

# Applying Item Response Theory to Develop a Shortened Version of the Need for Cognition Scale

Francesca Chiesi<sup>1</sup>, Kinga Morsanyi<sup>2</sup>, Maria Anna Donati<sup>1</sup>, and Caterina Primi<sup>1</sup>

<sup>1</sup> Department of Neuroscience, Psychology, Drug, and Child's Health (NEUROFARBA),  
Section of Psychology, University of Florence, Italy

<sup>2</sup> School of Psychology, Queen's University, Belfast, UK

## ABSTRACT

## KEYWORDS

assessment, item response theory, need for cognition, short form, validity

The 18-item Need for Cognition Scale (NFC-18) is the most commonly used tool to measure the need for cognition. The aim of this study was to explore the possibility of developing an abbreviated version of the scale, applying the item response theory (IRT). Item response theory analyses suggested the exclusion of eight items that did not perform well in measuring the latent trait. The resulting 10-item scale (NFC-10), which included highly discriminative items, covered the same range of the measured trait as the original scale and showed high measurement precision along various levels of the trait. Additionally, since IRT analyses can only confirm the accuracy of the short scale in measuring the underlying construct, we sought to replicate the nomological net of the NFC-18 using the shortened version of the scale. The results showed that the NFC-10 reflects an adequate operationalization of the construct, in line with the longer version. In particular, as expected, the NFC-10 showed moderate relations with various measures of cognitive skills and self-report measures of cognitive styles, confidence, and anxiety. These findings confirm that we have obtained a much shorter version of the NFC that maintains excellent reliability and validity

## INTRODUCTION

Need for cognition (NFC) is a stable individual characteristic defined as a tendency to engage in and enjoy effortful cognitive activities (Cacioppo & Petty, 1982). Interindividual variation in NFC has been conceptualized as falling along a continuum from low to high. Both individuals low in NFC and individuals high in NFC must make sense of the world, but they tend to derive meaning, adopt positions, and solve problems in somewhat different ways. Individuals high in NFC naturally tend to seek, acquire, think about, and reflect on information, and they engage in and enjoy effortful cognitive activity. Individuals with low NFC, by contrast, are more likely to rely on others (e.g., experts and instruments), they avoid detailed information about the world and find cognitively effortful problems or tasks stressful (Cacioppo & Petty, 1982). They also tend to adopt cognitive heuristics. In other words, individuals with low NFC are characterized by low intrinsic motivation to engage in effortful cognitive endeavors, whereas individuals with high NFC are characterized by high intrinsic motivation to exercise their mental faculties (Cacioppo, Petty, & Jarvis, 1996).

Need for cognition is a thinking disposition, similar to open-minded thinking, need for closure, reflectivity, superstitious thinking, and dogmatism (Cacioppo et al., 1996; Kruglanski & Webster, 1996; Liberali, Reyna, Furlan, Stein, & Pardo, 2012; Stanovich, 2011; Stanovich & West, 2007; Strathman, Gleicher, Boninger, & Edwards, 1994), which are important aspects of what is called the *reflective mind* (Sinatra & Pintrich, 2003; Stanovich, 2009, 2011; Sternberg, 2003). For example, people who are more prone to think hard and invest cognitive effort in the reasoning process are more likely to recognise the need to apply the correct rules and less inclined to use tempting, but incorrect heuristics (e.g., Handley, Capon, Beveridge, Dennis, & Evans, 2004; Kokis, Macpherson, Toplak, West, & Stanovich, 2002; Stanovich & West, 1999; however, see e.g., Morsanyi, Primi, Chiesi & Handley,

Corresponding author: Francesca Chiesi, Department of Neuroscience, Psychology, Drug, and Child's Health (NEUROFARBA), Section of Psychology, via di San Salvi 12- Padiglione 26, 50135, Firenze (Italy). Tel.: 055-2755054. E-mail: francesca.chiesi@unifi.it

2009). Moreover, differences in NFC could lead to different ways to process and interpret information and, consequently, different ways to choose and pursue individual goals. Specifically, higher NFC is linked to a higher number of goals that require reasoning and problem solving (Gollwitzer, Kappes, & Oettingen, 2012). For instance, Wu, Parker, and De Jong (2014) proposed that individuals with high NFC are more likely to engage in innovative processes that require dealing with complex situations, investing effort, and developing new strategies and solutions in the workplace.

To measure NFC, Cacioppo and Petty (1982) developed a 34-item inventory, but the most commonly used version includes 18 items from the original scale (the Need for Cognition Scale, NFC-18; Cacioppo, Petty, & Kao, 1984), which are rated on a 5-point Likert scale. The items describe a variety of situations in which people could choose to gather information, analyze available evidence, abstract from past experience, or synthesize ideas; items dealing with potentially noneffortful cognitive activities such as reveries, mystical or religious experiences, daydreaming, and artistic ruminations were intentionally excluded. The items were also worded to avoid responses limited to particular domains, problems, or situations and known groups, and cross-validations were used to select items. Thus, the scale was designed to distinguish between individuals varying along a continuum ranging from the extreme cognitive miser to the supreme cognizer (Cacioppo & Petty, 1982; Cacioppo et al., 1984).

The NFC-18 is characterized by one dominant factor (Cacioppo et al., 1984; Hevey et al., 2012; Sadowski, 1993; Soubelet & Salthouse, 2017; Verplanken, Hazenberg, & Palenew, 1992), high internal consistency (e.g., Cacioppo et al., 1984; Sadowski, 1993; Soubelet & Salthouse, 2017; Woo, Hans, & Kuncel, 2007), and good test-retest reliability (Sadowski & Gulgoz, 1992; Soubelet & Salthouse, 2017; Verplanken, 1991). Moreover, evidence of its validity was provided by testing the relationships between NFC scores and the tendency to seek out, scrutinize, and use relevant information when making decisions and solving problems (e.g., Berzonsky & Sullivan, 1992; Petty & Jarvis, 1996), the desire for new experiences that stimulate thinking (Venkatraman, Marlino, Kardes, & Sklar, 1990; Venkatraman & Price, 1990), and the tendency to ignore, avoid, or distort information (Petty & Jarvis, 1996; Webster & Kruglanski, 1994; Venkatraman et al., 1990).

Although the NFC-18 is a popular measure and it has good validity and reliability, it might be possible to measure NFC more quickly using a psychometrically sound shortened version. Indeed, with research questions becoming increasingly complex and involving a growing number of constructs, shorter scales potentially offer added value (Ziegler, Kemper, & Kruyen, 2014). Starting from these considerations, the aim of this study was to explore the possibility of developing an abbreviated version of the NFC-18 for research purposes. Specifically, the aim was to obtain a brief measure that could be more appropriate for large, multivariate studies in which many tests and scales need to be administered together.

To achieve this, we used item response theory (IRT) analyses, which make it possible to select items that offer the most information in measuring the targeted underlying trait. Specifically, IRT has

potential benefits in shortening a scale because it makes it possible to evaluate the amount of information provided by each item of the scale for each trait level on the trait dimension through the item information function (IIF). Whereas the statistical and psychometric meaning of information has a technical gist, the meaning is intuitive: If the amount of information is large, the trait level can be estimated with precision, if the amount of information is small, the trait cannot be accurately estimated. Thus, on the basis of item information, it is possible to select items that convey the higher amount of information along the entire range of the measured trait. When a single item is involved, the amount of information at any point on the trait scale is going to be rather small. However, through the selection of items that perform better and assure adequate information along the different levels of the trait, a well-performing shortened scale can be obtained.

