frontiers in Education

ORIGINAL RESEARCH
published: 13 February 2017
doi: 10.3389/feduc.2017.00002

Assessing the Willingness to Elaborate among Young Students: Psychometric Evaluation of a Swedish Need for Cognition Scale

Tova Stenlund and Bert Jonsson*

Department of Psychology, Umeå University, Umeå, Sweden

The personality trait Need for Cognition (NFC) has been studied for many years and found to be important for individuals' educational achievement. The original NFC-scale was developed in the eighties, and during the following decade the scale was translated and adapted into a number of other languages. A renewed interest for the personality trait of NFC has made these scales interesting to use. It is though vital that instruments used for studies of individual differences in the area of educational research, or in any other area, can portray valid results today. The aim of the present paper was to evaluate validity and reliability of the short version of the *Mental Effort Tolerance Questionnaire*, a Swedish adaption of the NFC-scale made in 1991, which has not been previously evaluated. This study involved 420 young students, and the evaluation of reliability includes a study of temporal stability (test–retest), as well as internal stability. Further, the evaluation of validity includes construct and criterion validity. Regarding reliability, the results showed a test–retest reliability coefficient of 0.88 (n = 108) and an internal stability (Cronbach's alpha) of 0.88 (n = 420). Evaluation of construct validity found evidence for a five factor dimensional structure (n = 420), discriminant validity to measures of general intelligence (r = 0.25; n = 122), working memory (r = 0.22; n = 164), and the personality trait Grit (r = 0.26; n = 169). Finally, criterion validity was found for grades (r = 0.35; n = 125). Overall, the results of the evaluation show that the inferences made from the results of the short version of the Swedish NFC-scale exhibits satisfactory reliability and validity, suggesting that the questionnaire can be used in educational contexts. The questionnaire might, however, benefit from being even more shortened.

Keywords: personality, need for cognition, grit, reliability, validity, educational achievement

OPEN ACCESS

Edited by:
Meryem Yilmaz Soylu, Formerly affiliated with Meliksah University, Turkey

Reviewed by:
Monika Fleischhauer, Private University of Applied Sciences, Germany
Anja Strobel, Dresden University of Technology, Germany

****Correspondence:***
Tova Stenlund
tova.stenlund@umu.se

Specialty section:
This article was submitted to Educational Psychology, a section of the journal Frontiers in Education

Received: *16 November 2016*
Accepted: *19 January 2017*
Published: *13 February 2017*

Citation:
Stenlund T and Jonsson B (2017) Assessing the Willingness to Elaborate among Young Students: Psychometric Evaluation of a Swedish Need for Cognition Scale. Front. Educ. 2:2. doi: 10.3389/feduc.2017.00002

INTRODUCTION

Need for Cognition (NFC) is an individual difference in the "tendency to engage in and enjoy thinking" (Cacioppo and Petty, 1982, p. 116). NFC, as a concept, has been studied by psychologists for many years [see, e.g., Maslow (1943) and Cohen et al. (1955)]. The definition has though changed from the need to understand and make reasonable the experiential world (Cohen et al., 1955) to the need to engage in and enjoy cognitively demanding activities (Cacioppo and Petty, 1982). In recent years, there has been a revival of interest in NFC. It is expected that individuals who receive high NFC-scores evaluate and elaborate on information to a higher degree, and they

also find this process more enjoyable when compared to those with low scores on the NFC-scale (Cacioppo and Petty, 1982; Cacioppo et al., 1996). NFC is not considered to be a proxy for intelligence as it indexes motivation to engage but is most likely moderately associated with intellectual ability (Cacioppo et al., 1996; Hill et al., 2013; Preckel, 2014). NFC is also a personality characteristic that is suggested to be stable over time (however, possible to mature) and found to be important for, and positively related to individuals' performance in an educational context (Richardson et al., 2012; Meier et al., 2014; Stenlund et al., 2016). For example, Meier et al. (2014) showed that NFC best predicted attendance of special classes over and above cognitive ability, academic achievement, sex, and parental level of education. NFC was also found to be the third strongest non-intellectual (such as personality traits, motivational factors, and self-regulated learning strategies) predictor for tertiary GPA in a recent meta-analysis (Richardson et al., 2012). Further, Stenlund et al. (2016) showed that individuals with high NFC-scores benefit more in participating in structured groups discussions and did in relation to information processing also remembered more on a subsequent recall task. This result confirmed previous studies on NFC and information processing [see Cacioppo et al. (1996) for a review]. Verplanken et al. (1992) did in addition discover that individuals with high NFC-scores were more prone to search for new information. The significance of NFC as a moderator of information processing was also shown in a study of information search and interactive web pages (Sicilia et al., 2005).

