Short-Scale Construction Using Meta-Analytic Ant Colony Optimization

A Demonstration With the Need for Cognition Scale

Ulrich Schroeders[1], Maximilian Morgenstern[1], Kristin Jankowsky[1], and Timo Gnambs[2]

[1]Institute of Psychology, University of Kassel, Germany
[2]Leibniz Institute for Educational Trajectories, Bamberg, Germany

Abstract: The Need for Cognition Scale (NCS) is a self-report scale measuring individual differences in the tendency to engage in and enjoy thinking. The shortened version with 18 items (NCS-18; Cacioppo et al., 1984) has widely been administered in research on persuasion, critical thinking, and educational achievement. Whereas most studies advocated for essential uni-dimensionality, the question remains which psychometric model yields the best representation of the NCS-18. In the present study, we compared six competing measurement models for the NCS-18 with meta-analytic structural equation models using summary data of 87 samples (N = 90,215). Results demonstrated that the negatively worded items introduced considerable measurement bias that was best accounted for with an acquiescence model. In a further analytical step, we showcased how the pooled correlation matrix can be used to compile short versions of the NCS-18 via Ant Colony Optimization. We examined model fit and reliability of short scales with varying item numbers (between 4 and 15) and a balanced ratio of positively and negatively worded items. We discuss the potentials and limits of the newly proposed method.

Keywords: Need for cognition, NCS-18, meta-analytic structural equation modeling, metaheuristic, ant colony optimization

Need For Cognition (NFC) describes people's "tendency to engage in and enjoy thinking" (Cacioppo & Petty, 1982, p. 116). The construct has initially been used in various fields of social psychology including decision-making (Levin et al., 2000), persuasion (DeSteno et al., 2004), and priming (Petty et al., 2008), and especially as a motivational factor in the context of the elaboration likelihood model (e.g., Petty et al., 1993). Since then, NFC has become a popular construct that has been applied to many other psychological disciplines (Petty et al., 2009), for example, in cognitive psychology, research on critical thinking (West et al., 2008), problem solving (Nair & Ramnarayan, 2000), as well as memory recall and recognition (Kardash & Noel, 2000). NFC has also become influential in educational psychology as an intellectual investment trait (Jebb et al., 2016; Mussel, 2013) for explaining interindividual differences in learning and educational achievement in primary and secondary school (Colling et al., 2022; Luong et al., 2017) as well in university (Grass et al., 2017).

The construct of NFC is strongly tied to a specific measurement instrument – the 18-item short version of the Need for Cognition Scale (NCS-18; Cacioppo et al., 1984) – which has been included in several hundreds of articles since its development. Whereas in most studies, a uni-dimensional structure is assumed (e.g., Cacioppo et al., 1984; Sadowski, 1993), there are also alternative multidimensional conceptualizations (e.g., Tanaka et al., 1988), or solutions that incorporate method-specific variance that is caused by negatively worded items. In the present study, we examined the factor structure of the NCS-18 with meta-analytical structural equation modeling (MASEM) by comparing competing measurement models including models that try to tap method-specific variances such as an acquiescence model (Billiet & McClendon, 2000) and bifactor models (Eid et al., 2017). Moreover, we used the NCS-18 as an example to demonstrate how the meta-analytically derived correlation matrix can be used as a starting point for short-scale construction using metaheuristics such as Ant Colony Optimization (ACO; Schroeders et al., 2016a).

Development and Dimensionality of the Need for Cognition Scale

The original Need for Cognition Scale was developed by Cacioppo and Petty (1982). Drawing on earlier work by Cohen et al. (1955), they initially developed a self-report measure with 45 items that were intended to capture a single major trait and selected the 34 items that discriminated best between a group of blue- versus white-collar workers (low vs. high in NFC). Already in this initial version, the NCS-34, several items were reversely scored (i.e., a negative answer indicating higher levels of NFC) to counteract a potential response bias of acquiescence. Only a few years later, this scale was revised and shortened to an 18-item version (NCS-18; Cacioppo et al., 1984) that correlated highly with the long version ($r = .95$). The NCS-18 became the most popular reference version for measuring NFC and was translated into several languages including Chinese (Kao, 1994), Portuguese (Gomes et al., 2013), Spanish (Maldonado et al., 1993), and Turkish (Gülgöz & Sadowski, 1995). In some instances, the development of instruments to assess NFC decoupled early on from the original test development. For example, Bless et al. (1994) compiled from the long version (Cacioppo & Petty, 1982) a 16-item German short scale that has 12 items in common with the NCS-18. There are also adaptions for children such as the 20-item version in French (Ginet & Py, 2000), a 14-item version in German (Keller et al., 2019), or – derived from the latter – the Polish version (Tanaś, 2021). The following explanations refer exclusively to the NCS-18.

The NCS-18 was designed to measure NFC uni-dimensionally (e.g., Cacioppo et al., 1984) and the majority of studies seem to support this construction rationale (Cacioppo et al., 1996; Culhane et al., 2004; Lins de Holanda Coelho et al., 2020; Perri & Wolfgang, 1988; Pieters et al., 1987; Sadowski, 1993). Multidimensional solutions often concern the longer versions of the scale (e.g., Tanaka et al., 1988; Waters & Zakrajsek, 1990). Because the items with the highest factor loading on the first factor were intentionally selected for the 18-item short version, likely, the resulting short scale is indeed uni-dimensional. Sometimes multidimensional structures of the short scale are reported for translated versions of the NCS-18 (e.g., Maldonado et al., 1993). For example, Gomes et al. (2013) reported a multidimensional structure for the Portuguese version with three moderately correlated factors (i.e., cognitive effort, preference for complexity, and desire for understanding). Lord and Putrevu (2006) even identified four dimensions for both the original and the abbreviated scale (i.e., enjoyment of cognitive stimulation, preference for complexity, commitment to cognitive effort, and desire for understanding). In summary, however, it must be ascertained that solutions with multiple

distinctive factors are rare in research on the structure of the NCS-18.

Much more attention, however, has been paid to the question of whether the strict form of uni-dimensionality of the scale can be maintained since half of the NCS-18 items are negatively worded (e.g., "Thinking is not my idea of fun"). These negatively worded items were intentionally developed and retained in the short version to control for response bias or careless/insufficient effort responding. The different item wording induces systematic method-specific variance in self-report scales that can be accounted for by several psychometric models that have been proposed in the literature (e.g., DiStefano & Motl, 2006; Gnambs & Schroeders, 2020): (a) a two-dimensional model, (b) a correlated-uniqueness model, (c) bifactor models with and without a reference factor, and (d) an acquiescence model. Some, but not all of these models have been applied to the NCS-18 (for a graphical representation see Figure 1). Please note that our meta-analytic investigation focused on the variable-side to acknowledge wording effects. A complementary line of research tries to identify heterogeneity in response patterns between positively and negatively worded items on the person side with factor mixture modeling (Kam & Fan, 2020) or mixture item response model (Jin et al., 2017).

A two-dimensional model in which the positively and negatively worded items load on separate but correlated factors showed a better fit than the uni-dimensional model (Forsterlee & Ho, 1999; Hevey et al., 2012). However, a two-dimensional model seems only appropriate if the additionally specified factor represents something substantively different rather than a mere negation of the first factor (cognizers vs. cognitive misers), which is a questionable assumption given the clear loading pattern and the test authors' clear intention to develop a uni-dimensional scale. Hereto, Zhang et al. (2016) examined the effect of reversely worded items on the factor structure of the NCS-18 by manipulating the extent of negatively worded items (none, half, all). As a result, the versions with homogeneously positively or negatively formulated items showed a clearer uni-dimensional structure than the original version for which the additional specification of a method factor was necessary.

To circumvent this argumentative problem of using positively and negatively worded items to measure a single, uni-dimensional construct, the correlated-uniqueness model has been suggested, in which correlated errors among the negatively worded items are introduced (Marsh, 1989; Marsh & Bailey, 1991). The idea is that the NCS-18 is principally uni-dimensional, but that response bias produces correlated residual variances. Systematic comparisons using confirmatory factor analysis repeatedly showed the superiority of the correlated-uniqueness model over a
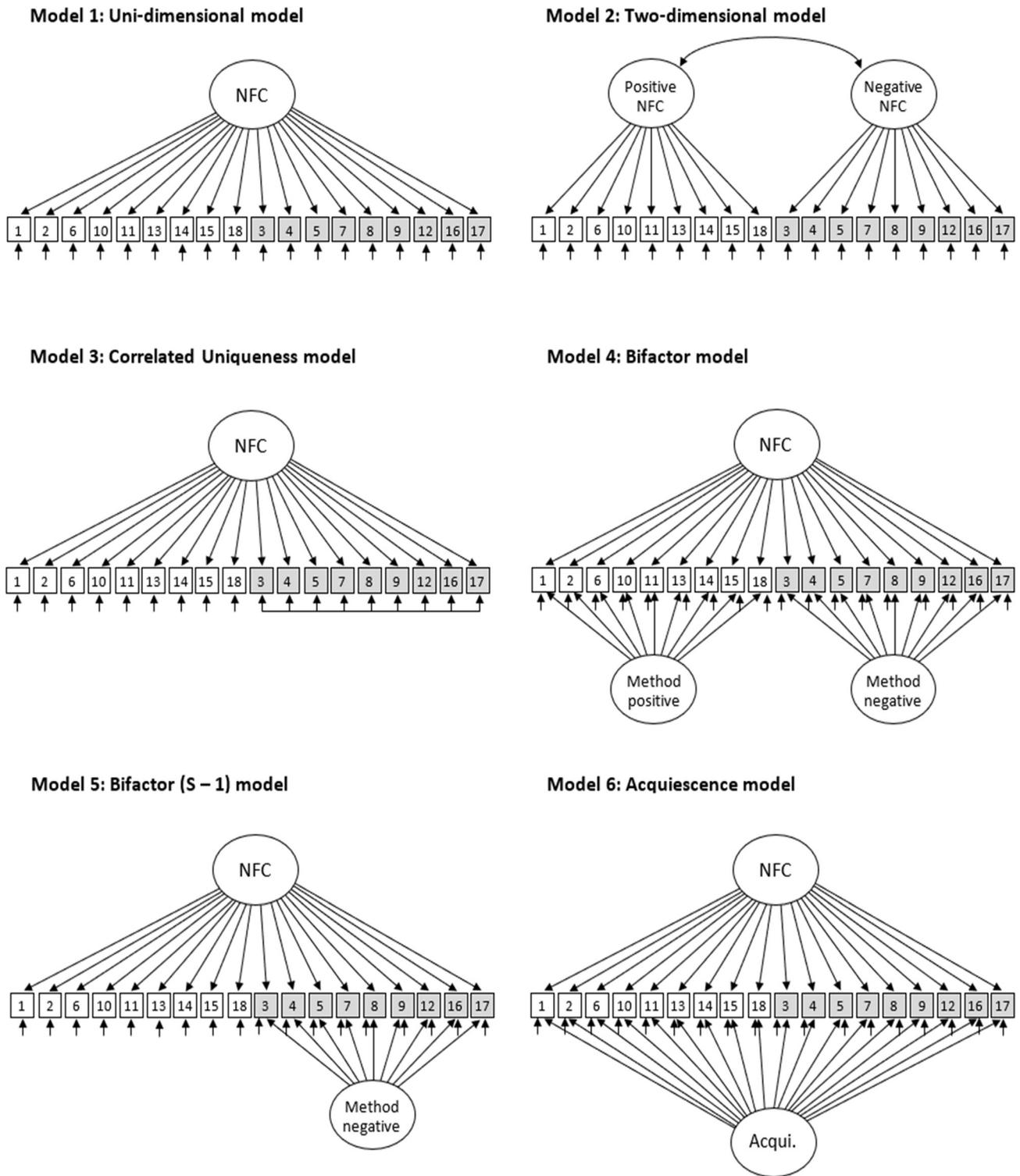
Model 1: Uni-dimensional model

Model 2: Two-dimensional model

Model 3: Correlated Uniqueness model

Model 4: Bifactor model

Model 5: Bifactor (S − 1) model

Model 6: Acquiescence model

Figure 1. Competing measurement models for the NCS-18

two-dimensional model in terms of fit indices for the NCS-18 (Forsterlee & Ho, 1999; Georgiou & Kyza, 2017; Hevey et al., 2012). However, specifying correlated residuals has several disadvantages (Conway et al., 2004; Lance et al., 2002), for example, that a person's individual bias cannot be expressed with this modeling approach and that the trait variance is likely to be overestimated (Kenny & Kashy, 1992).

Bifactor models consist of a general factor reflecting the common variance of all items and specific, uncorrelated

factors to capture additional variance among item sets. Bifactor models have recently experienced a renaissance as an important structural representation of multidimensionality within a uni-dimensional construct (Reise, 2012; Reise et al., 2010). However, bifactor models often lead to anomalous results such as negligible specific factors or irregular loading patterns. To overcome these shortcomings, Eid et al. (2017) proposed two alternative bifactor models in which either an indicator or a factor is set as a reference, whereas the remaining items constitute an uncorrelated method factor. Different bifactor models have been examined for strongly revised versions of the NCS-18 (e.g., Georgiou & Kyza, 2017; Preckel, 2014) and the original short scale (e.g., Bors et al., 2006; Zhang et al., 2016). The results indicated that a model with an additional method factor for the negatively keyed items – a bifactor (S−1) model – outperformed other modeling approaches.