Additionally, IRT provides the test information function (TIF), which evaluates the precision of the test at different levels of the measured construct instead of providing a single value (e.g., Cronbach's  $\alpha$ ) for reliability (Embretson & Reise, 2000; Hambleton, Swaminathan, & Rogers, 1991). More precisely, the TIF provides information on how accurate the test is at estimating a trait along the whole range of trait scores: The more information the test provides at a particular trait level, the smaller the error associated with ability estimation, and the higher the local reliability. Since the TIF is generated by aggregating the IIFs, in general, longer tests will measure an examinee's attribute with greater precision than shorter tests. Nonetheless, in the IRT framework, item selection can be done ensuring that the TIF of the shortened scale maintains an adequate amount of information along the trait continuum, which is similar to the original scale.

Finally, IRT allows the assessment of differential item functioning (DIF; Embretson & Reise, 2000). Differential item functioning analysis is used to study the performance of items within scales, and it examines whether or not the likelihood of endorsing each item is equal across subgroups that are matched on the measured trait. For example, a randomly selected man with a certain level of NFC and a randomly selected woman with the same level of NFC should have the same likelihood of endorsing a particular response option for each item on the scale. Differential item functioning analysis can be used for tailoring the length of psychological scales because it makes it possible to identify and exclude items that have differential functioning and, as a consequence, produce biased measures.

In the current study, we assessed the psychometric properties of the items as well as the characteristics of the whole NFC-18 to select the items that conveyed the largest amount of information along different levels of the NFC trait. Moreover, we investigated gender DIF aiming both to shorten the scale and to confirm the psychometric soundness of the obtained shortened scale. Gender invariance was investigated because, although gender differences for the NFC scale have not been reported (see, Cacioppo et al., 1996 for a review), the NFC has been found to be correlated with some measures where gender differences are common (e.g., anxiety and school achievement – with females typically scoring higher on both). Finally, given that IRT analyses can only confirm the accuracy of the obtained shortened scale in measuring the

underlying construct, validity measures were administered to provide evidence that the abbreviated scale still measures NFC adequately. Specifically, Smith, McCarthy, and Anderson (2000) stated that a short form of a test should meet the same standards of validity as the full form. Thus, an important goal in the development of any test's short form should be to replicate the pattern of relationships established for the construct as measured by the long form of the test. Following these recommendations, we sought to replicate the nomological net of the NFC-18, employing the shortened version of the scale.

Regarding validity, as NFC is a thinking disposition that may be related to but not equivalent to cognitive abilities, we investigated its relationships with a measure of fluid intelligence, (e.g., Greco & Walter, 2013; Handley et al., 2004; Hill et al., 2013; Kokis et al., 2002). Second, in line with the theoretical and operational definition of the construct and previous studies regarding the validity of the NFC scale, its relationships with reasoning abilities were investigated, assuming that people with high NFC are likely to use relevant information when making decisions and solving problems (Liberati et al., 2012; Stanovich, 2011) and less likely to adopt heuristic shortcuts and to ignore or distort information (Venkatraman et al., 1990). For this reason, we hypothesized that the Cognitive Reflection Test (CRT, Frederick, 2005), which measures the tendency to avoid incorrect heuristic shortcuts and to rely on careful deliberation instead, should be positively correlated with NFC.

We also investigated the correlations between NFC and the Probabilistic Reasoning Scale (PRS, Primi, Morsanyi, Donati, Galli, & Chiesi, 2017), as poor probabilistic reasoning is often the result of misconceptions (e.g., Fischbein, 1987) and previous studies (e.g., Clinton, Morsanyi, Alibali, & Nathan, 2016; Kokis et al., 2002; Morsanyi et al., 2009; West, Toplak, & Stanovich, 2008) found a positive relationship between NFC and probabilistic reasoning (although see Chiesi, Primi, & Morsanyi, 2011). Additionally, the PRS includes some arithmetic computations, and Dornic, Ekehammar, and Laaksonen (1991) found that people with high NFC reported that they found arithmetic tasks easier as compared to people with low NFC. According to Cacioppo et al. (1996), this is because people with high NFC are generally more practiced at performing familiar tasks, which leads to higher fluency and accuracy. Based on these premises, we expected that NFC would also be related to math fluency (i.e., the ability to correctly solve a large number of relatively simple arithmetic tasks within a short time frame).

In addition to measuring performance on cognitive tasks, we also investigated the relations between NFC and various self-report measures. Factor-analyses of longer versions of the scale have identified confidence as one of the underlying factors of NFC (Tanaka, Panter & Winborne, 1988; Waters & Zakrajsek, 1990). We have measured confidence and self-evaluated ability related to quantitative skills by administering the Subjective Numeracy Scale (Fagerlin et al., 2007). In addition to confidence, several studies investigated the relation between anxiety and NFC. Negative correlations have been reported between NFC and both state and trait anxiety (Olson, Camp, & Fuller, 1984), social anxiety (Mueller & Johnson 1990), math anxiety (Lin, Durbin & Rancer, 2016), anxiety about cognitive stressors (Cacioppo & Petty,

1984), and perceived stress in college undergraduates (Petty & Jarvis, 1996). Nevertheless, Cacioppo and Petty (1982) found no relation between NFC and test anxiety. In the current study, we administered a math anxiety scale to investigate its relation to NFC.

The last scale that we used for validation purposes was the Object-Spatial Imagery and Verbal Questionnaire (Blazhenkova & Kozhevnikov, 2009) which measures three types of cognitive style: verbal, object-visualizer, and spatial. Cognitive styles represent regularities in cognitive functioning, particularly in the acquisition and processing of information. Those who possess a verbal cognitive style mainly use verbal-analytical strategies when performing cognitive tasks. Verbalizers score higher on tests that require participants to formulate thoughts in a verbal format (Blazhenkova & Kozhevnikov, 2009) and on tasks that require the fluent retrieval of declarative knowledge (Morsanyi, O'Mahony, & McCormack, 2017). By contrast, spatial visualizers preferentially rely on imagery to represent and transform spatial relations. Whereas a verbal cognitive style is more common in humanities students, spatial thinking is more characteristic of science students, and it has been found to be associated with good performance on quantitative and mechanical reasoning tests (Blazhenkova & Kozhevnikov, 2009; Morsanyi et al., 2017). Finally, object visualizers use imagery to construct vibrant, clear images of individual objects, whereas they tend to score lower on tasks that require reasoning about structure and spatial relations (Hegarty & Kozhevnikov, 1999; Pitta-Pantazi, Sophocleous, & Christou, 2013). This thinking style has been found to be characteristic of fine arts students, and it is linked to some aspects of artistic production (Pérez-Fabello, Campos, & Campos-Juanatey, 2016). Although these thinking styles tend to be independent (Blazhenkova & Kozhevnikov, 2009), we expected that both verbal and spatial thinking styles should be positively related to NFC. That is, NFC should be related to high-level skills in cognitive activities, and the enjoyment of these, regardless of the nature of these activities (i.e., whether they predominantly require verbal or quantitative abilities).