Cacioppo and Petty (1982) developed a scale to measure NFC, which has been explored in relation to a variety of individual differences, such as intelligence, personality, and academic achievement [see, e.g., Cacioppo et al. (1996), Fleischhauer et al. (2010), and Preckel (2014)]. The original and short version of the NFC-scale is, according to Cacioppo et al. (1996), reasonable, reliable, and valid. Over the years the NFC-scale has been adopted and translated into several other languages [see, e.g., Dornic et al. (1991), Bless et al. (1994), and Curseu (2011)]. The Swedish adaption of the NFC, the *Mental Effort Tolerance Questionnaire* (METQ) was made in 1991 and has not been psychometrically evaluated since (Dornic et al., 1991). In the Dornic et al. (1991) study, a short version of the METQ also was suggested; however, it was neither further examined nor evaluated. Further, this study only involved 70 participants with a mean age of 30 years, and according to Preckel (2014) NFC has predominantly been studied in adult groups. Hence, there is a lack of empirical evidence for the Swedish NFC-scale, and more research is needed to establish whether the Swedish adaption of the NFC-scale can portray valid and reliable results in a younger population, and as such be useful in an educational context. The main purpose of this study was to examine the quality of the results from the short version of the METQ in a relatively large group of young students.

In the present study, the reliability, in terms of temporal and internal stability, of the results of the METQ was examined, and further, construct and criteria-related validity. Little attention has been paid to the NFC-scale's temporal stability, but studies found show high stability, i.e., a correlation coefficient exceeding 0.80 (Sadowski and Gulgoz, 1992; Bertrams and Dickhäuser, 2010; Fleischhauer et al., 2015). Based on these studies we expected high temporal stability (i.e., test–retest reliability) of the METQ. In contrast to temporal stability, the internal stability of NFC has been frequently examined, with most studies showing high internal stability that is above 0.80 [see, e.g., Venkatraman and Price (1990), Cacioppo et al. (1996), Fleischhauer et al. (2010), and Preckel (2014)]. A similar result is expected in the present study. With regards to construct validity, the results of the METQ were expected to be unidimensional, in line with the original NCF-scale (Dornic et al., 1991; Cacioppo et al., 1996). However, studies have shown that more than one factor might be involved [see, e.g., Cacioppo et al. (1984), Tanaka et al. (1988), and Fosterlee and Ho (1999)], suggesting that this is not a clear case. For example, Fosterlee and Ho (1999), who examined the short version of the NFC-scale, found two factors, while Tanaka et al. (1988) found three factors when examining the original 34-item NFC-scale. The latter result was also confirmed by Waters and Zakrajsek (1990). However, previous studies have often argued for a unidimensional structure based on the findings of a dominant factor and the proportion variance explained (Fosterlee and Ho, 1999). In the present study, relationships with other constructs were also examined as regarding construct validity. First, a measure of general fluid intelligence was used to examine conceptual autonomy. The relationship between the METQ and general fluid intelligence was expected to be relatively weak or moderate, but statistically significant, in line with earlier studies (typically the coefficients range between 0.20 and 0.30) and theories about NFC [see, e.g., Cacioppo et al. (1996), Fleischhauer et al. (2010), Hill et al. (2013), and Preckel (2014)]. Second, another measure of cognitive ability, a complex working memory task, was also used to examine the construct validity of the results of the METQ. Earlier studies (Hill et al., 2013, 2016) argue that working memory, because of its strong relation with intelligence, should be associated with NFC. However, the result from these studies have only shown a weak relation ($r = 0.12$–0.13), and we expect to find a similar result in this study. Third, in the present paper we examined the relationship between the METQ and Grit, which is another non-cognitive personality trait predicting success in an educational context. Grit is defined by Duckworth et al. (2007) as "perseverance and passion for long term-goals" (p. 1,087). A weak to moderate positive relation ($r = 0.30$) between NFC and personality traits, indicating goal orientation, such as Grit, has been found in earlier studies (Fleischhauer et al., 2010). We expected a similar relation, but, as for general fluid intelligence and working memory, weak enough to show a discrepancy between the two constructs. Finally, to examine criterion validity, the relation between the METQ and educational achievement was analyzed. Cacioppo et al. (1996) argue that the NFC is linked to effortful cognitive processing, and additional studies have also found that NFC-scores seem be related to different areas important to educational achievement, such as comprehension performance (Dai and Wang, 2007; Bråten et al., 2014), and math achievements (Preckel, 2014). Moreover, recent meta-analytic studies, primarily examining college students, have found positive

relations ($r = 0.19–0.28$) between grade point average and NFC (Richardson et al., 2012; von Stumm and Ackerman, 2013). Based on these studies we expected to find a positive relation between the METQ and educational achievement, as measured by grades.

METHOD

Participants and Procedure

The participants in this study included students ($n = 434$) in upper-secondary schools in the northern part of Sweden. The data from the participants were collected on three different occasions during the period 2012–2016. The participants' ages ranged from 16 to 20 years, and the mean age among the students was 17.2 years (SD = 0.9), with 240 females (55.2%) and 194 males. The sample was homogenous with respect to ethnicity, and that they all were enrolled in theoretical upper-secondary programs. Of the 434 participants, 14 (8 females and 6 males) did not answer all items in the short version of METQ, and they were therefore excluded from the analyses.