The last model to be mentioned in this context is the acquiescence model (Billiet & McClendon, 2000). Acquiescence describes the tendency to agree to an item independent of its content (Ferrando & Lorenzo-Seva, 2010). In such a model, all items load on a method factor in addition to the trait factor with factor loadings (in the recoded data set) fixed at −1 for negatively worded items and +1 for positively worded items, whereas the variance of the method factor is freely estimated to reflect individual differences in acquiescence (Aichholzer, 2014; Billiet & McClendon, 2000). In the context of the NCS-18, the acquiescence model has only been applied once, showing a better model fit than the two-factor solution (Bruinsma & Crutzen, 2018).

Short Scale Construction

The rise of longitudinal and multivariate studies in psychological research has created a greater need for psychometrically sound short scales (Dörendahl & Greiff, 2020). This trend is not exclusive to psychological research but also extends to related fields such as educational research, sociology, and economics. Employing short scales is advantageous in large-scale assessments, where even minor reductions in test length can lead to significant cost savings and potentially higher participant response rates (Schoeni et al., 2012), or in longitudinal experience sampling studies, where individual assessments should be kept short (Burchert et al., 2021). Typically, short scales are constructed by abbreviating existing scales based on some naïve, reliability-based item selection strategy (e.g., part-whole corrected item-total correlation or "alpha if item deleted" statistics, for an overview see Kruyen et al., 2013). However, it has repeatedly been shown that metaheuristics such as Ant Colony Optimization (ACO) are superior to traditional item selection procedures (Jankowsky et al., 2020; Leite et al., 2008; Olaru et al., 2015, 2019;

Schroeders et al., 2016a), because they offer the possibility to consider multiple criteria simultaneously (Steger et al., 2022), operate on the level of scale statistics rather than item statistics (Schroeders et al., 2016b), and are not prone to sequence effects of item removal (Olaru et al., 2019).

Selecting appropriate items from a larger item pool can be conceptualized as a combinatorial problem: Which items should be selected to suffice some preset criteria such as good model fit and high predictive validity? In general, the complexity of the combinatorial tasks increases dramatically with an increased number of items in the pool and may not be solvable in a reasonable amount of time with deterministic algorithms (i.e., exhaustive search). To tackle such combinatorial tasks, applied computer scientists have developed metaheuristics that can identify an optimal (or almost optimal) solution in a reasonable amount of time (Dorigo & Stützle, 2010). Metaheuristics are often inspired by biological processes (e.g., evolution) or natural adaptations (e.g., the foraging behavior of ants). Due to their versatility, they have also been proven as a highly effective tool for shortening scales in psychological assessment (Janssen et al., 2017; Leite et al., 2008; Olaru et al., 2015, 2019; Schroeders et al., 2016a). In the present context, we utilize the ACO algorithm which mimics the behavior of some ant species when foraging (Deneubourg et al., 1983, 1990).

We will provide a brief overview of the fundamental principles and analogies to enhance comprehension of the functioning of ACO (for a more detailed description, see Olaru et al., 2019). Ants employ pheromone trails to find the shortest path from the nest to the food source. On shorter routes, these trails tend to accumulate more rapidly, thereby attracting more ants. In a short period of time, the routes are refined until an efficient path is identified. In the context of short-scale construction, the different paths correspond to the various item sets that are randomly selected from a larger pool of items. Each set is evaluated based on a predefined optimization function, for example, maximizing reliability (i.e., minimizing route length). Just like pheromones accumulate more quickly on shorter routes, enticing more ants to follow them, items belonging to the best set in each iteration are assigned higher virtual pheromone values. Higher pheromones are equivalent to a higher drawing probability of these items being selected in subsequent iterations. The search process continues until no further improvements can be achieved.

Although the NCS-18 is already the short version of the initial 34-item measures, the demand for even shorter versions was repeatedly expressed. For example, Chiesi et al. (2018) introduced a 10-item version selecting the most informative items by means of item response theory, without any loss of critical validity compared to the long version. Similarly, Lins de Holanda Coelho et al. (2020) constructed a 6-item ultra-short scale (NCS-6) by manually selecting

moderately difficult and informative items that had a high item-total correlation.

The Present Study

We used meta-analytic structural equation modeling (MASEM; Cheung, 2014, Cheung & Cheung, 2016) to summarize the existing empirical work on the dimensional structure of the NCS-18 and to evaluate competing measurement models (see Figure 1). One advantage of MASEM is that the results of individual small, heterogeneous studies are combined (and weighted) so that more robust statements about the dimensionality of the NCS-18 can be made beyond a specific sample. More precisely, we employed a two-stage structural equation modeling (TSSEM; Cheung & Chan, 2005). In this method, correlation coefficients between the item scores are initially extracted from primary studies and meta-analytically combined into a pooled correlation matrix. Subsequently, confirmatory factor models are fitted to the pooled correlation matrix.

With the present research, we pursue two research goals: The first deals with the optimal factor analytic representation of the NCS-18. Most studies used the NCS-18 to include a short, prominent measure of intellectual engagement as a personality trait, thus, finding the correct underlying structure was not a primary concern. Since the NCS-18 was designed to capture a single latent construct, its uni-dimensionality has often been presupposed rather than examined in a confirmatory factor analytic manner. Thus, we examine the different psychometric conceptualizations that have been proposed in the literature. More specifically, we specified a two-dimensional model, a correlated uniqueness model (Marsh, 1989), two different types of bifactor models (Eid et al., 2017), and an acquiescence model (Billiet & McClendon, 2000) to find an optimal structural representation of the NCS-18.

Our second goal is to propose a new method that combines meta-analytic SEM and metaheuristics to compile short scales. In more detail, the above-mentioned pooled correlation matrix was used as a starting point for metaheuristic optimization (Schroeders et al., 2016a). Although the search for the best model could be exhaustive in the present context (i.e., selecting subsets of items out of a pool of 18 items), we showcase the more generic approach in a proof-of-concept study that can easily be adapted to more complex scenarios. Thus, MASEM-ACO combines the advantages of meta-analytic aggregation of structural information using MASEM and item sampling techniques using ACO that are eligible to consider multiple criteria simultaneously.

Methods

In an open data repository, we provide relevant material including the codebook, the coded data, and annotated syntax for all analyses to reproduce the reported findings (see Schroeders et al., 2024, https://osf.io/tbrdv). Furthermore, we present the results of supplemental analyses which are briefly referenced in the main text.

Meta-Analytic Database

Search Strategy

The search for primary studies reporting on the factor structure of the NCS-18 covered Google Scholar, main scientific databases (e.g., PsycArticles, PsycINFO, and PSYNDEX), open data repositories (e.g., OSF, Mendeley Data), and major journals sharing primary data (e.g., PLOS ONE, Data in Brief).[1] In May 2023, we identified 4,024 potentially relevant journal articles and data sets using the Boolean expression ("NFC scale" OR "NC scale" OR "need for cognition scale") AND ("factor analysis" OR "factor structure" OR "principal component analysis" OR "item analysis"). After scanning the titles, abstracts, and, subsequently, tables and figures of these manuscripts or the raw data, we reviewed the full text of 77 studies. We retained all studies that met the following criteria:

(a) In the study, the original or a translated version of the NCS-18 was administered (i.e., all studies that altered the wording/meaning of the items or used the extended NCS-34 were excluded, despite an overlap in items).

(b) The necessary item-level statistics were available either as raw data, as the full correlation (or covariance) matrix, or as the loading pattern from an exploratory (or confirmatory) factor analysis. In case the raw data of a study was available, we calculated the respective correlation matrix.

(c) The sample size was reported.

We excluded studies that reported the results of a factor analysis that was jointly conducted with items of another

---

[1] Open Science Framework: https://osf.io, PsychArchives: https://www.psycharchives.org, Harvard Dataverse: https://dataverse.harvard.edu, Mendeley Data: https://data.mendeley.com, Kaggle: https://www.kaggle.com/datasets, Google Dataset Search: https://datasetsearch.research.google.com, Journal of Open Psychology Data: https://openpsychologydata.metajnl.com, Scientific Data: https://www.nature.com/sdata/, Data in Brief: https://www.sciencedirect.com/journal/data-in-brief, eLife: https://elifesciences.org, PLOS: https://journals.plos.org/plosone/.

measure besides the NCS-18. We also excluded studies with factor analytic results that did not accurately describe the empirical data (i.e., explained variance below R^2 = .30 in exploratory factor analysis or insufficient model fit in confirmatory factor analysis). No further exclusions were made based on sample characteristics, publication year, type of publication (e.g., peer-reviewed or not), or the language of publication. We also made an open social media call for unpublished studies including the NCS-18 and asked colleagues directly via email to send raw data or the item correlation matrices. Several authors were responsive to our request (Barceló, 2023; Cartwright et al., 2009; Edwards, 2009; Gomes et al., 2013; Grădinaru et al., 2023; Karagiannopoulou et al., 2020; Koutsogiorgi, 2020; Lee et al., 2020; Pilli & Mazzon, 2016; Powell et al., 2016; Sousa et al., 2018; van Tilburg et al., 2019; Weigold & Weigold, 2022; Weng & DeMarree, 2019). This literature search and screening process resulted in 57 publications with 87 samples that were included in our meta-analysis (see Figure 2 for an overview).

Coding Procedure

We defined all relevant information to be extracted from each publication accompanied by relevant coding guidelines in a coding protocol (see Electronic Supplementary Material, ESM 1). The focal information pertained to correlations of the 18 items of the NCS either calculated from the raw data or extracted from the publication as well as the factor loading patterns. If different factor solutions for the same sample were available, we used the factor loading pattern with the highest number of factors. In addition, descriptive information was collected on the publication (e.g., publication year, type of publication), the sample (e.g., sample size, country, language, mean age, percentage of women), and the reported factor analysis (e.g., number of extracted factors, factor analytic method). If raw data were available, the respective information was calculated from the data. All studies' characteristics were coded by the first author and a second time by the last author independently to evaluate the coding process. Data extraction was mostly script-based, with the correctness of the scripts also double-checked. The intercoder agreement was quantified using Krippendorff's α (Krippendorff, 2013) which indicated very good agreement with values between 0.88 and 1.00. Deviations in the rating were resolved by consensus.

Evaluation of Risk of Bias

The quality of the available studies was assessed using eight slightly adapted items from the risk of bias scale which was developed to evaluate the potential biases of primary studies included in systematic reviews and meta-analyses (Nudelman & Otto, 2020). This quality screening was specifically designed for observational studies that did not involve any interventions. The items included in the evaluation covered various aspects such as the recruitment of participants (i.e., whether appropriate methods were used to select respondents), the size of the sample, and data management procedures (i.e., whether data cleaning procedures were reported including the handling of invalid responses or outliers). The specific items, along with the modifications made, can be found in ESM 1, Table E1. The risk of bias was determined by the sum score across the eight items with higher scores indicating a greater risk of bias. The first and last author independently rated all of the studies. There was a high level of agreement between the ratings as indicated by Krippendorff's α coefficient of .89, which is why the mean of both raters' scores was used for the primary analyses.

Meta-Analytic Procedure

Meta-Analytic Factor Analyses

We examined the factor structure of the NCS-18 with MASEM that integrates two established techniques that have a long-standing tradition but limited mutual exchange (Cheung, 2013; Jak, 2015). In more detail, we used the two-stage structural equation modeling approach (TSSEM; Cheung & Chan, 2005). Recently, a one-stage MASEM (OSMASEM) has been introduced (Jak & Cheung, 2020). OSMASEM and TSSEM (without moderators) typically result in highly comparable point estimates and standard errors for the SEM parameters (e.g., Gnambs & Sengewald, 2023; Jak & Cheung, 2023) and, thus, can be used interchangeably. One advantage of TSSEM is that it is computationally more efficient and faster. Moreover, meta-analytic exploratory factor analyses and the here implemented Ant Colony Optimization require the pooled correlation matrix, similar to TSSEM. In the first stage of TSSEM, the item-level correlation matrices were pooled using a random-effects meta-analysis with a maximum likelihood estimator (Cheung & Cheung, 2016). In doing so, we used the zero-order Pearson product-moment correlations between the items as effect size measures (for a graphical representation of the correlation matrix of all correlation matrices, see ESM 1, Figure E1 which can be used for visual detection of outliers). For the majority of publications, raw data was available (57 samples), whereas correlation matrices were rarely depicted (14 samples). For 16 samples, we calculated the model-implied item-level correlations based on the reported factor pattern matrices from exploratory or confirmatory factor analyses (Gnambs & Staufenbiel, 2016).