Regarding the size of the correlations, in their review of the literature, Cacioppo et al. (1996) reported low to moderate correlations between NFC and other constructs, with low correlations for ability measures and moderate correlations for self-report scales (with most correlations below .35). The only exception, where stronger correlations were found, were self-report scales that measured constructs that are conceptually very close to NFC, such as academic curiosity ( $r = .68$ ), and information style orientation (i.e., a tendency to seek out and elaborate on self-relevant information,  $r = .50$ ). A recent study by Soubelet and Salthouse (2017) also reported low to moderate correlations (typically ranging from .1 to .3) between NFC and various other constructs, with the exception of a strong correlation between NFC and the Big Five personality dimension of openness ( $r = .60$ ). Given these earlier findings, we expected modest correlations between NFC and the various validity measures, and all these correlations were expected to be positive, with the exception of the relation between NFC and the math anxiety scale. Additionally, employing the shortened scale, we expected to replicate the same pattern of correlations as with the original scale. Indeed, the NFC-10 was developed excluding the

NFC-18 items that provided the smaller amount of information on the measured trait (i.e., the items that did not provide additional information when measuring NFC).

## METHOD

### Participants

The participants were 634 undergraduate students ( $M_{\text{age}} = 21.80$  years;  $SD = 5.45$ ; 318 females) attending various courses at three different universities in the South-East and South-West of England and in Northern Ireland. About 60% of the participants studied psychology (they were recruited from two different universities), about 30% studied medicine (they were all recruited from the same university), whereas the remaining participants were recruited from the third university by placing adverts around the university campus, and they attended various courses (mostly related to science and engineering). All students participated on a voluntary basis, and they received ungraded course credit for their participation. Ethical approval was obtained separately from the institutional ethics committees of each university where data collection took place. Participants also provided written informed consent before participating in the study.

### Materials

#### MEASURES OF COGNITIVE SKILLS

To measure fluid intelligence, A 12-item short form of the Raven's Advanced Progressive Matrices (APM; Arthur & Day, 1994) was administered. This short form has been shown to have adequate psychometric properties, and it is a valid and reliable instrument (Chiesi, Ciancaleoni, Galli, Morsanyi, & Primi, 2012). Three practice items from the Raven's Standard Progressive Matrices (Raven, 1962) were administered before the test items. Cronbach's  $\alpha$  in the current sample was .74.

The CRT (Frederick, 2005) or its extended version (the CRT-Long, Primi, Morsanyi, Chiesi, Donati, & Hamilton, 2016) was administered to all participants<sup>1</sup>. The original test includes three open-ended problems, and the extended version includes 3 additional tasks. The CRT measures the ability to resist intuitive response tendencies and to rely on effortful reasoning instead. An example item is the following: "If it takes 5 minutes for five machines to make five widgets, how long would it take for 100 machines to make 100 widgets?" The typical incorrect heuristic response that first comes to mind is "100 minutes", but the correct response is "5 minutes". The number of correct responses was summed to obtain a total score for both the CRT and the CRT-Long. In our sample, Cronbach's  $\alpha$  was .73 for the CRT, and .80 for the CRT-Long.

The PRS (Primi et al., 2017) is a 16-item multiple choice questionnaire that provides a comprehensive assessment of basic aspects of probabilistic reasoning, including basic and conditional probabilities presented in text and tables, reasoning about random sequences of events, and the ability to resist some typical fallacies and biases.

Participants have to select the correct response from three options. An example item is the following: "60% of the population in a city are men and 40% are women. 50% of the men and 30% of the women smoke. We select a person from the city at random. What is the probability that this person is a smoker?" Response options are (a) 42%, (b) 50%, and (c) 85% (The correct response is 42%). The number of correct responses was summed to obtain a total score. Cronbach's  $\alpha$  in the current sample was .66.

The Math Fluency (MF) subtest of the Woodcock-Johnson III Tests of Achievement (Woodcock, McGrew, & Mather, 2001) was used as a measure of arithmetic skills. This test assesses the ability to solve simple addition, subtraction and multiplication problems quickly (e.g.,  $5 + 3 =$ ;  $8 \times 6 =$ ; with the numbers presented vertically). Participants were asked to work through a series of problems as quickly and accurately as possible within a 3-minute time limit. The total number of items that were correctly solved was calculated.

#### SELF-REPORT MEASURES OF COGNITIVE STYLES AND ANXIETY

The 18-item version of the NFC-18 (Cacioppo et al., 1984) consists of nine positively worded items and nine negatively worded items. Respondents indicate the extent to which they agree with each statement on a 5-point Likert scale ranging from 1 (*extremely uncharacteristic*) to 5 (*extremely characteristic*). An example item is "I find satisfaction in deliberating hard and for long hours." Cronbach's  $\alpha$  for the current sample was .83.

The Subjective Numeracy Scale (SNS, Fagerlin et al., 2007) is a subjective measure of quantitative ability, which was developed with the aim of distinguishing between low-numerate and high-numerate individuals. An example item is "How good are you at working with fractions?" The items are rated on a 6-point Likert scale, which are labelled differently, depending on the question asked (e.g., ranging from 1, *not good at all*, to 6, *extremely good*; or 1, *never*, to 6, *very often*). A single composite score was obtained based on participants' ratings of each item. Cronbach's  $\alpha$  for the current sample was .74.

The Abbreviated Math Anxiety Scale (AMAS; Hopko, Mahadevan, Bare, & Hunt, 2003) is a short valid math anxiety scale with only 9 items, using a 5-point scale ranging from 1 (*low anxiety*) to 5 (*high anxiety*). An example item is "Thinking about an upcoming math test 1 day before." Cronbach's  $\alpha$  for the current sample was .88.

The 45-item Object-Spatial Imagery and Verbal Questionnaire (OSIVQ, Blazhenkova & Kozhevnikov, 2009) was administered to assess individual differences in the object-visualizer, spatial, and verbal cognitive styles. The following is an example of an item from the verbal subscale: "My verbal skills are excellent." An example item from the object imagery scale is "I have a photographic memory." An example of an item from the spatial imagery scale is "I can easily imagine and mentally rotate three-dimensional geometric figures." Participants rated their agreement with each item on a 5-point scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). Totals for each subscale were calculated. In the current sample, Cronbach's  $\alpha$  for the verbal scale was .76, for the object visualizer scale it was .70, and for the spatial scale it was .80.

## Procedure

The participants were recruited from various university courses at three different UK universities. Each group completed a slightly different set of measures, although all participants completed the 18-item NFC scale and either the original or the long version of the CRT. Specifically, one group of participants ( $n = 390$ ; 253 females) completed the NFC, the CRT-Long, the PRS, the SNS, and the AMAS. A subsample of this group ( $n = 58$ ; 49 females) also took part in an additional testing session, where they completed the Raven's APM, the MF test and the OSIVQ. Another group of students ( $n = 216$ ; 172 females) completed the NFC, the Raven's APM and the original CRT. Finally, a small group ( $n = 28$ ; 17 females) completed the NFC and the original CRT only. Each task was briefly introduced and instructions for completion were given. All answers were collected in a paper-and-pencil format.

## Analysis strategy

The IRT analyses were performed using IRTPRO 2.0 (Cai, Thissen, & du Toit, 2011) and, according to the NFC-18 response format, Samejima's (1969) graded response model (GRM) was used. The model makes three key assumptions about the data: (a) unidimensionality, (b) local independence, and (c) that the IRT model fits the data (Reeve & Fayes, 2005). In the current study, these assumptions were verified as follows.

The single factor structure of the NFC-18 was tested using confirmatory factor analysis (CFA) in order to demonstrate that the NFC-18 set of items measures a single latent construct ( $\theta$ ). After checking the distribution of the NFC-18 items for assessment of normality, that is, if skewness and kurtosis indices were between the values of -1 and 1 (Marcoulides & Hershberger, 1997), the CFA was conducted with AMOS 5.0 (Arbuckle, 2003) using maximum likelihood estimation on the variance-covariance matrix.

