity. *Proceed. Natl. Acad. Sci.* 109 (42), 17135–17140.

- 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 Cogniti.* 39 (7), 1275–1289.
- Urbina, S., 2011. Tests of intelligence. *The Cambridge Handbook of Intelligence*, pp. 20–38.
- Young, A.G., Powers, A., Pilgrim, L., Shtulman, A., 2018. Developing a cognitive reflection test for school-age children. *CogSci*.
- Young, A.G., Shtulman, A., 2020. How children’s cognitive reflection shapes their science understanding. *Front. Psychol.* 11, 1247.