In the second stage of MASEM, the derived pooled correlation matrix was subjected to weighted least square factor analyses, because simply taking a pooled correlation matrix as input for a structural equation model is inaccurate (see
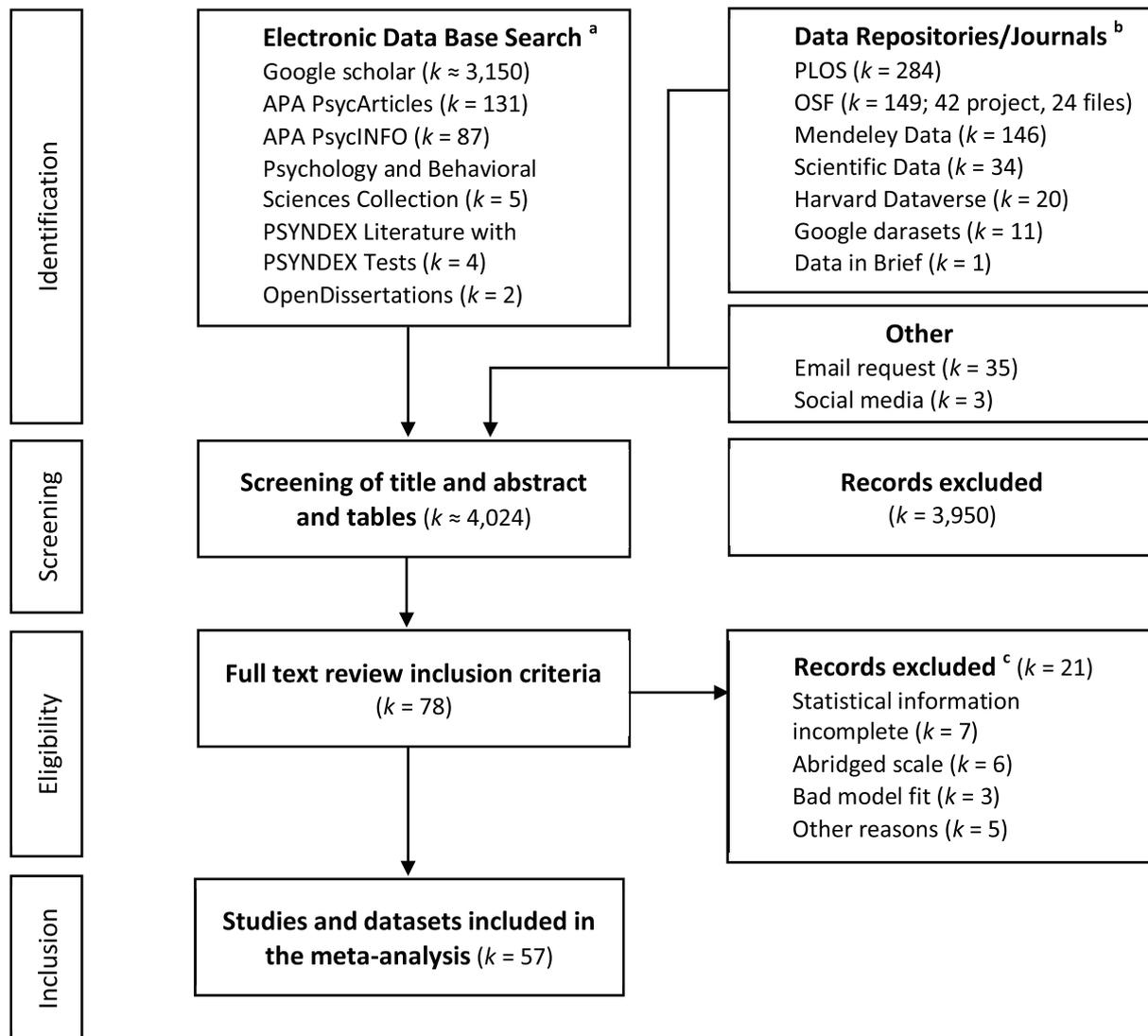
Figure 2. Overview of the literature search process. [a]The search term was a Boolean expression: ("NFC scale" OR "NC scale" OR "Need for cognition scale") AND ("factor analysis" OR "factor structure" OR "principal component analysis" OR "item analysis"). [b]For screening the data repositories the search term was reduced to "need for cognition" AND data. [c]For more detailed information on the reasons for exclusion see screening_studies.xlsx in the OSF deposit.

Cheung & Chan, 2005 for a full account). We first report the results of an exploratory factor analysis with oblimin rotation (δ = 0). Following the recommendations of Auerswald and Moshagen (2019), we used several criteria to decide on the number of factors to retain (e.g., Horn's parallel analysis, Bayesian information criteria, and sequential χ^2 model tests). The main focus, however, is on testing the competing measurement models by means of confirmatory factor analysis with a weighted least square estimator using the asymptotic variance-covariance matrix of the pooled correlations from the first step as weights (Cheung & Chan, 2005). In line with conventional standards (see Schermelleh-Engel et al., 2003) and current recommendations (Bader & Moshagen, 2022), the following cut-off criteria were used

as an indication of acceptable model fit: Comparative fit index (CFI) ≥ .95, non-normed fit index (NNFI; also known as Tucker-Lewis Index) ≥ .95, root mean square error of approximation (RMSEA) ≤ .08, and a standardized root mean square residual (SRMR) ≤ .10. Model fit was considered good for a CFI ≥ .97, NNFI ≥ .97, RMSEA ≤ .05, and SRMR ≤ .05. Additionally, the relational fit values Akaike information criterion (AIC) and Bayesian information criterion (BIC) were also provided. Our evaluation of model fit was not based on model fit indices alone but also took into account model complexity and the pattern/magnitude of factor loadings (see Heene et al., 2011). The significance of (nested) factors was determined with the reliability coefficient ω (Flora, 2020).

Sensitivity Analyses

We conducted several sensitivity analyses to examine the robustness of the correlation matrices: First, we used the correlation among the correlation matrices to find the studies that deviated most from all others. These outlier studies were excluded to study their impact on the pooled correlation matrix. Second, we examined the influence of a few very large sample-sized studies – that accounted for approximately 60% of the sample – on the meta-analytic results by splitting the database into two subsets (large vs. the remaining studies). Finally, we considered the study quality as another biasing influence that might affect the factor analytic results. Therefore, we weighted each correlation matrix by the inverse of the risk of bias score using a Gaussian kernel function (see also Hildebrandt et al., 2016), which means, high-quality studies were included in the recalculations with a proportionally larger sample size. Put differently, we reran the MASEM analyses for a set of hypothetical samples of the highest quality (see the supplement information in Gnambs & Schroeders, 2024, for details on this approach).

Ant Colony Optimization Procedure

The core of every metaheuristic search is the optimization function which is modular in design and often contains several criteria to evaluate the quality of the models. In the present context, we limit our examination to (a) model fit and (b) the reliability of the trait factor. Please note that in principle this optimization function can accommodate any quantifiable additional criteria such as maximizing validity (Steger et al., 2022) or cross-cultural measurement invariance (Jankowsky et al., 2020). All criteria were logit-transformed, on the one hand, to bring the values onto a common scale [0;1] and, on the other hand, to maximally differentiate between models around a preset cutoff value (Janssen et al., 2017; Schroeders et al., 2016a).

With respect to the first criterion, model fit, we used a combination of an incremental fit index, the comparative fit index (CFI $\geq$ .97), and an absolute fit index, the root mean square error of approximation (RMSEA $\leq$ .05) – as proposed with the two-index presentation strategy (Hu & Bentler, 1999). The inflection points of the logit functions were set to the cut-off values mentioned above as an indication of good model fit.

$$\varphi_{\text{CFI}} = \frac{1}{1 + \exp(100 \cdot (.97 - \text{CFI}))}, \qquad (1)$$

$$\varphi_{\text{RMSEA}} = \frac{1}{1 + \exp(-100 \cdot (.05 - \text{RMSEA}))}. \qquad (2)$$

With respect to the reliability of the scale, we used McDonald's ω (Flora, 2020) with values larger than .75 considered desirable.

$$\varphi_{\text{Rel}} = \frac{1}{1 + \exp(25 \cdot (.75 - \omega_{\text{NFC}}))}. \qquad (3)$$

Both criteria of model fit and reliability were combined equally in an overall optimization function:

$$\varphi_{\text{overall}} = \frac{\varphi_{\text{CFI}} + \varphi_{\text{RMSEA}} + \varphi_{\text{Rel}}}{3}. \qquad (4)$$

Furthermore, we systematically varied the number of items from 4 to 15, paying attention to a balance between positively and negatively worded items (in case of an odd item number, one positively worded item more was drawn). We tested the originally intended uni-dimensional measurement model and the acquiescence model to deal with the method variance introduced by the negatively worded items.

Results

Study Characteristics

The meta-analysis included 87 samples nested in 57 publications that were published between 1993 and 2023. The median sample size was Mdn = 354 participants (total N = 90,215; Min = 117, Max = 33,784) with approximately 59.3% women and a reported mean age of 29.3 years (SD = 9.5). The NCS-18 has been translated into several different languages (Catalan, Chinese, Dutch, Croatian, French, Greek, Icelandic, Indonesian, Korean, Portuguese, Romanian, Russian, Serbian, Spanish, and Turkish), but two-thirds of the samples included in our meta-analytic data set relied on the original English version, followed by Dutch (13 samples)[2] and Portuguese (4 samples). Most studies consisted of either undergraduate university students or online samples from crowdworking platforms (mostly MTurk). The three largest studies, which together accounted for approximately 60% of the total sample size, were an extensive survey conducted during the Catalan independence movement with a total of 33,784 participants (Barceló, 2023), the Dutch panel survey LISS (Longitudinal Internet Studies for Social science; Scherpenzeel & Das, 2010) with 13,503 participants, and the AIID study (Attitudes, Identities, and Individual Differences), which was conducted via the Project Implicit website with 6,851 participants (Hussey & Hughes, 2020). The study characteristics of all samples are given in Table 1.

---

[2] Ten out of the 13 Dutch samples were consecutive waves of the LISS panel (Longitudinal Internet Studies for Social science; Bruinsma & Crutzen, 2018). For the analyses, only those participants were included who were not already considered in previous waves (disjoint samples).

Exploratory Factor Analyses

The pooled correlations between the 18 NCS items showed moderate item correlations between .06 and .55 (Mdn = .29; for the pooled correlation matrix see ESM 1, Table E2). The different criteria that can be used to determine the number of factors in exploratory factor analysis (Auerswald & Moshagen, 2019; Ruscio & Roche, 2012) came to the same conclusion: The empirical Kaiser criterion (EKC; Braeken & van Assen, 2017), the Hull method (Lorenzo-Seva et al., 2011), the minimum average partial method (MAP; Velicer, 1976), and Horn's parallel analyses (PA_{PCA}; Garrido et al., 2013) all suggested a two-factor solution. Also, the sequential χ^2 model tests came to the same conclusion when the sample size was reduced to the usual orders of magnitude ($n < 1{,}000$). Accordingly, we extracted two factors in an exploratory factor analysis with oblimin rotation. The two-factor structure reflected the division into negatively and positively worded items with high factor loadings on the corresponding factors (Mdn_{pos} = .63; Mdn_{neg} = .60), whereas the cross-loadings were close to zero (Mdn = .00, Min = −.09, Max = .07). The factors were substantially correlated at r = .59, but far from unity. Parameter estimates of the uni- and the two-dimensional solution are listed in ESM 1, Table E3.

Confirmatory Factor Analyses

We estimated six measurement models (see Figure 1) based on the pooled correlation matrix to examine which of the different psychometric modeling approaches adequately captured the structure of the NCS-18 (model fit values are listed in Table 2, the parameter estimates of all models are provided in ESM 1, Table E4). The confirmatory factor analyses showed that the uni-dimensional model with a single general factor for all items did not adequately describe the empirical data with respect to the above-mentioned cutoff values. Although all items had substantial loadings on the latent factor (Mdn = .56, Min = .29, Max = .67), the absolute model fit indices (RMSEA and SRMR) were insufficient. In contrast, all other models that accounted for the method effects related to item wording showed good model fit values. More specifically, the two-factor model with two separate factors for the positively and negatively worded items (Model 2), the correlated uniqueness model with residual correlations between all negatively worded items (Model 3), the two bifactor models accounting for method variance without and with reference factor (Models 4 and 5), and the acquiescence model (Model 6) only slightly differed in terms of model fit values. Taking a closer look at the models' assumptions, one would rule out the two-dimensional model because the two factors do not represent distinct facets. Moreover, in many applied

settings one is interested in calculating a single-person estimate for NFC. The correlated-uniqueness model also has substantive shortcomings (Conway et al., 2004; Lance et al., 2002), because among other things the amount of individual bias cannot be quantified, and it is less parsimonious than the competing models.