The absence of local dependence (LD) is important for the fit of unidimensional IRT modelling because the item parameter estimates reflect the latent trait adequately only if there is no association among item responses when  $\theta$  is held constant. The LD, that is, an excess of covariation among item responses that is not accounted for by a unidimensional IRT model, was assessed using the  $\chi^2$  LD statistic (Chen & Tiessen, 1997), computed by comparing the observed and the expected frequencies in each of the two-way cross tabulations between responses to each pair of items. This diagnostic statistic is approximately distributed as standardized  $\chi^2$ . Given this approximation, as a rule of thumb, values of 10 or greater indicate the presence of LD.

The model's goodness of fit was evaluated using the M2 statistic and the associated root mean square error of the approximation (RMSEA) value. As the M2 statistic, similar to other  $\chi^2$  statistics, is generally unrealistic because some error will be present in any strong parametric model, the RMSEA provides a metric for model errors (Cai, Maydeu-Olivares, Coffman, & Thissen, 2006). RMSEA values of .05 or less indicate a good fit.

Item response theory models use the original response data for estimating probabilities of responses as a function of the latent trait  $\theta$ , which is defined as a continuous variable that conventionally has a

mean of zero and *SD* of 1.0. This function describes the relation between the probability of endorsing a response given not only the respondent's level of  $\theta$  but also the item characteristics. The item characteristics in the GRM model are estimated by the discrimination ( $a$ ) and location ( $b_i$ ) parameters. The first parameter describes the ability of an item to discriminate among people with different levels of the underlying trait (e.g., the higher  $a$  is, the higher the item's ability to differentiate between people with different levels of NFC). The  $b$  parameters represent an item's sensitivity in differentiating among the various levels of the target trait (i.e., if  $b$ s are evenly spaced along the trait, the item categories provide a better differentiation in measuring NFC). The IIF, graphically represented by the item information curve (IIC), describes the amount of information that a particular item provides across the entire continuum of the latent construct, and it depends on both the discrimination and location parameters. Thus, we used IIFs to select the items that conveyed the higher amount of information along the range of the trait measured by the NFC-18, looking at the area above the IICs, which equals both the size of the  $a$  parameter and the spread of the  $b$  parameters.

To shorten the scale, we also performed gender DIF applying the IRT likelihood ratio test approach (IRTLR; Thissen, Steinberg, & Wainer, 1988) implemented in the IRTPRO software (Cai et al., 2011). This procedure involves comparing differences in log-likelihoods (distributed as  $\chi^2$ ) associated with nested models. Since DIF analyses examine differences in item parameters, for the GRM model two types of DIF can be detected: *uniform* DIF for the location parameters and *nonuniform* DIF for the discrimination parameter.

Once the shortened scale was defined, all the above described analyses were repeated for the brief scale in order to confirm the item and test psychometric properties. In particular, we investigated the precision of the shortened scale as a whole comparing its TIF with the one of the original NFC scale. The TIF is generated by aggregating the IIFs of items in a single measure and it allows to compute the information (I), that is, the expected value of the inverse of the standard error (*SE*), provided by the test at each level of the trait. Thus, the more information the test provides at a particular trait level, the smaller the error associated with trait estimation. Test information can be related to traditional reliability coefficients. For short scales, it has been suggested to use McDonald's  $\omega$  instead of Cronbach's  $\alpha$  (Schipolowski, Schroeders, & Wilhelm, 2014). Thus, we converted I in  $\omega$  applying the formula [ $= I/(I+1)$ ], proposed by McDonald (2013).

Regarding validity, the Pearson product-moment correlations were computed and compared using Steiger's (1980)  $z$  tests in order to identify substantive differences in the correlations observed for the two different scales.

## RESULTS

Preliminarily, an examination of the item distributions revealed that they did not deviate significantly from normality: The absolute value of skewness indices ranged from 0.08 to 0.96, and kurtosis indices ranged from 0.01 to 0.98. Then, the CFA showed that the NFC set of

items measured one dimension (CFI = .91, TLI = .90, and RMSEA = .053 [C.I. = .046-.060]). All factor loadings were significant ( $p < .001$ ), ranging from .31 to .70. None of the LD statistics were greater than 10, attesting that there was not an excess of covariation among item responses when  $\theta$  was held constant. The fit statistics of the IRT model indicated an adequate fit for the GRM ( $M_2 = 4207.09$ ,  $df = 2430$ ,  $p < .001$ ; RMSEA = .03).

Having verified the preliminary assumptions for IRT modelling, we looked at the IICs to select the items that conveyed the higher amount of information along the range of the trait measured by the NFC-18 (see Figure 1). The figure clearly shows that items 8, 9, 16, and 18 can be excluded because they convey a very small amount of information. Moreover, since we aimed to select the better performing items, items 7, 12, 14, and 17 were also candidates for removal. Finally, the gender DIF analysis revealed that seventeen out of eighteen items did not show differential functioning across male and female respondents. Only item 7 showed uniform DIF ( $p = .047$ ) and this finding provided an additional reason for its exclusion.

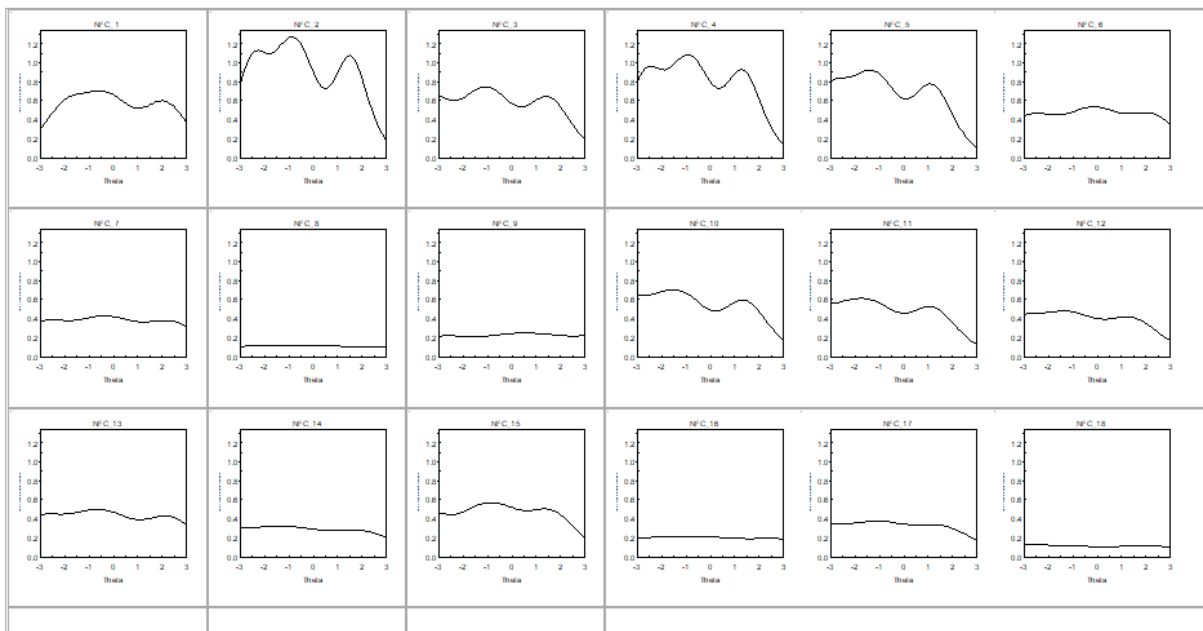
As a result, we retained 10 items and we repeated the analyses for this shortened version of the NFC scale (NFC-10). The CFA confirmed the one-factor structure of the NFC-10 (CFI = .95, TLI = .93, and RMSEA = .062 [C.I. = .049-.074]). All factor loadings were significant ( $p < .001$ ), ranging from .53 to .72. None of the LD statistics were greater than 10, indicating the absence of LD. The fit for the GRM was also adequate ( $M_2 = 1386.47$ ,  $df = 710$ ,  $p < .001$ ; RMSEA = .04). Overall, these results showed that the IRT item parameter estimates properly reflect the latent trait and describe the psychometric properties of the items. Specifically, parameter estimates (see Table 1) indicated that each item of the shortened scale performed well at measuring the latent

trait. Specifically, all item discrimination values were above 1.30, indicating large discriminative power (Baker & Kim, 2004), and location parameters were quite evenly spaced (with  $b_1$  and  $b_2$  well below the

**TABLE 1.** Discrimination (a) and Location (b) Parameters for Each Item of the Abbreviated Need for Cognition Scale (NFC-10)

Item*	a (SE)	b <sub>1</sub> (SE)	b <sub>2</sub> (SE)	b <sub>3</sub> (SE)	b <sub>4</sub> (SE)
NFC_1	1.63 (0.13)	-1.82 (0.13)	-0.59 (0.07)	-0.11 (0.07)	2.00 (0.15)
NFC_2	2.21 (0.18)	-2.36 (0.16)	-1.08 (0.07)	-0.54 (0.06)	1.47 (0.11)
NFC_3	1.54 (0.13)	-3.41 (0.30)	-1.40 (0.11)	-0.77 (0.08)	1.49 (0.13)
NFC_4	1.78 (0.15)	-2.72 (0.20)	-1.17 (0.09)	-0.66 (0.07)	1.36 (0.11)
NFC_5	1.65 (0.14)	-3.07 (0.25)	-1.69 (0.12)	-1.01 (0.08)	1.17 (0.11)
NFC_6	1.31 (0.11)	-2.75 (0.22)	-0.56 (0.08)	0.20 (0.08)	2.26 (0.19)
NFC_10	1.58 (0.14)	-3.36 (0.29)	-1.85 (0.14)	-0.98 (0.08)	1.35 (0.12)
NFC_11	1.42 (0.13)	-3.89 (0.39)	-2.18 (0.17)	-1.20 (0.10)	1.14 (0.12)
NFC_13	1.34 (0.12)	-2.96 (0.25)	-1.06 (0.10)	-0.17 (0.08)	2.20 (0.19)
NFC_15	1.44 (0.12)	-3.56 (0.32)	-1.31 (0.11)	-0.46 (0.07)	1.52 (0.14)

Note. Numbers refer to the original 18-item version.



**FIGURE 1.** Item Information Curve (IIC) of the original 18-item Need For Cognition (NFC) scale (left) and the 10-item NFC scale (right). Latent trait (Theta) is shown on the horizontal axis (higher values mean higher NFC), and the amount of information and the SE yielded by the test at each trait level is shown on the vertical axis.

**TABLE 2.**

Standard Error (*SE*) and Reliability ( $\omega$ ) Indices Yielded by the Original (NFC-18) and the Shortened (NFC-10) Scales for Each Level of the Theta ( $\theta$ ) Trait

$\theta$	<i>SE</i>			$\omega$		
	NFC-18	NFC-10	% change	NFC-18	NFC-10	% change
-3.0	.33	.38	15.1	.90	.87	3.3
-2.5	.32	.36	12.5	.91	.88	3.3
-2.0	.32	.35	9.3	.91	.89	2.2
-1.5	.31	.35	12.9	.91	.89	2.2
-1.0	.30	.34	13.3	.91	.90	1.1
-0.5	.31	.35	12.9	.91	.89	2.2
0.0	.33	.37	12.1	.90	.88	2.2
0.5	.34	.37	8.8	.90	.87	3.3
1.0	.33	.37	12.1	.90	.88	2.2
1.5	.33	.36	9.1	.90	.88	2.2
2.0	.35	.40	14.3	.89	.86	3.4
2.5	.40	.46	15.0	.86	.82	4.6
3.0	.46	.56	21.7	.82	.76	7.3

mean trait,  $b_3$  around the mean trait, and  $b_4$  well above it<sup>2</sup>), indicating that the item categories provide an adequate differentiation in measuring the trait.

The NFC-10 is highly reliable in measuring the different levels of the NFC construct continuum. Importantly, the original and shortened scales covered exactly the same range of the trait and, compared to the 18-item NFC scale, the shortened scale was less precise only at the extreme high levels of the NFC latent trait, as displayed in Figure 2. To quantify the change in reliability between the original and the shortened versions, we compared *SE*s and  $\omega$ s of the two versions (see Table 2). Overall, the percent change in *SE* was around 12% along the different trait levels, with a higher level of decrease (22%) observed for the extreme levels of the trait. The percent change in  $\omega$  was around 2-3% along the different trait levels and, again, the maximum decrease (7%) was observed for the extreme high levels of the trait. These results confirm that after excluding almost half of the items, the abbreviated scale maintained good accuracy from low to high levels of the trait, that is, the unavoidable loss of information did not reduce dramatically the reliability of the NFC-10 scale.

Finally, the gender DIF analysis confirmed that none of the ten items showed either *uniform* DIF (*p* values ranged from .14 to .94) or *nonuniform* DIF (*p* values ranged from .30 to .89). Thus, the NFC-10 was proven to be gender invariant, that is, the scale functions in the same way for male and female respondents and it provides unbiased measures of NFC.

**TABLE 3.**

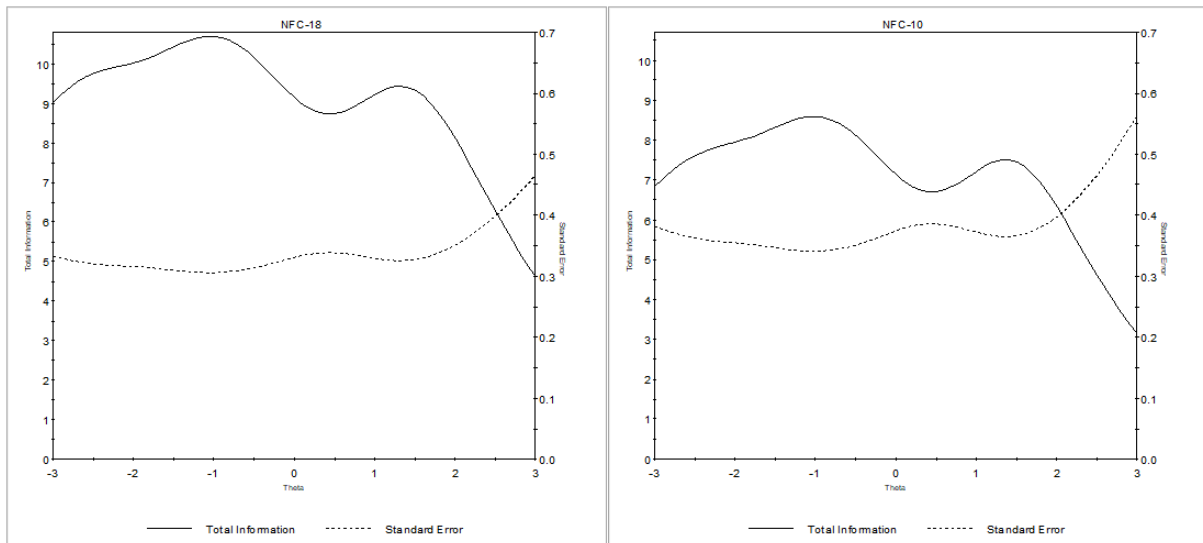
Correlations Between the 18- and 10-item Versions of the Need for Cognition Scale (NFC-18 and NFC-10) and All Other Variables in the Study