Considering model complexity, the bifactor with two orthogonal method factors for the different item wording besides a trait factor performed best, as shown by the highest NNFI and the lowest BIC. However, what is problematic about this solution is the range and magnitude of the factor loadings on the method factor for negatively keyed items which varied between -.09 and .30 (see also Table S4). To overcome this inconsistent loading pattern indicating over factorization, we estimated a bifactor (S−1) model with the positively keyed items of the NCS-18 set as a reference (Eid et al., 2017). In this model, all factor loadings on the nested method factor for negatively keyed items were more pronounced (M_{neg} = 0.49) and even slightly larger than the loadings of these items on the trait factor ($M_{NFC/neg}$ = 0.36). These results indicate that negatively keyed items functioned somewhat differently as compared to positively keyed items. They captured non-negligible, unique variance in addition to the common trait.

The acquiescence model which captures the tendency to agree to an item independent of its content (Ferrando & Lorenzo-Seva, 2010) is in line with the underlying uni-dimensional conceptualization of the NCS-18 while simultaneously addressing response style bias. Kam and Zhou (2015) have argued that the acquiescence model is based on the assumption that all items are equally affected by response bias, which is hard to test. Taking the model fit values, the factor loading patterns, and model parsimony into account, we recommend the acquiescence model to represent the structure of the NCS-18. In the subsequent short-scale construction via ACO, we compare the uni-dimensional model (with its insufficient model fit for the long form) and the acquiescence model.

Short-Scale Construction With Ant Colony Optimization

Figure 3 shows the incremental model fit index CFI (Figure 3A) and the reliability coefficient ω (Figure 3B) for short scales with 4–15 items. Figure 3 shows the results for a uni-dimensional measurement model that had poor model fit in the long version (blue lines) and the results for the acquiescence model (black line). In both models, the number of positively and negatively worded items was balanced (in the case of an odd number of items, one positively formulated item more was drawn). The results of the short scales compiled via MASEM-ACO (solid

Table 1. Overview of samples and coded data

Study	Year	Country	Language	N	Women	Age	Data	Outlier	RoB
Alarcon & Lee	2022	USA	English	741	38.00	36.60	Raw data	No	3.5
Bakker et al.	2020	USA	English	1,209	57.99	41.00	Raw data	No	5.0
	2020	USA	English	1,324	51.91	47.77	Raw data	No	5.0
Barceló	2023	Spain	Spanish, Catalan	33,784	–	–	Correlations	No	4.0
Broniatowski et al.	2023	USA	English	1,762	38.90	38.90	Raw data	No	5.0
Calloway et al.	2023	–	English	204	45.10	31.96	Raw data	No	4.5
	2023	USA	English	187	65.78	19.09	Raw data	No	4.5
Cartwright et al.	2009	USA	English	307	83.67	–	Correlations	No	6.0
Cazan	2016	Romania	Romanian	297	72.39	18.16	EFA loadings	No	6.0
Clay et al.	2022	USA	English	187	63.10	40.56	Raw data	No	5.0
Culhane et al.	2004	USA	English	237	58.40	20.17	Correlations	No	4.0
	2004	USA	English	185	65.10	19.62	CFA loadings	No	5.5
Culhane et al.	2006	USA	English	289	59.80	19.33	Correlations	No	3.0
	2006	USA	English	175	64.80	19.61	Correlations	No	4.0
Damer et al.	2019	USA	English	315	42.54	29.87	Raw data	No	4.0
	2019	UK	English	288	68.40	24.20	Raw data	No	4.0
	2019	UK	English	262	63.36	31.66	Raw data	No	4.0
DeMarree et al.	2020	USA	English	859	62.70	19.34	Raw data	No	5.0
	2020	USA	English	482	47.72	19.51	Raw data	No	5.0
	2020	USA	English	551	71.69	48.29	Raw data	No	5.0
Dennin et al.	2022	–	English	420	52.29	–	Raw data	No	5.5
Ebersole et al.	2017	USA	English	152	59.21	18.68	Raw data	No	6.0
	2017	USA	English	117	70.69	19.37	Raw data	No	6.0
	2017	USA	English	182	71.19	18.87	Raw data	No	6.0
	2017	USA	English	174	71.68	19.70	Raw data	No	6.0
	2017	USA	English	189	75.13	20.11	Raw data	No	6.0
	2017	USA	English	178	80.34	18.74	Raw data	No	6.0
	2017	USA	English	140	86.33	21.33	Raw data	No	6.0
Eck & Gebauer	2022	USA	English	2,542	57.53	35.53	Raw data	No	4.0
Edwards	2009	USA	English	3,364	60.00	–	Correlations	No	6.0
Elias & Loomis	2002	USA	English	138	61.59	19.00	EFA loadings	No	6.0
Gomes et al.	2013	Portugal	Portuguese	344	51.20	28.55	Correlations	No	4.0
Gouveia et al.	2015	Brazil	Portuguese	215	61.00	20.00	Correlations	No	5.0
Grădinaru et al.	2023	Romania	Romanian	747	75.30	37.71	Correlations	No	4.0
Gústavsson et al.	2020	Iceland	Icelandic	602	67.28	–	EFA loadings	No	4.5
Hallahan	2009	USA	English	329	56.53	–	EFA loadings	No	7.0
Hanel & Wolf	2020	UK	English	476	53.36	38.91	Raw data	No	4.5
Hussey & Hughes	2020	–	English	6,851	64.63	31.57	Raw data	No	5.0
Janssen et al.	2020	Netherlands	Dutch	930	55.00	18.84	Raw data	No	3.5
	2020	Netherlands	Dutch	509	50.00	18.82	Raw data	No	3.5
Jin	2016	South Korea	Korean	288	–	–	EFA loadings	Yes	7.5
Karagiannopoulou et al.	2020	Greece	Greek	406	86.30	–	Correlations	No	5.0
Koutsogiorgi	2020	Cyprus	Greek	409	–	–	Raw data	No	6.0
Lantos & Harris	2021	–	English	248	60.17	33.91	Raw data	No	4.0
Laroche et al.	2009	Canada, Hong Kong, Taiwan	English, Chinese	373	57.10	–	EFA loadings	No	7.0
Lee et al.	2020	USA	English	502	–	–	Raw data	No	3.0

(Continued on next page)

Table 1. (Continued)

Study	Year	Country	Language	N	Women	Age	Data	Outlier	RoB
Lins de Holanda	2020	USA	English	821	45.07	32.12	EFA loadings	No	6.0
Coelho et al.	2020	UK	English	476	46.43	38.91	EFA loadings	No	7.0
LISS-Wave 1	2008	Netherlands	Dutch	6,755	54.23	46.01	Raw data	No	2.5
LISS-Wave 2	2009	Netherlands	Dutch	434	50.69	33.96	Raw data	No	3.0
LISS-Wave 3	2010	Netherlands	Dutch	1,366	55.49	48.25	Raw data	No	3.0
LISS-Wave 4	2011	Netherlands	Dutch	194	50.78	31.88	Raw data	No	4.0
LISS-Wave 5	2012	Netherlands	Dutch	1,154	53.91	45.36	Raw data	No	3.0
LISS-Wave 6	2013	Netherlands	Dutch	172	56.14	28.08	Raw data	No	4.0
LISS-Wave 7	2014	Netherlands	Dutch	1,537	54.17	41.02	Raw data	No	3.0
LISS-Wave 8	2015	Netherlands	Dutch	219	47.95	27.97	Raw data	No	4.0
LISS-Wave 9	2017	Netherlands	Dutch	1,179	56.23	43.09	Raw data	No	3.0
LISS-Wave 10	2018	Netherlands	Dutch	493	57.76	37.43	Raw data	No	3.0
Loose et al.	2023	Uruguay	Spanish	357	56.02	23.12	Raw data	No	5.0
Ludwig et al.	2018	USA	English	410	49.00	36.41	Raw data	No	4.5
Luong & Lomanowska	2022	–	English	532	44.09	–	Raw data	No	4.0
Maldonado et al.	1993	Spain	Spanish	232	57.76	23.97	EFA loadings	Yes	6.0
Malmberg	2010	Netherlands	Dutch	146	60.47	30.00	EFA loadings	Yes	7.0
Menendez et al.	2023	USA	English	240	65.83	–	Raw data	No	5.0
Minson et al.	2020	–	English	202	53.47	39.93	Raw data	No	7.0
	2020	–	English	467	50.11	37.73	Raw data	No	4.0
	2020	–	English	233	48.50	37.31	Raw data	No	3.5
Newman et al.	2020	USA	English	325	–	–	Raw data	No	5.5
	2020	USA	English	392	50.00	36.56	Raw data	No	5.0
	2020	USA	English	563	–	35.23	Raw data	No	4.5
Park	2012	USA	English	425	100.00	–	EFA loadings	No	6.5
Petrović & Žeželj	2022	Serbia	Serbian	167	78.40	21.16	Raw data	No	4.5
Pilli & Mazzon	2016	Brazil	Portuguese	1,008	48.40	–	Raw data	No	5.0
Powell et al.	2016	Australia	English	396	63.89	20.20	Correlations	No	3.5
Pryor et al.	2000	USA	English	559	–	–	EFA loadings	No	4.0
Salama-Younes et al.	2014	France	French	123	–	–	EFA loadings	Yes	8.0
Shchebetenko	2011	Russia	Russian	482	65.90	20.83	CFA loadings	Yes	7.5
Sousa et al.	2018	Portugal	Portuguese	442	33.94	–	Correlations	Yes	6.0
Tobin & Guadagno	2022	–	English	308	48.01	27.78	Raw data	No	4.0
Türker et al.	2015	Turkey	Turkish	143	47.60	–	EFA loadings	Yes	7.0
Van Tilburg et al.	2019	USA	English	320	52.07	35.01	Raw data	No	5.0
Vaughan-Johnston & Jacobson	2020	Canada	English	201	84.65	18.10	Raw data	No	6.0
	2020	Canada	English	335	69.34	18.10	Raw data	No	5.0
Weigold & Weigold	2022	USA	English	1,111	46.51	–	Correlations	No	4.0
Weng & DeMarree	2019	USA	English	354	44.92	19.15	Correlations	No	5.0
Yamamoto & Maeder	2019	Canada	English	169	46.90	36.00	Raw data	No	4.0
Zhang et al.	2016	Canada	English	303	81.03	19.97	Raw data	No	5.0

Note. Women = Percentage of women; Age = Mean age of sample; Data = Data type for meta-analysis; RoB = Risk of bias assessment (adapted version of Nudelman & Otto, 2020), higher values indicate a higher risk. For additional information on the NCS data from the LISS panel (Longitudinal Internet Studies for Social science) see Bruinsma and Crutzen (2018).

lines) were compared to the best model of 10 randomly selected short scales (dashed lines). The more sophisticated modeling of the acquiescence model yielded excellent model fit with CFI values above .985, regardless of the length of the short scale. In contrast, the uni-dimensional models with more than five items did not describe the data sufficiently well. Although there was a slight advantage of ACO over a random selection, the item pool was apparently

Table 2. Goodness of fit statistics for different meta-analytic confirmatory factor models of the NCS-18

No	Model	χ^2	df	CFI	NNFI	RMSEA	SRMR	AIC	BIC
1	Uni-dimensional model	9,515.5	135	.828	.805	.028 [.027; .028]	.081	9,245.5	7,975.1
2	Two-dimensional model	1,486.4	134	.975	.972	.011 [.010; .011]	.025	1,218.4	−42.5
3	Correlated uniqueness model	927.7	99	.985	.976	.010 [.009; .010]	.019	729.7	−201.9
4	Bifactor model with two method factors	895.2	117	.986	.981	.009 [.008; .009]	.019	661.2	−439.8
5	Bifactor (S−1) model for the neg. keyed items	1,243.5	126	.979	.975	.010 [.009; .010]	.023	991.5	−194.2
6	Acquiescence model	1,242.0	134	.980	.977	.010 [.009; .010]	.023	974.0	−286.9

Note. N = 90,215. k = 87. CFI = comparative fit index; NNFI = non-normed fit index; RMSEA = root mean square error of approximation; SRMR = standardized root mean residual; AIC = Akaike information criterion; BIC = Bayesian information criterion.
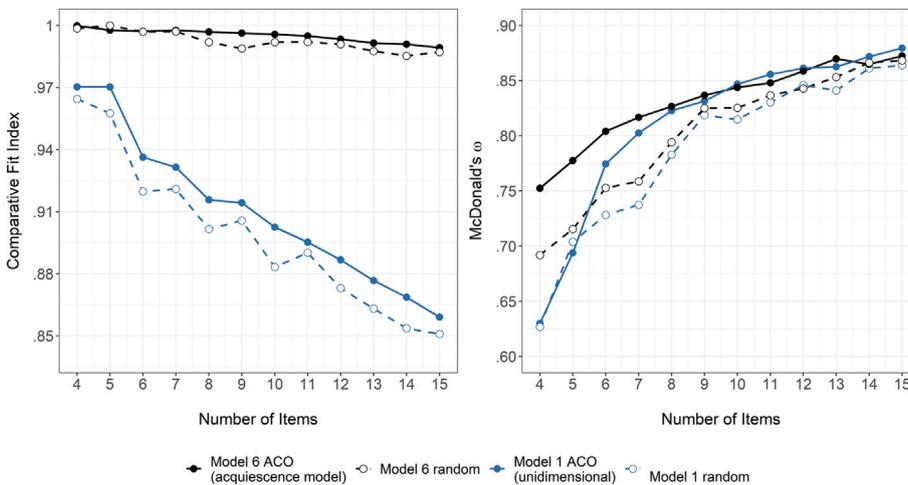
Figure 3. Comparison of short scales derived via MASEM-ACO vs. randomly selected. The dashed lines represent the best of 10 randomly drawn short scales.

not diverse and large enough to compensate for the incorrect modeling of the uni-dimensional model. As expected, the reliability coefficient increased with an increasing number of items; if at least half of the items were selected, values of ω were larger than .80. Again, there are small advantages of ACO selection over a random selection for the reliability coefficients. The acquiescence model achieved higher reliability values than the uni-dimensional model because the bias is separated from the trait. For an overview of the items included in the short versions of the acquiescence model, see ESM 1, Table E5.