	<i>M</i> ( <i>SD</i> ) Range	NFC-18	NFC-10	<i>z</i>
Raven's APM (short form)	5.88 (2.19) 0-11	.18** ( <i>N</i> =274)	.20*** ( <i>N</i> =274)	-1.06
Cognitive Reflection Test (CRT)	1.30 (1.15) 0-3	.22*** ( <i>N</i> =621)	.24*** ( <i>N</i> =621)	-1.62
CRT-Long	3.44 (1.98) 0-6	.23*** ( <i>N</i> =369)	.27*** ( <i>N</i> =369)	-2.50*
Probabilistic Reasoning Scale	13.99 (2.05) 5-16	.15** ( <i>N</i> =362)	.15** ( <i>N</i> =362)	0.00
Math fluency	109.95 (21.95) 67-158	.30* ( <i>N</i> =58)	.33* ( <i>N</i> =58)	-0.70
Verbal thinking style	46.86 (7.99) 31-63	.39** ( <i>N</i> =58)	.37** ( <i>N</i> =58)	0.51
Spatial thinking style	42.53 (7.96) 25-65	.31* ( <i>N</i> =58)	.32* ( <i>N</i> =58)	-0.25
Object Visualizer thinking style	52.68 (9.99) 36-90	-.25* ( <i>N</i> =58)	-.34** ( <i>N</i> =58)	-2.22*
Abbreviated Math Anxiety Scale	14.63 (5.97) 7-34	-.36*** ( <i>N</i> =379)	-.37*** ( <i>N</i> =379)	-0.66
Subjective Numeracy Scale	34.53 (7.90) 12-87	.43*** ( <i>N</i> =380)	.42*** ( <i>N</i> =380)	0.68

Note. The *z*-comparisons were between the correlation coefficients obtained for the original (NFC-18) and the shortened (NFC-10) scales.

\* *p* < .05. \*\* *p* < .01. \*\*\* *p* < .001.

Table 3 presents descriptive statistics for all measures in the study, and the relations between the NFC-18, the NFC-10, and all other variables. Despite the relative homogeneity of the sample, there was a good range of scores on all measures, including the NFC-18 (*M* = 62.55, *SD* = 10.09; range: 30-88) and the NFC-10 (*M* = 35.58, *SD* = 6.36; range: 15-50). The Pearson product-moment correlations for the shortened NFC scale demonstrated that all the investigated relationships were significant and in the expected direction, and the original and the shortened scales showed very similar relationships with the investigated constructs. Steiger's *z* tests showed no significant differences when comparing the original and the abbreviated scales, except for the correlations with cognitive reflection and the object visualizer thinking style, which were stronger for the 10-item scale. These results confirm that the shortened form replicates the pattern of relationships established for the construct as measured by the long form of the test, that is, employing the NFC-10, we were able to replicate the nomological net of the NFC-18. Additionally, some expected relationships were even stronger when measured with the abbreviated scale. This finding might suggest that some items of the original version did not adequately represent the underlying construct and, as a consequence, their exclusion resulted in an improved scale.

**FIGURE 2.**

Test information function (TIF) of the original 18-item Need For Cognition (NFC) scale (left) and the 10-item NFC scale (right). Latent trait (Theta) is shown on the horizontal axis (higher values mean higher NFC), and the amount of information and the SE yielded by the test at each trait level is shown on the vertical axis.

## DISCUSSION

The construct of NFC was first introduced by Cacioppo and Petty (1982), and it continues to be commonly used in educational, cognitive and personality psychology research. The most popular version of the scale is the NFC-18, introduced by Cacioppo et al. (1984). Item response theory analyses offer a way to analyze the contribution of each item to measuring NFC, and to select items which ensure precise measurement across a range of trait levels.

Applying IRT, we demonstrated that some items of the NFC-18 scale did not perform well in measuring the latent trait. After excluding these items, all remaining items had high discriminative power and they measured a large spectrum of the trait, which covered the same range as the original scale. The scale also showed gender invariance, which demonstrates its suitability to measure the NFC trait among male and female participants. This result complements recent findings regarding the age invariance of NFC between 18-99 years of age (Soubelet & Salthouse, 2017).

The validity results demonstrated that the short version of the scale (see Appendix for the final list of items) showed the same (or even stronger) relationships with several related constructs as the original scale. In particular, as expected, the NFC-10 showed moderate relations with various measures of cognitive skills (i.e., fluid intelligence, cognitive reflection, probabilistic reasoning skills, arithmetic fluency) and self-report measures of cognitive styles (i.e., verbal, spatial and object visualizer), confidence in one's numerical skills (i.e., subjective numeracy) and mathematics anxiety. In line with previous studies regarding the validity of the NFC scale (Cacioppo et al., 1996; Soubelet & Salthouse, 2017), these correlations were low to moderate.

An interesting and somewhat unexpected finding was that, whereas the verbal and spatial thinking styles were positively related to NFC,

the object-visualizer thinking style showed a negative relationship, which was even stronger in the case of the NFC-10. This thinking style has been found to be associated with holistic processing, spontaneity, and artistic creations which offer a multiplicity of meanings (Pérez-Fabello et al., 2016). At the same time, this thinking style is negatively related to mathematical and visual-spatial reasoning (e.g., Hegarty & Kozhevnikov, 1999; Pitta-Pantazi et al., 2013). Although it is easy to see how this cognitive style might contrast with an analytical approach, this result is still somewhat puzzling given the strong links between NFC and curiosity (Cacioppo et al., 1996), as well as openness (Soubelet & Salthouse, 2017), which could be expected to relate to artistic interests. It would be interesting to conduct further studies regarding the links between NFC and both the production and appreciation of art, as this seems to be a neglected area so far.

Regarding the limitations of our study, we could mention that the participants were all university students, representing a relatively homogeneous population in terms of their age and intellectual ability. Future studies with a more heterogeneous sample could be useful for conducting further analyses regarding differential item functioning, for example, by comparing participants from different age groups, or participants with different levels of education. Nevertheless, even in the current sample, we have obtained a good range of scores on all measures, including both versions of the NFC scale. Another limitation is the range of validity measures, which mostly included measures of intelligence, quantitative skills and related thinking styles. Future studies could use a broader range of measures, including the measurement of constructs that could be expected to correlate strongly with NFC, such as the personality dimension of openness.

Notwithstanding these limitations, overall, our results confirm that we have obtained a much shorter version of the NFC that maintains excellent reliability and validity. Thus, the NFC-10 could be a very convenient and useful instrument for future studies.



## FOOTNOTES

<sup>1</sup> An advantage of using the extended version of the scale is that, due to the popularity of the scale, the original items are widely known, although a recent paper has argued that this does not affect the correlation between the Cognitive Reflection Test and other measures (Bialek & Pennycook, 2017).

<sup>2</sup> Location parameters are the thresholds that separate the response options of the Likert scale indicating on which part of the range of  $\theta$  people have a 50% chance of selecting each option given their  $\theta$  level. The need for cognition items are rated on a 5-point Likert scale. Therefore, each item has four (the number of response options minus 1) b parameters (i.e., the thresholds that separate 1 from 2–5 options, 1–2 from 3–5, 1–3 from 4–5, and 1–4 from 5).

## REFERENCES

- Arbuckle, J. L. (2003). *Amos 5.0* [Computer software]. Chicago, IL: Smallwaters.
- Arthur Jr, W., & Day, D. V. (1994). Development of a short form for the Raven Advanced Progressive Matrices Test. *Educational and Psychological Measurement, 54*, 394–403. doi: 10.1177/0013164494054002013
- Baker, F. B., & Kim, S. H. (Eds.). (2004). *Item response theory: Parameter estimation techniques*. New York, NY: Marcel Dekker.
- Berzonsky, M. D., & Sullivan, C. (1992). Social-cognitive aspects of identity style: Need for cognition, experiential openness, and introspection. *Journal of Adolescent Research, 7*, 140–155. doi: 10.1177/074355489272002
- Bialek, M., & Pennycook, G. (2017). The Cognitive Reflection Test is robust to multiple exposures. *Behavior Research Methods, 1*–7. doi: 10.3758/s13428-017-0963-x
- Blazhenkova, O., & Kozhevnikov, M. (2009). The new object-spatial-verbal cognitive style model: Theory and measurement. *Applied Cognitive Psychology, 23*, 638–663. doi: 10.1002/acp.1473
- Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131. doi: 10.1037/0022-3514.42.1.116
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of “need for cognition”. *Journal of Personality Assessment, 48*, 306–307. doi: 10.1207/s15327752jpa4803\_13
- Cacioppo, J. T., Petty, R. E., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*, 197–253. doi: 10.1037/0033-2909.119.2.197
- Cai, L., Maydeu-Olivares, A., Coffman, D. L., & Thissen, D. (2006). Limited information goodness-of-fit testing of item response theory models for sparse 2p tables. *British Journal of Mathematical and Statistical Psychology, 59*, 173–194. doi: 10.1348/000711005X66419
- Cai, L., Thissen, D., & du Toit, S. H. C. (2011). *IRTPRO 2.1 for Windows*. Chicago, IL: Scientific Software International.
- Chen, W. H., & Tiessen, D., (1997). Local dependence indices for items pairs using item response theory. *Journal of Educational and Behavioral Statistics, 22*, 265–289. doi: 10.2307/1165285
- Chiesi, F., Ciancaleoni, M., Galli, S., Morsanyi, K., & Primi, C. (2012). Item response theory analysis and differential item functioning across age, gender and country of a short form of the advanced progressive matrices. *Learning and Individual Differences, 22*, 390–396. doi: 10.1016/j.lindif.2011.12.007
- Chiesi, F., Primi, C., & Morsanyi, K. (2011). Developmental changes in probabilistic reasoning: The role of cognitive capacity, instructions, thinking styles and relevant knowledge. *Thinking & Reasoning, 17*, 315–350. doi: 10.1080/13546783.2011.598401.
- Clinton, V., Morsanyi, K., Alibali, M. W., & Nathan, M. J. (2016). Learning about probability from text and tables: Do color coding and labeling through an interactive-user interface help? *Applied Cognitive Psychology, 30*, 440–453. doi: 10.1002/acp.3223
- Dornic, S., Ekehammar, B., & Laaksonen, T. (1991). Tolerance for mental effort: Self-ratings related to perception, performance and personality. *Personality and Individual Differences, 12*, 313–319. doi: 10.1016/0191-8869(91)90118-U
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Lawrence Erlbaum.
- Fagerlin, A., Zikmund-Fisher, B. J., Ubel, P. A., Jankovic, A., Derry, H. A., & Smith, D. M. (2007). Measuring numeracy without a math test: development of the Subjective Numeracy Scale. *Medical Decision Making, 27*, 672–680.
- Fischbein, E. (1987). *Intuition in Science and Mathematics*. Dordrecht, The Netherlands: D. Reidel Publishing Company.
- Frederick, S. (2005). Cognitive reflection and decision making. *The Journal of Economic Perspectives, 19*, 25–42. doi: 10.1257/089533005775196732
- Gollwitzer, P. M., Kappes, H. B., & Oettingen, G. (2012). Needs and incentives as sources of goals. In H. Aarts & A. J. Elliot (Eds.), *Goal-directed behavior* (pp. 115–150). New York, NY: Psychology Press.
- Greco, L., & Walter, S. (2013). The need for cognition: A meta-analysis clarifying the link to intelligence and personality. In G. Atinc (Ed.), *Academy of Management Proceedings* (13718). Briarcliff Manor, NY: Academy of Management.
- Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory*. Newbury Park, CA: Sage Publications, Inc.
- Handley, S., Capon, A., Beveridge, M., Dennis, I., & Evans, J. St. B. T. (2004). Working memory, inhibitory control and the development of children's reasoning. *Thinking and Reasoning, 10*, 175–195. doi: 10.1080/13546780442000051
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and mathematical problem solving. *Journal of Educational Psychology, 91*, 684–689. doi: 10.1037/0022-0663.91.4.684