Sensitivity Analyses

The influence of the outlier studies on the pooled correlations matrix was small. The differences in the pooled correlations as compared to the full sample (Min = −.015, Max = .012) were unsystematic around 0 (M = 0, Mdn = 0), which is why we have abstained from repeating the confirmatory factor analyses (see ESM 1, Figure E2). We reran the main analyses separately for the three big studies (Barceló, 2023; Hussey & Hughes, 2020; Scherpenzeel & Das, 2010) versus all other studies. Although there were differences in the pooled correlation matrices (Min = −.170, Max = .113), their average was also close to 0 (M = −0.029,

Mdn = −.025, see ESM 1, Table E6). The results of the measurement models were similar (see ESM 1, Table E7) and did not change any conclusions drawn. Finally, although the risk of bias for the included studies varied considerably (see last column in Table 1), controlling for the study quality did not affect the factor analytic results. Figure 4 shows that the pooled correlations and factor loadings of the acquiescence model were similar, regardless of whether we controlled for the study quality or not. The average difference in factor loadings between the two analyses was small (M = −0.012; range: −0.029 to 0.002), indicating that differences in the quality of scientific reporting did not affect the statistics that underlie the results of the present meta-analysis (for quite similar results see Gnambs & Schroeders, 2024).

Discussion

NFC is the tendency "to seek, acquire, think about, and reflect back on information to make sense of stimuli, relationships, and events in their world" (Cacioppo et al., 1996, p. 198) is a personality trait often considered in psychological research because it helps in explaining why people under the same circumstances decide or behave
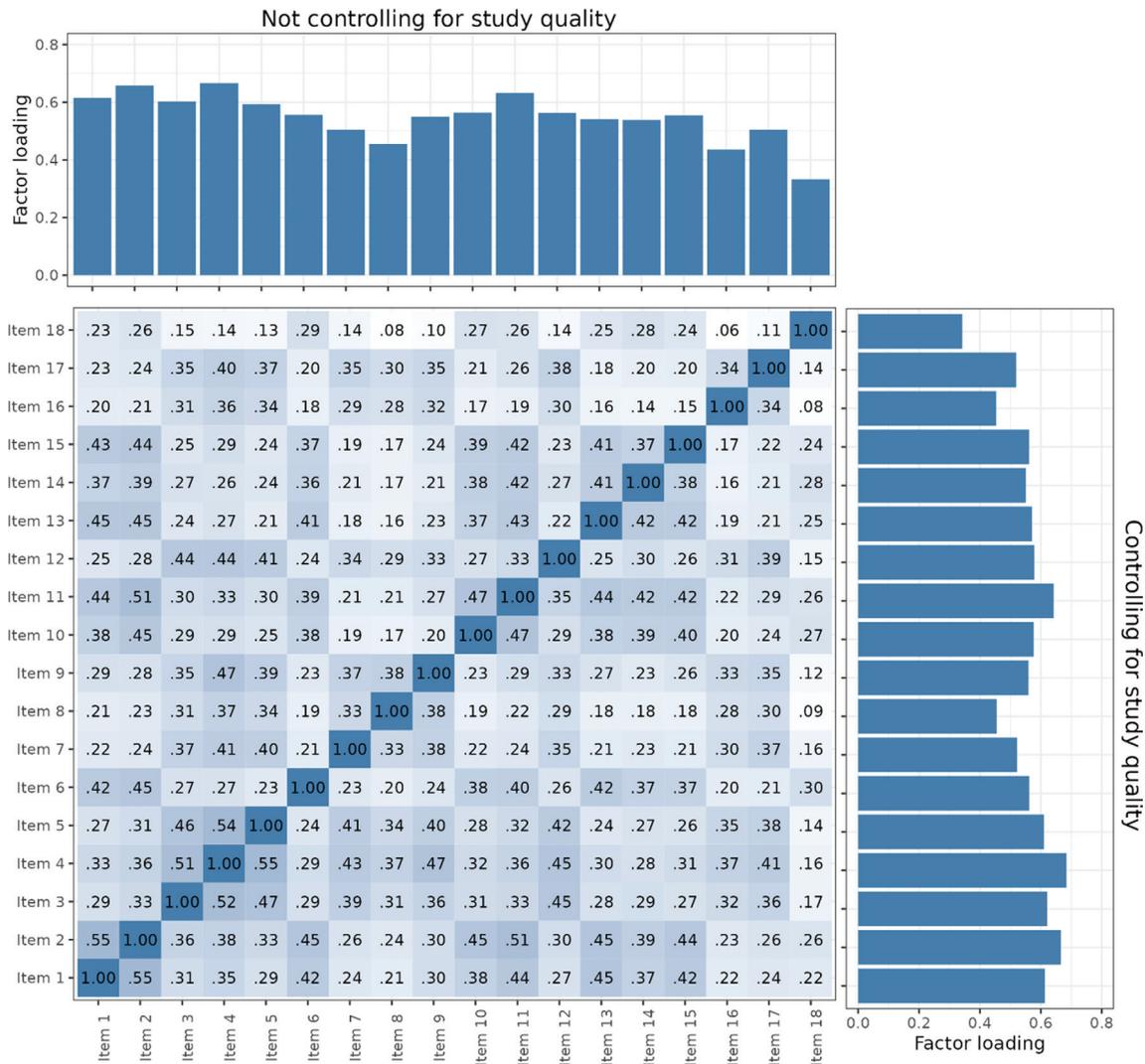
Figure 4. Pooled correlations and factor loadings for the NCS-18 with and without controlling for study quality. Presented are pooled correlations between the items of the NCS-18 and factor loadings of the meta-analytic acquiescence model. Upper-diagonal results on the top do not control for study quality, while lower diagonal results on the right control for study quality.

differently. It is closely related to typical intellectual engagement, openness to new ideas, and epistemic curiosity, which is why these constructs are grouped in the Seek-Think cluster of Mussel's Intellect Framework (Mussel, 2013). The most popular scale to measure NFC is by far the NCS-18. Concerning the internal structure of the scale, uni-dimensionality is often assumed, but rarely tested with confirmatory factor analyses. In the present meta-analysis, we gathered all available raw data, correlation matrices, and factor loading matrices to address the question of the best psychometric modeling. It is also the first psychometric review of the NCS-18 in which bifactor models and the acquiescence model have been compared. One of the main findings was that the consideration of specific method variance for the negatively formulated items was essential since a uni-dimensional measurement model did not adequately describe the data. Although all models accounting for this

method-related variance achieved similar good fit values, we prefer the acquiescence model because it is both parsimonious (in comparison to the correlated uniqueness and the bifactor models) and yielded a more sensible pattern of factor loadings (in comparison to the bifactor models).

Limitations and Future Research

Some limitations of the present meta-analysis have to be taken into account: First, the recovery of population factors in individual studies can be impeded by sampling error (MacCallum et al., 2001). Although pooling results across diverse samples should provide more robust inferences on the population factor structure, the meta-analytical basis is decisive for the quality of the conclusions drawn. The high proportion of online studies (see Table 1), mostly conducted via the panel provider MTurk, was striking,

despite the known limitations (e.g., Douglas et al., 2023; Kennedy et al., 2020). One study included in our meta-analysis (Weigold & Weigold, 2022) directly compared the results of convenience samples commonly used in psychological research (i.e., traditional college students, MTurk workers, and an MTurk sample of college students), and found various differences between the samples, not only in the sample characteristics but also in the correlations at the scale level and at the item level. On the one hand, the sometimes-reported low study quality of online studies is even more alarming because many studies included in the present meta-analysis did not report any cleaning procedures (ca. 70% of the studies according to item 8 of the risk of bias scale). On the other hand, we found only negligible differences in the correlation matrix depending on the study quality. Second, we could not investigate moderation effects because, despite the large database, it was not possible to look more closely at language differences, age differences, or the effects of the assessment setting.

We think that the present article could initiate further research because it introduces a new method that can help to shorten measurement instruments, a frequent demand in large-scale psychological assessment (Kruyen et al., 2013). For this, MASEM-ACO uses meta-analytic aggregation of statistical information across a large number of studies, to then use the pooled correlation matrix as a starting point for scale abbreviation via ACO. The present study is to be understood as a proof-of-concept because for the NCS-18 even a complete search would have been possible. The first reason why an exhaustive search is possible is that the measure is already relatively short: For example, for the ten-item versions, there are only 15,876 = $\binom{9}{5} \cdot \binom{9}{5}$ models, if the same number of positively and negatively keyed items are drawn. However, if the number of items is larger, the number of models to be estimated increases exponentially, a phenomenon which is known as combinatorial explosion. For example, selecting 18 items out of 34 items, as was done for the NCS-18, is combinatorically more demanding: 590,976,100 = $\binom{17}{9} \cdot \binom{17}{9}$. With increasing item numbers, ACO can increasingly demonstrate its advantages in short-scale construction.

The second reason why a complete search is possible in the case of the NCS-18 is that the questionnaire's structure is quite clear compared to other scales. If the dimensional structure is unclear, another metaheuristic that was recently introduced – Bee Swarm Optimization (BSO) – might be helpful in finding the dimensional structure while simultaneously selecting items for the final scale (Schroeders et al., 2023). The authors outlined the advantages of the BSO over and above traditional methods (e.g., exploratory factor analysis with sequential item selection) and demonstrated its usefulness in finding the underlying structure in two empirical data sets. Possible candidates for such a meta-analytic BSO investigation would be the 27-item Short Dark Triad Questionnaire (SD3; Jones & Paulhus, 2014) or the Actively Open-minded Thinking scale (AOT; Stanovich & West, 2007).

From a meta-analytic perspective, one might object that it is unlikely to obtain sufficient data such as correlation matrices and factor loadings when more than 20 items are presented. This may be true for aggregate data; however, the Open Science movement has significantly improved the availability of raw data (Hardwicke et al, 2021; Nosek et al., 2022). This paradigm shift in sharing raw data is also reflected in the literature search of the present study: Whereas only three studies (with 10 samples) provided raw data before 2017, there is a sharp increase thereafter (28 studies with 47 samples). This cultural change in the research community has once and for all altered the database on which to rely in meta-analyses, which is why meta-analyses will continue to thrive in the future.

Conclusion

In the present meta-analysis, we compared competing measurement models for the 18-item NCS using summary data of 87 samples (N = 90,215). After considering content-related and various psychometric criteria such as loading patterns, model fit, and parsimony, we found that an acquiescence model appeared particularly well-suited to account for the method variance caused by the negatively keyed items (which also hinders a uni-dimensional model). This study is the first to combine meta-analysis with metaheuristics (MASEM-ACO), thereby providing a flexible and promising new tool for the method toolbox of psychometricians. Leveraging the growing meta-analytic database at the item level, MASEM-ACO enables the construction of sound short scales.

Electronic Supplementary Materials

The following electronic supplementary material is available with this article at https://doi.org/10.1027/1015-5759/a000818

ESM 1. The ESM includes the adapted Risk of Bias scale (Table E1), the pooled correlation matrix (Table E2), the meta-analytic EFA loading pattern (Table E3), the parameter estimates of the NCS-18 (Table E4), the compiled short version using MASEM-ACO (Table E5), additional sensitivity checks (Table E6, Table E7, and Figure E2), and the correlation matrix of all correlation matrices (Figure E1).

References

References marked with * were included in the meta-analysis.