- Hevey, D., Thomas, K., Pertl, M., Maher, L., Craig, A., & Ni Chuinneagain, S. (2012). Method effects and the Need for Cognition scale. *International Journal of Educational and Psychological Assessment, 12*, 20–33.
- Hill, B. D., Foster, J. D., Elliott, E. M., Shelton, J. T., McCain, J., & Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *Journal of Research in Personality, 47*, 22–25. doi: 10.1016/j.jrp.2012.11.001
- Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS) construction, validity, and reliability. *Assessment, 10*, 178–182. doi: 10.1177/1073191103010002008
- Kokis, J. V., Macpherson, R., Toplak, M. E., West, R. F., Stanovich, K. E. (2002). Heuristic and analytic processing: Age trends and associations with cognitive ability and cognitive styles. *Journal of Experimental Child Psychology, 83*, 26–52. doi: 10.1016/S0022-0965(02)00121-2
- Kruglanski, A. W., & Webster, D. M. (1996). Motivated closing the mind: “Seizing” and “freezing”. *Psychological Review, 103*, 263–283. doi: 10.1037/0033-295X.103.2.263
- Liberali, J. M., Reyna, V. F., Furlan, S., Stein, L. M., & Pardo, S. T. (2012). 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
- Lin, Y., Durbin, J. M., & Rancer, A. S. (2016). Math anxiety, need for cognition, and learning strategies in quantitative communication research methods courses. *Communication Quarterly, 64*, 390–409. doi: 10.1080/01463373.2015.1103294
- Marcoulides, G. A., & Hershberger, S. L. (1997). *Multivariate statistical methods. A first course*. Mahawa, NJ: Lawrence Erlbaum Associates.
- McDonald, R. P. (2013). *Test theory: A unified treatment*. London, England: Psychology Press.
- Morsanyi, K., O'Mahony, E., & McCormack, T. (2017). Number comparison and number ordering as predictors of arithmetic performance in adults: Exploring the link between the two skills, and investigating the question of domain-specificity. *The Quarterly Journal of Experimental Psychology, 70*, 2497–2517. doi: 10.1080/17470218.2016.1246577
- Morsanyi, K., Primi, C., Chiesi, F., & Handley, S. (2009). The effects and side-effects of statistics education. Psychology students' (mis-)conceptions of probability. *Contemporary Educational Psychology, 34*, 210–220. doi: 10.1016/j.cedpsych.2009.05.001
- Mueller, J. H., & Johnson, W. C. (1990). Trait distinctiveness and age specificity in self-referent information processing. *Bulletin of the Psychonomic Society, 28*, 119–122. doi: 10.3758/BF03333980
- Olson, K., Camp, C., & Fuller, D. (1984). Curiosity and need for cognition. *Psychological Reports, 54*, 71–74. doi: 10.2466/pr0.1984.54.1.71
- Pérez-Fabello, M. J., Campos, A., & Campos-Juanatey, D. (2016). Is object imagery central to artistic performance? *Thinking Skills and Creativity, 21*, 67–74. doi: 10.1016/j.tsc.2016.05.006
- Petty, R. E., & Jarvis, B. G. (1996). An individual difference perspective on assessing cognitive processes. In N. Schwarz & S. Sudman (Eds.), *Answering questions: Methodology for determining cognitive and communicative processes in survey research*, (pp. 221–257). San Francisco, CA: Jossey-Bass.
- Pitta-Pantazi, D., Sophocleous, P., & Christou, C. (2013). Spatial visualizers, object visualizers and verbalizers: Their mathematical creative abilities. *ZDM, 45*, 199–213. doi: 10.1007/s11858-012-0475-1
- Primi, C., Morsanyi, K., Chiesi, F., Donati, M. A., & Hamilton, J. (2016). The development and testing of a new version of the cognitive reflection test applying item response theory (IRT). *Journal of Behavioral Decision Making, 29*, 453–469. doi: 10.1002/bdm.1883
- Primi, C., Morsanyi, K., Donati, M. A., Galli, S., & Chiesi, F. (2017). Measuring probabilistic reasoning: The construction of a new scale applying item response theory. *Journal of Behavioral Decision Making, 30*, 933–950. doi: 10.1002/bdm.2011
- Raven, J. C. (1962). *Advanced progressive matrices*. London, England: Lewis & Co. Ltd.
- Reeve, B. B., & Fayers, P. (2005). Applying item response theory modeling for evaluating questionnaire item and scale properties. *Assessing Quality of Life in Clinical Trials: Methods of Practice, 2*, 55–73.
- Sadowski, C. J. (1993). An examination of the short Need for Cognition Scale. *Journal of Psychology, 127*, 451–454. doi: 10.1080/00223980.1993.9915581
- Sadowski, C. J., & Gulgoz, S. (1992). Internal consistency and test-retest reliability of the Need for Cognition Scale. *Perceptual and Motor Skills, 74*, 610. doi: 10.2466/pms.1992.74.2.610
- Samejima, F. (1969). Estimation of latent ability using response pattern of graded scores. *Psychometrika Monograph Supplement, 34*, 100.
- Schipolowski, S., Schroeders, U., & Wilhelm, O. (2014). Pitfalls and challenges in constructing short forms of cognitive ability measures. *Journal of Individual Differences, 35*, 190–200. doi: 10.1027/1614-0001/a000134
- Sinatra, G. M., & Pintrich, P. R. (2003). *Intentional conceptual change*. Mahwah, NJ: Erlbaum.
- Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short form development. *Psychological Assessment, 12*, 102–111. doi: 10.1037//1040-3590.12.1.102
- Soubelet, A., & Salthouse, T. A. (2017). Does Need for Cognition have the same meaning at different ages? *Assessment, 24*, 987–998. doi: 10.1177/1073191116636449
- Stanovich, K. E. (2009). *What intelligence tests miss: The psychology of rational thought*. New Haven, CT: Yale University Press.
- Stanovich, K. E. (2011). *Rationality and the reflective mind*. New York, NY: Oxford University Press.

- Stanovich, K. E., & West, R. F. (1999). Discrepancies between normative and descriptive models of decision making and the understanding acceptance principle. *Cognitive Psychology, 38*, 349–385. doi: 10.1006/cogp.1998.0700
- Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning, 13*, 225–247. doi: 10.1080/13546780600780796
- Steiger, J. H. (1980). Tests for comparing elements of a correlation matrix. *Psychological Bulletin, 87*, 245–251. doi: 10.1037/0033-2909.87.2.245
- Sternberg, R. J. (2003). *Wisdom, intelligence, and creativity synthesized*. Cambridge, MA: Cambridge University Press.
- Strathman, A., Gleicher, F., Boninger, D. S., & Edwards, S. (1994). The consideration of future consequences: Weighing immediate and distant outcomes of behavior. *Journal of Personality and Social Psychology, 66*, 742–752. doi: 10.1037/0022-3514.66.4.742
- Tanaka, J. S., Panter, A. T., & Winborne, W. C. (1988). Dimensions of the need for cognition: Subscales and gender differences. *Multivariate Behavioral Research, 23*, 35–50. doi: 10.1207/s15327906mbr2301\_2.
- Thissen, D., Steinberg, L., & Wainer, H. (1988). Use of item response theory in the study of group differences in trace lines. In Wainer, H. & Braun, H. I. (Eds.), *Test Validity*, (pp. 147–169). Hillsdale, NJ: Erlbaum.
- Venkatraman, M. P., Marlino, D., Kardes, F. R., & Sklar, K. B. (1990). The interactive effects of message appeal and individual differences on information processing and persuasion. *Psychology and Marketing, 7*, 85–96. doi: 10.1002/mar.4220070202
- Venkatraman, M. P., & Price, L. L. (1990). Differentiating between cognitive and sensory innovativeness. *Journal of Business Research, 20*, 293–315. doi: 10.1016/0148-2963(90)90008-2
- Verplanken, B. (1991). Persuasive communication of risk information: A test of cue versus message processing effects in a field experiment. *Personality and Social Psychology Bulletin, 17*, 188–193. doi: 10.1177/014616729101700211
- Verplanken, B., Hazenberg, P. T., & Palenewen, G. R. (1992). Need for cognition and external information search effort. *Journal of Research in Personality, 26*, 128–136. doi: 10.1016/0092-6566-(92)90049-A
- Waters, L. K., & Zakrajsek, T. (1990). Correlates of need for cognition total and subscale scores. *Educational and Psychological Measurement, 50*, 213–217. doi: 10.1177/0013164490501026
- Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology, 67*, 1049–1062. doi: 10.1037/0022-3514.67.6.1049
- West, R. F., Toplak, M. E., & Stanovich, K. E. (2008). Heuristics and biases as measures of critical thinking: associations with cognitive ability and thinking dispositions. *Journal of Educational Psychology, 100*, 930. doi: 10.1037/a0012842
- Woo, S. E., Harms, P. D., & Kuncel, N. R. (2007). Integrating personality and intelligence: Typical intellectual engagement and need for cognition. *Personality and Individual Differences, 43*, 1635–1639. doi: 10.1016/j.paid.2007.04.022
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson tests of achievement*. Itasca, IL: Riverside Publishing.
- Wu, C. H., Parker, S. K., & De Jong, J. P. (2014). Need for cognition as an antecedent of individual innovation behavior. *Journal of Management, 40*, 1511–1534. doi: 10.1177/0149206311429862
- Ziegler, M., Kemper, C. J., & Kruyen, P. (2014). Short scales—Five misunderstandings and ways to overcome them. *Journal of Individual Differences, 35*, 185–189. doi: 10.1027/1614-0001/a000148

RECEIVED 17.12.2017 | ACCEPTED 24.04.2018

## APPENDIX A

The 10-item Need for Cognition Scale. Items that are reverse scored are marked by asterisks.

1. I would prefer complex to simple problems.
2. I like to have the responsibility of handling a situation that requires a lot of thinking.
3. Thinking is not my idea of fun.\*
4. I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.\*
5. I try to anticipate and avoid situations where there is likely a chance I will have to think in depth about something.\*
6. I find satisfaction in deliberating hard and for long hours.
7. The idea of relying on thought to make my way to the top appeals to me.
8. I really enjoy a task that involves coming up with new solutions to problems.
9. I prefer my life to be filled with puzzles that I must solve.
10. I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.