Aichholzer, J. (2014). Random intercept EFA of personality scales. *Journal of Research in Personality, 53*, 1–4. https://doi.org/10.1016/j.jrp.2014.07.001

*Alarcon, G. M., & Lee, M. A. (2022). The relationship of insufficient effort responding and response styles: An online experiment. *Frontiers in Psychology, 12*, Article 784375. https://doi.org/10.3389/fpsyg.2021.784375

Auerswald, M., & Moshagen, M. (2019). How to determine the number of factors to retain in exploratory factor analysis: A comparison of extraction methods under realistic conditions. *Psychological Methods, 24*(4), 468–491. https://doi.org/10.1037/met0000200

Bader, M., & Moshagen, M. (2022). Assessing the fitting propensity of factor models. *Psychological Methods*. Advance online publication. https://doi.org/10.1037/met0000529

*Bakker, B. N., Lelkes, Y., & Malka, A. (2020). Understanding partisan cue receptivity: Tests of predictions from the bounded rationality and expressive utility perspectives. *The Journal of Politics, 82*(3), 1061–1077. https://doi.org/10.1086/707616

*Barceló, J. (2023). Need for affect, need for cognition, and the desire for independence. *PLoS One, 18*(2), Article e0280457. https://doi.org/10.1371/journal.pone.0280457

Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling: A Multidisciplinary Journal, 7*(4), 608–628. https://doi.org/10.1207/S15328007SEM0704_5

Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., & Schwarz, N. (1994). Need for Cognition: Eine Skala zur Erfassung von Engagement und Freude bei Denkaufgaben [Need for Cognition: A scale measuring engagement and happiness in cognitive tasks]. *Zeitschrift für Sozialpsychologie, 25*, 147–154.

Bors, D. A., Vigneau, F., & Lalande, F. (2006). Measuring the need for cognition: Item polarity, dimensionality, and the relation with ability. *Personality and Individual Differences, 40*, 819–828. https://doi.org/10.1016/j.paid.2005.09.007

Braeken, J., & van Assen, M. A. (2017). An empirical Kaiser criterion. *Psychological Methods, 22*(3), 450–466. https://doi.org/10.1037/met0000074

*Broniatowski, D., Hosseini, P., Porter, E., & Wood, T. J. (2023). *The role of mental representation in sharing misinformation online*. PsyArXiv. https://doi.org/10.31234/osf.io/htkr7

Bruinsma, J., & Crutzen, R. (2018). A longitudinal study on the stability of the need for cognition. *Personality and Individual Differences, 127*, 151–161. https://doi.org/10.1016/j.paid.2018.02.001

Burchert, S., Kerber, A., Zimmermann, J., & Knaevelsrud, C. (2021). Screening accuracy of a 14-day smartphone ambulatory assessment of depression symptoms and mood dynamics in a general population sample: Comparison with the PHQ-9 depression screening. *PLoS One, 16*(1), Article e0244955. https://doi.org/10.1371/journal.pone.0244955

Cacioppo, J. T., & Petty, R. E. (1982). The need for cognition. *Journal of Personality and Social Psychology, 42*, 116–131. https://doi.org/10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., & Feng Kao, C. (1984). The efficient assessment of need for cognition. *Journal of Personality Assessment, 48*(3), 306–307. https://doi.org/10.1207/s15327752jpa4803_13

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., & Jarvis, W. B. G. (1996). Dispositional differences in cognitive motivation: The life and times of individuals varying in need for cognition. *Psychological Bulletin, 119*(2), 197–253. https://doi.org/10.1037/0033-2909.119.2.197

*Calloway, R. C., Helder, A., & Perfetti, C. A. (2023). A measure of individual differences in readers' approaches to text and its relation to reading experience and reading comprehension. *Behavior Research Methods, 55*(2), 899–931. https://doi.org/10.3758/s13428-022-01852-1

*Cartwright, K. B., Galupo, M. P., Tyree, S., & Jennings, J. L. (2009). Reliability and validity of the complex postformal thought questionnaire: Assessing adults' cognitive development. *Journal of Adult Development, 16*(3), 183–189. https://doi.org/10.1007/s10804-009-9055-1

*Cazan, A. M. (2016). The factor structure of the short Need for Cognition Scale. *Bulletin of the Transilvania University of Braşov, 9*(58), 19–28. https://webbut.unitbv.ro/index.php/Series_VII/article/view/3757

Cheung, M. W.-L. (2013). Multivariate meta-analysis as structural equation modeling. *Structural Equation Modeling, 20*(3), 429–454. https://doi.org/10.1080/10705511.2013.797827

Cheung, M. W.-L. (2014). Fixed- and random-effects meta-analytic structural equation modeling: examples and analyses in R. *Behavior Research Methods, 46*(1), 29–40. https://doi.org/10.3758/s13428-013-0361-y

Cheung, M. W.-L., & Chan, W. (2005). Meta-analytic structural equation modeling: A two-stage approach. *Psychological Methods, 10*(1), 40–64. https://doi.org/10.1037/1082-989X.10.1.40

Cheung, M. W.-L., & Cheung, S. F. (2016). Random-effects models for meta-analytic structural equation modeling: Review, issues, and illustrations. *Research Synthesis Methods, 7*(2), 140–155. https://doi.org/10.1002/jrsm.1166

Chiesi, F., Morsanyi, K., Donati, M. A., & Primi, C. (2018). Applying item response theory to develop a shortened version of the Need for Cognition Scale. *Advances in Cognitive Psychology, 14*(3), 75–86. https://doi.org/10.5709/acp-0240-z

*Clay, G., Dumitrescu, C., Habenicht, J., Kmiecik, I., Musetti, M., & Domachowska, I. (2022). Who is satisfied with effort? Individual differences as determinants of satisfaction with effort and reward. *European Journal of Psychological Assessment, 38*(6), 452–462. https://doi.org/10.1027/1015-5759/a000742

Cohen, A. R., Stotland, E., & Wolfe, D. M. (1955). An experimental investigation of need for cognition. *The Journal of Abnormal and Social Psychology, 51*(2), 291–294. https://doi.org/10.1037/h0042761

Colling, J., Wollschläger, R., Keller, U., Preckel, F., & Fischbach, A. (2022). Need for cognition and its relation to academic achievement in different learning environments. *Learning and Individual Differences, 93*, 1–14. https://doi.org/10.1016/j.lindif.2021.102110

Conway, J. F., Lievens, F., Scullen, S. E., & Lance, C. E. (2004). Bias in the correlated uniqueness model for MTMM data. *Structural Equation Modeling, 11*(4), 535–559. https://doi.org/10.1207/s15328007sem1104_3

*Culhane, S. E., Morera, O. F., & Hosch, H. M. (2004). The factor structure of the need for cognition short form in a Hispanic sample. *The Journal of Psychology, 138*(1), 77–90. https://doi.org/10.3200/jrlp.138.1.77-90

*Culhane, S. E., Morera, O. F., & Watson, P. J. (2006). The assessment of factorial invariance in need for cognition using Hispanic and Anglo samples. *The Journal of Psychology, 140*(1), 53–67. https://doi.org/10.3200/jrlp.140.1.53-67

*Damer, E., Webb, T. L., & Crisp, R. J. (2019). Diversity may help the uninterested: Evidence that exposure to counter-stereotypes promotes cognitive reflection for people low (but not high) in need for cognition. *Group Processes & Intergroup Relations, 22*(8), 1079–1093. https://doi.org/10.1177/1368430218811250

*DeMarree, K. G., Petty, R. E., Briñol, P., & Xia, J. (2020). Documenting individual differences in the propensity to hold attitudes with certainty. *Journal of Personality and Social*

Psychology, 119(6), 1239–1265. https://doi.org/10.1037/pspa0000241

Deneubourg, J. L., Pasteels, J. M., & Verhaeghe, J. C. (1983). Probabilistic behaviour in ants: A strategy of errors? *Journal of Theoretical Biology, 105*(2), 259–271. https://doi.org/10.1016/S0022-5193(83)80007-1

Deneubourg, J.-L., Aron, S., Goss, S., & Pasteels, J. M. (1990). The self-organizing exploratory pattern of the argentine ant. *Journal of Insect Behavior, 3*(2), 159–168. https://doi.org/10.1007/BF01417909

*Dennin, A., Furman, K., Pretz, J. E., & Roy, M. J. (2022). The relationship of types of intuition to thinking styles, beliefs, and cognitions. *Journal of Behavioral Decision Making, 35*(5), Article e2283. https://doi.org/10.1002/bdm.2283

DeSteno, D., Petty, R. E., Rucker, D. D., Wegener, D. T., & Braverman, J. (2004). Discrete emotions and persuasion: The role of emotion-induced expectancies. *Journal of Personality and Social Psychology, 86*(1), 43–56. https://doi.org/10.1037/0022-3514.86.1.43

DiStefano, C., & Motl, R. W. (2006). Further investigating method effects associated with negatively worded items on self-report surveys. *Structural Equation Modeling, 13*(3), 440–464. https://doi.org/10.1207/s15328007sem1303_6

Dörendahl, J., & Greiff, S. (2020). Are the machines taking over? Benefits and challenges of using algorithms in (short) scale construction. *European Journal of Psychological Assessment, 36*(2), 217–219. https://doi.org/10.1027/1015-5759/a000597

Dorigo, M., & Stützle, T. (2010). Ant colony optimization: Overview and recent advances. In M. Gendreau & J.-Y. Potvin (Eds.), *Handbook of metaheuristics* (pp. 227–263). Springer.

Douglas, B. D., Ewell, P. J., & Brauer, M. (2023). Data quality in online human-subjects research: Comparisons between MTurk, Prolific, CloudResearch, Qualtrics, and SONA. *PLoS One, 18*(3), Article e0279720. https://doi.org/10.1371/journal.pone.0279720

*Ebersole, C. R., Alaei, R., Atherton, O. E., Bernstein, M. S., Brown, M., Chartier, C. R., Chung, L., Hermann, A. D., Joy-Gaba, J. A., Line, M. J., Rule, N. O., Sacco, D. F., Vaughn, L. A., & Nosek, B. A. (2017). Observe, hypothesize, test, repeat: Luttrell, Petty and Xu (2017) demonstrate good science. *Journal of Experimental Social Psychology, 69*, 184–186. https://doi.org/10.1016/j.jesp.2016.12.005

*Eck, J., & Gebauer, J. E. (2022). A sociocultural norm perspective on Big Five prediction. *Journal of Personality and Social Psychology, 122*(3), 554–575. https://doi.org/10.1037/pspp0000387

*Edwards, M. C. (2009). An introduction to item response theory using the Need for Cognition Scale. *Social and Personality Psychology Compass, 3*(4), 507–529. https://doi.org/10.1111/j.1751-9004.2009.00194.x

Eid, M., Geiser, C., Koch, T., & Heene, M. (2017). Anomalous results in G-factor models: Explanations and alternatives. *Psychological Methods, 22*(3), 541–562. https://doi.org/10.1037/met0000083

*Elias, S. M., & Loomis, R. J. (2002). Utilizing need for cognition and perceived self-efficacy to predict academic performance. *Journal of Applied Social Psychology, 32*(8), 1687–1702. https://doi.org/10.1111/j.1559-1816.2002.tb02770.x

Ferrando, P. J., & Lorenzo-Seva, U. (2010). Acquiescence as a source of bias and model and person misfit: A theoretical and empirical analysis. *British Journal of Mathematical and Statistical Psychology, 63*(2), 427–448. https://doi.org/10.1348/000711009X470740

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using r to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science, 3*(4), 484–501. https://doi.org/10.1177/2515245920951747

Forsterlee, R., & Ho, R. (1999). An examination of the short form of the Need for Cognition Scale applied in an Australian sample. *Educational and Psychological Measurement, 59*(3), 471–480. https://doi.org/10.1177/00131649921969983

Garrido, L. E., Abad, F. J., & Ponsoda, V. (2013). A new look at Horn's parallel analysis with ordinal variables. *Psychological Methods, 18*(4), 454–474. https://doi.org/10.1037/a0030005

Georgiou, Y., & Kyza, E. A. (2017). Translation, adaptation, and validation of the need for cognition scale – short form in the greek language for secondary school students. *Journal of Psychoeducational Assessment, 36*(5), 523–531. https://doi.org/10.1177/0734282916686005

Ginet, A., & Py, J. (2000). Le besoin de cognition: Une échelle française pour enfants et ses conséquences au plan sociocognitif [Need for cognition: A French scale for children and its consequences on a sociocognitive level]. *L'Année Psychologique, 100*(4), 585–628. https://doi.org/10.3406/psy.2000.28665

Gnambs, T., & Schroeders, U. (2020). Cognitive abilities explain wording effects in the Rosenberg Self-Esteem Scale. *Assessment, 27*(2), 404–418. https://doi.org/10.1177/1073191117746503

Gnambs, T., & Schroeders, U. (2024). Reliability and factorial validity of the Core Self-Evaluations Scale: A meta-analytic investigation of wording effects. *European Journal of Psychological Assessment*, Advance online publication. https://doi.org/10.1027/1015-5759/a000783

Gnambs, T., & Sengewald, M.-A. (2023). Meta-analytic structural equation modeling with fallible measurements. *Zeitschrift für Psychologie, 231*(1), 39–52. https://doi.org/10.1027/2151-2604/a000511

Gnambs, T., & Staufenbiel, T. (2016). Parameter accuracy in meta-analyses of factor structures. *Research Synthesis Methods, 7*(2), 168–186. https://doi.org/10.1002/jrsm.1190

*Gomes, A., Santos, J. D., Gonçalves, G., Orgambídez-Ramos, A., & Giger, J. (2013). Estudo de validação da Escala de Necessidade de Cognição com amostra portuguesa [Validation study of the Need for Cognition Scale with a Portuguese sample]. *Avaliação Psicológica, 12*(2), 179–192. https://doaj.org/article/24e07dd2537f4cf085d8adf614482701

*Gouveia, V. V., Mendes, L. A. D. C., Soares, A. K. S., Monteiro, R. P., Santos, L. C., & d. O., (2015). Escala de Necessidade de Cognição (NCS-18): Efeito de itens negativos em sua estrutura fatorial [Need for Cognition Scale (NCS-18): Effect of negative items in its factorial structure]. *Psicologia: Reflexão e Crítica, 28*(3), 425–433. https://doi.org/10.1590/1678-7153.201528301

*Grădinaru, D., Constantin, T., & Sorin, C. (2023). Psychometric properties of the Romanian version of the borderline personality questionnaire in a sample of nonclinical adults. *Psihologija*. Advance online publication. https://doi.org/10.2298/psi210624033g

Grass, J., Strobel, A., & Strobel, A. (2017). Cognitive investments in academic success: The role of need for cognition at university. *Frontiers in Psychology, 8*, Article 790. https://doi.org/10.3389/fpsyg.2017.00790

Gülgöz, S., & Sadowski, C. J. (1995). Turkish adaptation of the Need for Cognition Scale and its correlation with academic performance measures. *Türk Psikoloji Dergisi, 10*(35), 15–24.

*Gústavsson, M. F., Ólafsdóttir, R. Ó., & Holm, Þ. G. (2020). *Próffræðilegir eiginleikar Þankaþarfakvarðans í nýrri íslenskri Þýðingu* [Psychometric properties of the Need for Cognition Scale in a new Icelandic translation] [Bachelor's thesis]. Skemman, University of Akureyri. http://hdl.handle.net/1946/36222

*Hallahan, K. (2009). Need for cognition as motivation to process publicity and advertising. *Journal of Promotion Management, 14*(3–4), 169–194. https://doi.org/10.1080/10496490802353790

*Hanel, P. H. P., & Wolf, L. J. (2020). Leavers and Remainers after the Brexit referendum: More united than divided after all? *British Journal of Social Psychology, 59*(2), 470–493. https://doi.org/10.1111/bjso.12359

Hardwicke, T. E., Thibault, R. T., Kosie, J. E., Wallach, J. D., Kidwell, M. C., & Ioannidis, J. P. (2021). Estimating the prevalence of transparency and reproducibility-related research practices in psychology (2014–2017). *Perspectives on Psychological Science, 17*(1), 239–251. https://doi.org/10.1177/1745691620979806

Heene, M., Hilbert, S., Draxler, C., Ziegler, M., & Bühner, M. (2011). Masking misfit in confirmatory factor analysis by increasing unique variances: A cautionary note on the usefulness of cutoff values of fit indices. *Psychological Methods, 16*(3), 319–336. https://doi.org/10.1037/a0024917

Hevey, D., Thomas, K., Pertl, M., Maher, L., Craig, A., & Chuinneagain, S. N. (2012). Method effects and the Need for Cognition Scale. *The International Journal of Educational and Psychological Assessment, 12*(1), 20–33.

Hildebrandt, A., Lüdtke, O., Robitzsch, A., Sommer, C., & Wilhelm, O. (2016). Exploring factor model parameters across continuous variables with local structural equation models. *Multivariate Behavioral Research, 51*(2–3), 257–258. https://doi.org/10.1080/00273171.2016.1142856

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118

*Hussey, I., & Hughes, S. (2020). Hidden invalidity among 15 commonly used measures in social and personality psychology. *Advances in Methods and Practices in Psychological Science, 3*(2), 166–184. https://doi.org/10.1177/2515245919882903

Jak, S. (2015). *Meta-analytic structural equation modelling.* Springer. https://doi.org/10.1007/978–3-319–27174-3

Jak, S., & Cheung, M. W. L. (2020). Meta-analytic structural equation modeling with moderating effects on SEM parameters. *Psychological Methods, 25*(4), 430–455. https://doi.org/10.1037/met0000245

Jak, S., & Cheung, M. W. L. (2023). *Can findings from meta-analytic structural equation modeling in management and organizational psychology be trusted?* PsyArxiv. https://doi.org/10.31234/osf.io/b3qvn

Jankowsky, K., Olaru, G., & Schroeders, U. (2020). Compiling measurement invariant short scales in cross-cultural personality assessment using Ant Colony Optimization. *European Journal of Personality, 34*(3), 470–485. https://doi.org/10.1002/per.2260

Janssen, A. B., Schultze, M., & Grötsch, A. (2017). Following the ants: Development of short scales for proactive personality and supervisor support by ant colony optimization. *European Journal of Psychological Assessment, 33*(6), 409–421. https://doi.org/10.1027/1015-5759/a000299

*Janssen, E., Verkoeijen, P. P. J. L., Heijltjes, A., Mainhard, T., Van Peppen, L. M., & Van Gog, T. (2020). Psychometric properties of the Actively Open-minded Thinking scale. *Thinking Skills and Creativity, 36*, Article 100659. https://doi.org/10.1016/j.tsc.2020.100659

Jebb, A. T., Saef, R., Parrigon, S., & Woo, S. E. (2016). The need for cognition: Key concepts, assessment, and role in educational outcomes. In A. A. Lipnevich, F. Preckel, & R. D. Roberts (Eds.), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice* (pp. 115–132). Springer. https://doi.org/10.1007/978-3-319-28606-8_5

*Jin, C. H. (2016). The effects of mental simulations, innovativeness on intention to adopt brand application. *Computers in Human Behavior, 54*, 682–690. https://doi.org/10.1016/j.chb.2015.08.013

Jin, K.-Y., Chen, H.-F., & Wang, W.-C. (2017). Mixture item response models for inattentive responding behavior. *Organizational Research Methods, 21*(1), 197–225. https://doi.org/10.1177/1094428117725792

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A brief measure of dark personality traits. *Assessment, 21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Kam, C. C. S., & Zhou, M. (2015). Does acquiescence affect individual items consistently? *Educational and Psychological Measurement, 75*(5), 764–784. https://doi.org/10.1177/0013164414560817

Kam, C. C. S., & Fan, X. (2020). Investigating response heterogeneity in the context of positively and negatively worded items by using factor mixture modeling. *Organizational Research Methods, 23*(2), 322–341. https://doi.org/10.1177/1094428118790371

Kao, C. (1994). The concept and measurement of need for cognition: The concept and measurement of need for cognition. *Chinese Journal of Psychology, 36*, 1–20.

*Karagiannopoulou, E., Milienos, F. S., & Rentzios, C. (2020). Grouping learning approaches and emotional factors to predict students' academic progress. *International Journal of School and Educational Psychology, 10*(2), 258–275. https://doi.org/10.1080/21683603.2020.1832941

Kardash, C. M., & Noel, L. K. (2000). How organizational signals, need for cognition, and verbal ability affect text recall and recognition. *Contemporary Educational Psychology, 25*(3), 317–331. https://doi.org/10.1006/ceps.1999.1011

Keller, U., Strobel, A., Wollschläger, R., Greiff, S., Martin, R., Vainikainen, M., & Preckel, F. (2019). A need for cognition scale for children and adolescents. *European Journal of Psychological Assessment, 35*(1), 137–149. https://doi.org/10.1027/1015-5759/a000370

Kennedy, R., Clifford, S., Burleigh, T., Waggoner, P. D., Jewell, R., & Winter, N. J. G. (2020). The shape of and solutions to the MTurk quality crisis. *Political Science Research and Methods, 8*(4), 614–629. https://doi.org/10.1017/psrm.2020.6

Kenny, D. A., & Kashy, D. A. (1992). Analysis of the multitrait-multimethod matrix by confirmatory factor analysis. *Psychological Bulletin, 112*(1), 165–172. https://doi.org/10.1037/0033-2909.112.1.165

*Koutsogiorgi, C. C. (2020). *Responding to positively and negatively worded items: Correlational and experimental evidence in conceptually distinct areas* [Unpublished doctoral dissertation]. University of Cyprus.

Krippendorff, K. (2013). *Content analysis: An introduction to its methodology.* Sage.

Kruyen, P. M., Emons, W. H. M., & Sijtsma, K. (2013). On the shortcomings of shortened tests: A literature review. *International Journal of Testing, 13*(3), 223–248. https://doi.org/10.1080/15305058.2012.703734

Lance, C. E., Noble, C. L., & Scullen, S. E. (2002). A critique of the correlated trait-correlated method and correlated uniqueness models for multitrait-multimethod data. *Psychological Methods, 7*(2), 228–244. https://doi.org/10.1037/1082-989x.7.2.228

*Lantos, D., & Harris, L. T. (2021). The humanity inventory: Developing and validating an individual difference measure of dehumanization propensity. *Journal of Theoretical Social Psychology, 5*(4), 502–518. https://doi.org/10.1002/jts5.114

*Laroche, M., Tomiuk, M. A., Toffoli, R., & Richard, M. (2009). Analyses traditionnelles et FDI des échelles de mesure: application à l'échelle de l'intensité du raisonnement cognitif [Traditional and DIF analyses of measurement scales: Application to the Need for Cognition Scale]. *Canadian Journal of Administrative Sciences, 21*(4), 344–360. https://doi.org/10.1111/j.1936-4490.2004.tb00350.x

*Lee, K. Y., Reis, H. T., & Rogge, R. D. (2020). Seeing the world in pink and blue: Developing and exploring a new measure of essentialistic thinking about gender. *Sex Roles, 83*(11–12), 685–705. https://doi.org/10.1007/s11199-020-01141-1

Leite, W. L., Huang, I.-C., & Marcoulides, G. A. (2008). Item selection for the development of short forms of scales using an Ant Colony Optimization algorithm. *Multivariate Behavioral Research, 43*(3), 411–431. https://doi.org/10.1080/00273170802285743

Levin, I. P., Huneke, M. E., & Jasper, J. D. (2000). Information processing at successive stages of decision making: Need for cognition and inclusion–exclusion effects. *Organizational Behavior and Human Decision Processes, 82*(2), 171–193. https://doi.org/10.1006/obhd.2000.2881

*Lins de Holanda Coelho, G., Hanel, P. H. P., & Wolf, L. J. (2020). The very efficient assessment of need for cognition: Developing a six-item version. *Assessment, 27*(8), 1870–1885. https://doi.org/10.1177/1073191118793208

*Loose, T., Vásquez-Echeverría, A., & Alvarez-Nuñez, L. (2023). Spanish version of Need for Cognition Scale: Evidence of reliability, validity and factorial invariance of the very efficient short-form. *Current Psychology, 42*(17), 14440–14451. https://doi.org/10.1007/s12144-022-02739-2

Lord, K. A., & Putrevu, S. (2006). Exploring the dimensionality of the Need for Cognition Scale. *Psychology & Marketing, 23*(1), 11–34. https://doi.org/10.1002/mar.20108

Lorenzo-Seva, U., Timmerman, M. E., & Kiers, H. A. (2011). The Hull method for selecting the number of common factors. *Multivariate Behavioral Research, 46*(2), 340–364. https://doi.org/10.1080/00273171.2011.564527

*Ludwig, R. M., Srivastava, S. K., & Berkman, E. T. (2018). Planfulness: A process-focused construct of individual differences in goal achievement. *Collabra, 4*(1), Article 28. https://doi.org/10.1525/collabra.136

Luong, C., Strobel, A., Wollschläger, R., Greiff, S., Vainikainen, M., & Preckel, F. (2017). Need for cognition in children and adolescents: Behavioral correlates and relations to academic achievement and potential. *Learning and Individual Differences, 53*, 103–113. https://doi.org/10.1016/j.lindif.2016.10.019

*Luong, R., & Lomanowska, A. M. (2022). Evaluating Reddit as a crowdsourcing platform for psychology research projects. *Teaching of Psychology, 49*(4), 329–337. https://doi.org/10.1177/00986283211020739

MacCallum, R. C., Widaman, K. F., Preacher, K. J., & Hong, S. (2001). Sample size in factor analysis: The role of model error. *Multivariate Behavioral Research, 36*(4), 611–637. https://doi.org/10.1207/S15327906MBR3604_06

*Maldonado, J. C., García, M. L. S., Sintas, F., & Amat, M. E. (1993). Evaluación de la tendencia al esfuerzo cognitivo [Evaluation of the tendency to cognitive effort]. *Anuario De Psicología, 58*, 53–68. http://diposit.ub.edu/dspace/bitstream/2445/98931/1/103240.pdf

*Malmberg, J. (2010). *Pleasure & duty: Are there differences to store choice criteria between hedonic and functional stores? [Master's thesis]*. University of Rotterdam. https://hdl.handle.net/2105/8563

Marsh, H. W. (1989). Confirmatory factor analyses of multitrait-multimethod data: Many problems and a few solutions. *Applied Psychological Measurement, 13*(4), 335–361. https://doi.org/10.1177/014662168901300402

Marsh, H. W., & Bailey, M. (1991). Confirmatory factor analyses of multitrait-multimethod data: A comparison of alternative models. *Applied Psychological Measurement, 15*(1), 47–70. https://doi.org/10.1177/014662169101500106

*Menendez, D., Brown, S. A., & Alibali, M. W. (2023). Some correct strategies are better than others: Individual differences in strategy evaluations are related to strategy adoption. *Cognitive Science, 47*(3), Article e13269. https://doi.org/10.1111/cogs.13269

*Minson, J. A., Chen, F. S., & Tinsley, C. H. (2020). Why won't you listen to me? Measuring receptiveness to opposing views. *Management Science, 66*(7), 3069–3094. https://doi.org/10.1287/mnsc.2019.3362

Mussel, P. (2013). Intellect: A theoretical framework for personality traits related to intellectual achievements. *Journal of Personality and Social Psychology, 104*(5), 885–906. https://doi.org/10.1037/a0031918

Nair, K. U., & Ramnarayan, S. (2000). Individual differences in need for cognition and complex problem solving. *Journal of Research in Personality, 34*(3), 305–328. https://doi.org/10.1006/jrpe.1999.2274

*Newman, E. J., Jalbert, M., Schwarz, N., & Ly, D. P. (2020). Truthiness, the illusory truth effect, and the role of need for cognition. *Consciousness and Cognition, 78*, Article 102866. https://doi.org/10.1016/j.concog.2019.102866

Nosek, B. A., Hardwicke, T. E., Moshontz, H., Allard, A., Corker, K. S., Dreber, A., Fidler, F., Hilgard, J., Struhl, M. K., Nuijten, M. B., Rohrer, J. M., Romero, F., Scheel, A. M., Scherer, L. D., Schönbrodt, F. D., & Vazire, S. (2022). Replicability, robustness, and reproducibility in psychological science. *Annual Review of Psychology, 73*(1), 719–748. https://doi.org/10.1146/annurev-psych-020821-114157

Nudelman, G., & Otto, K. (2020). The development of a new generic risk-of-bias measure for systematic reviews of surveys. *Methodology, 16*(4), 278–298. https://doi.org/10.5964/meth.4329

Olaru, G., Schroeders, U., Hartung, J., & Wilhelm, O. (2019). Ant Colony Optimization and local weighted structural equation modeling. A tutorial on novel item and person sampling procedures for personality research. *European Journal of Personality, 33*(3), 400–419. https://doi.org/10.1002/per.2195

Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale Big-Five assessments. *Journal of Research in Personality, 59*, 56–68. https://doi.org/10.1016/j.jrp.2015.09.001

*Park, J. S. (2012). *Effects of online consumer reviews on attitudes and behavioral intentions toward products and retailers [Doctoral dissertation]*. University of Tennessee. http://trace.tennessee.edu/utk_graddiss/1552/

Perri, M., & Wolfgang, A. P. (1988). A modified measure of need for cognition. *Psychological Reports, 62*(3), 955–957. https://doi.org/10.2466/pr0.1988.62.3.955

*Petrović, M. B., & Žeželj, I. (2022). Thinking inconsistently: Development and validation of an instrument for assessing proneness to doublethink. *European Journal of Psychological Assessment, 38*(6), 463–475. https://doi.org/10.1027/1015-5759/a000645

Petty, R. E., Briñol, P., Loersch, C., & McCaslin, M. J. (2009). The need for cognition. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 318–329). Guilford Press.

Petty, R. E., DeMarree, K. G., Briñol, P., Horcajo, J., & Strathman, A. (2008). Need for cognition can magnify or attenuate priming effects in social judgment. *Personality and Social Psychology Bulletin, 34*(7), 900–912. https://doi.org/10.1177/0146167208316692

Petty, R. E., Schumann, D. W., Richman, S. A., & Strathman, A. J. (1993). Positive mood and persuasion: Different roles for affect under high- and low-elaboration conditions. *Journal of Personality and Social Psychology, 64*(1), 5–20. https://doi.org/10.1037/0022-3514.64.1.5

Pieters, R. G. M., Verplanken, B., & Modde, J. M. (1987). "Neiging tot nadenken": Samenhang met beredeneerd gedrag ["Need for cognition": Relationship with reasoned action]. *Nederlands Tijdschrift voor de Psycholoy, 42*, 62–70.

*Pilli, L. E., & Mazzon, J. A. (2016). Information overload, choice deferral, and moderating role of need for cognition: Empirical evidence. *Revista De Administração, 51*(1), 036–055. https://doi.org/10.5700/rausp1222

*Powell, C. F., Nettelbeck, T., & Burns, N. R. (2016). Deconstructing intellectual curiosity. *Personality and Individual Differences, 95*, 147–151. https://doi.org/10.1016/j.paid.2016.02.037

Preckel, F. (2014). Assessing need for cognition in early adolescence. *European Journal of Psychological Assessment, 30*(1), 65–72. https://doi.org/10.1027/1015-5759/a000170

*Pryor, P. L., McGahan, J. R., McDougal, B., Haire, S. M., & Marashi, H. (2000). Association of need for cognition with judgments of height, weight, and body fat covariation. *Psychological Reports, 87*(3 suppl), 1147–1157. https://doi.org/10.2466/pr0.2000.87.3 f.1147

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research, 47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment, 92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Ruscio, J., & Roche, B. (2012). Determining the number of factors to retain in an exploratory factor analysis using comparison data of known factorial structure. *Psychological Assessment, 24*(2), 282–292. https://doi.org/10.1037/a0025697

Sadowski, C. J. (1993). An examination of the short Need for Cognition Scale. *The Journal of Psychology: Interdisciplinary and Applied, 127*(4), 451–454. https://doi.org/10.1080/00223980.1993.9915581

*Salama-Younes, M., Guingouain, G., Le Floch, V., & Somat, A. (2014). Besoin de cognition, besoin d'évaluer, besoin de clôture: Proposition d'échelles en langue Française et approche socio-normative des besoins dits fondamentaux [Need for cognition, need for closing, need to evaluate: Proposal of scales in French and socio-normative approach of fundamental needs]. *European Review of Applied Psychology/Revue Européenne de Psychologie Appliquée, 64*(2), 63–75. https://doi.org/10.1016/j.erap.2014.01.001

Schermelleh-Engel, K., Moosbrugger, H., & Müller, H. (2003). Evaluating the fit of structural equation models: Tests of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research, 8*(2), 23–74. https://doi.org/10.23668/psycharchives.12784.

Scherpenzeel, A. C., & Das, M. (2010). "True" longitudinal and probability-based internet panels: Evidence from the Netherlands. In M. Das, P. Ester, & L. Kaczmirek (Eds.), *Social and Behavioral Research and the Internet: Advances in Applied Methods and Research Strategies* (pp. 77–104) Taylor & Francis.

Schoeni, R. F., Stafford, F. P., McGonagle, K. A., & Andreski, P. (2012). Response rates in national panel surveys. *Annals of the American Academy of Political and Social Science, 645*(1), 60–87. https://doi.org/10.1177/0002716212456363

Schroeders, U., Morgenstern, M., Jankowsky, K., & Gnambs, T. (2024). *Short-scale construction using meta-analytic Ant Colony Optimization: A demonstration with the Need for Cognition Scale* [Data, Materials]. https://osf.io/tbrdv

Schroeders, U., Scharf, F., & Olaru, G. (2023). Model specification searches in structural equation modeling using Bee Swarm Optimization. *Educational and Psychological Measurement, 84*(1), 40–61. https://doi.org/10.1177/00131644231160552

Schroeders, U., Wilhelm, O., & Olaru, G. (2016a). Meta-heuristics in short scale construction: Ant Colony Optimization and Genetic Algorithm. *PLoS One, 11*(11), Article e0167110. https://doi.org/10.1371/journal.pone.0167110

Schroeders, U., Wilhelm, O., & Olaru, G. (2016b). The influence of item sampling on sex differences in knowledge tests. *Intelligence, 58*(3), 22–32. https://doi.org/10.1016/j.intell.2016.06.003

*Shchebetenko, S. A. (2011). Psihometrika russkoj versii Škaly potrebnosti v poznanii. Vestnik Permskogo universiteta [Russian version of 18-item Need for Cognition Scale]. *Filosofija. Psihologija. Sociologija, 2*(6), 87–100.

*Sousa, C., Palácios, H., Gonçalves, C., Santana Fernandes, J., & Gonçalves, G. (2018). Need for cognition in a Portuguese managers sample: Invariance across gender and professional activity. *The Psychologist-Manager Journal, 21*(4), 249–271. https://doi.org/10.1037/mgr0000077

Stanovich, K. E., & West, R. F. (2007). Natural myside bias is independent of cognitive ability. *Thinking & Reasoning, 13*(3), 225–247. https://doi.org/10.1080/13546780600780796

Steger, D., Jankowsky, K., Schroeders, U., & Wilhelm, O. (2023). The road to hell is paved with good intentions: How common practices in scale construction hurt validity. *Assessment, 30*(6), 1811–1824. https://doi.org/10.1177/10731911221124846

Tanaka, J. S., Panter, A. T., & Winborne, W. C. (1988). Dimensions of the need for cognition: Subscales and gender differences. *Multivariate Behavioral Research, 23*(1), 35–50. https://doi.org/10.1207/s15327906mbr2301_2

Tanaś, Ł. (2021). Curiosity in children and adolescents: Data from the Polish adaptation of the Need for Cognition Scale. *Psychological Test Adaptation and Development, 2*(1), 24–34. https://doi.org/10.1027/2698-1866/a000007

*Tobin, S. J., & Guadagno, R. E. (2022). Why people listen: Motivations and outcomes of podcast listening. *PLoS One, 17*(4), Article e0265806. https://doi.org/10.1371/journal.pone.0265806

*Türker, A., İşçi, C., & Özaltın Türker, G. (2015). Biliş ihtiyacının satış performansı üzerine etkisi: acente temsilcileri üzerine bir uygulama [The effect of need for cognition on sales performance: An application on agency representatives]. *Akademik Bakış Dergisi, 47*, 108–125.

*Van Tilburg, W. A. P., Igou, E. R., Maher, P. J., Moynihan, A. B., & Martin, D. G. (2019). Bored like hell: Religiosity reduces boredom and tempers the quest for meaning. *Emotion, 19*(2), 255–269. https://doi.org/10.1037/emo0000439

*Vaughan-Johnston, T. I., & Jacobson, J. A. (2020). "Need" personality constructs and preferences for different types of self-relevant feedback. *Personality and Individual Differences, 154*, Article 109671. https://doi.org/10.1016/j.paid.2019.109671

Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika, 41*(3), 321–327. https://doi.org/10.1007/BF02293557

Waters, L. K., & Zakrajsek, T. D. (1990). Correlates of need for cognition total and subscale scores. *Educational and Psychological Measurement, 50*(1), 213–217. https://doi.org/10.1177/0013164490501026

*Weigold, A., & Weigold, I. K. (2022). Traditional and modern convenience samples: An investigation of college student, Mechanical Turk, and Mechanical Turk college student samples. *Social Science Computer Review, 40*(5), 1302–1322. https://doi.org/10.1177/08944393211006847

*Weng, J., & DeMarree, K. G. (2019). An examination of whether mindfulness can predict the relationship between objective and subjective attitudinal ambivalence. *Frontiers in Psychology, 10*, Article 854. https://doi.org/10.3389/fpsyg.2019.00854

West, R. G., 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*(4), 930–941. https://doi.org/10.1037/a0012842

*Yamamoto, S., & Maeder, E. M. (2019). Creating the punishment orientation questionnaire: An item response theory approach. *Personality and Social Psychology Bulletin, 45*(8), 1283–1294. https://doi.org/10.1177/0146167218818485

*Zhang, X., Noor, R., & Savalei, V. (2016). Examining the effect of reverse worded items on the factor structure of the Need for Cognition Scale. *PLoS One, 11*(6), Article e0157795. https://doi.org/10.1371/journal.pone.0157795

Conflict of Interest

We have no conflicts of interest to disclose. We confirm that the work conforms to Standard 8 of the American Psychological Association's Ethical Principles of Psychologists and Code of Conduct.

Open Science

Open Data: We confirm that there is sufficient information for an independent researcher to reproduce all of the reported results (Schroeders et al., 2024).
Open Materials: We confirm that there is sufficient information for an independent researcher to reproduce the reported methodology (Schroeders et al., 2024).
Preregistration of Studies and Analysis Plans: This study was not preregistered.

ORCID

Ulrich Schroeders
 https://orcid.org/0000-0002-5225-1122
Kristin Jankowsky
 https://orcid.org/0000-0002-4847-0760
Timo Gnambs
 https://orcid.org/0000-0002-6984-1276

Ulrich Schroeders
Institute of Psychology
University of Kassel
Holländische Str. 36-38
34127 Kassel
Germany
schroeders@psychologie.uni-kassel.de