All participants in the three samples received information about the aim and procedures of the studies, they signed an informed consent form, and they all filled in the METQ questionnaire. The three samples participated in different parts of the psychometric evaluation of the results of the METQ. The first sample was part of a larger study about memory and learning in the autumn of 2012. This sample consisted of 134 participants (108 girls, 80.5%; mean age: 17.3 years; SD = 0.60). A part of this sample completed a task that measured general fluid intelligence ($n = 126$). Information about grades from 9-year compulsory school was obtained from 131 of the participants in this sample (the grades were obtained from the Upper Secondary School Admission Office during the larger study about memory and learning). Further, in this sample, test–retest reliability of the METQ was examined, and 108 students completed the short version of the METQ twice (4 weeks apart). The second sample ($n = 119$, 64 girls, 54%; mean age: 17.7 years; SD = 1.00) only completed the METQ during the period 2014–2015. The third sample consisted of 181 participants (68 girls, 38%; mean age: 16.7 years; SD = 0.75) and was a part of a larger study about individual differences in memory and learning. A part of this sample filled in a questionnaire (GRIT-S; $n = 170$), measuring a different personality trait related to educational achievements, and completed a complex working memory task ($n = 164$).

Instruments

Mental Effort Tolerance Questionnaire

The original 40-item version of METQ, a Swedish adaptation of Cacioppo and Petty's original scale—i.e., the *NFC Scale* (Cacioppo and Petty, 1982)—was developed by Dornic et al. (1991) and also considered to be reliable and valid. The short version of the METQ, suggested in Dornic et al. (1991), consists of 30 items that are rated on a 5-point Likert-like scale (1 = strongly disagree, 3 = neutral, and 5 = strongly agree). Twelve of the statements indicated positive attitudes toward engaging and enjoying thinking, and 18 indicated negative attitudes. Thus, items that indicated negative attitudes required reverse scoring in order to conclude that high scores indicate a high NFC.

GRIT-S

The short version of GRIT (GRIT-S; Duckworth and Quinn, 2009) was used in the present study. The GRIT-S was translated (and back-translated to ensure quality) from English to Swedish. GRIT-S contains eight items, half of the items reflect an individual's ability to maintain interest (for example, "I often set a goal but later choose to pursue a different one") and half of the items reflects an individual's tendency to maintain effort (for example, "I have achieved a goal that took years of work"). The items are rated on a 5-point Likert-like scale (1 = strongly disagree 3 = neutral, and 5 = strongly agree). The GRIT-S has shown to have good reliability and validity (Duckworth and Quinn, 2009).

Raven Advanced Progressive Matrices

To measure general fluid intelligence, and to further evaluate the construct validity of the results of the METQ, the Raven advanced progressive matrices was used, which is a well-known and standardized instrument (Raven, 1990). The Raven consisted of 36 items, which were divided into 18 odd-numbered and 18 even-numbered items. One-half of the group completed the odd-numbered items, and the second half of the group completed the even-numbered items. All items were presented with an escalating degree of difficulty. The items consist of 3×3 matrices of geometric patterns in which the bottom-right pattern is missing. Eight alternatives are visible, and the participants are asked to select the alternative that is missing. The total number of correct solutions was tallied and represented the participant's score.

Operation Span Task

To measure working memory capacity (WMC) a standardized complex working memory task was used (Unsworth et al., 2005) denoted as operation span. In the Unsworth et al. (2005) operation span task, participants are required to perform mathematical operations (processing demand) while retaining letters in long-term memory (storage demand). The letters to recall consisted of three sets of each set size, with set size ranging from three to seven letters. The participants had to recall these letters in the same order as they were presented. The sum of all perfectly recalled sets was used as the dependent variable.

Grades

In order to relate the METQ scores to educational achievement (i.e., to evaluate criteria-related validity), the combined grades across 16 subjects from the participants' final (ninth) year in compulsory school were collected. In Sweden, a four-level point scale (0 = failed; 10 = pass; 15 = pass with distinction; and 20 = pass with special distinction) was used at this time, and the combined grades could range from 0 to 320 points.

Data Analysis

The quality of the results of the short version of the METQ was evaluated using the criteria to determine good reliability and

validity in psychological and educational tests developed by the European Federation of Psychologists' Associations (2013). According to EFPA, a sample size exceeding 100 is rated as adequate, and a sample size exceeding 200 is regarded as good, when examining reliability and validity. The EFPA rates the reliability and validity of studies across a 5-point scale with scores ranging from 0 to 4 (0 = not possible to rate, 1 = inadequate, 2 = adequate, 3 = good, and 4 = excellent).

The test–retest reliability was evaluated using intra-class correlation in the present study; a correlation coefficient at or above 0.70 was rated as good and above 0.80 as excellent. Internal consistency was estimated using Cronbach's alpha; a correlation coefficient at or above 0.80 was rated as good and above 0.90 as excellent. Criteria-related validity was evaluated by calculating Pearson's correlation; a correlation coefficient at or above 0.35 was rated as good and a coefficient above 0.50 as excellent. Construct validity was evaluated by examining relationships with other constructs, exploratory factor analysis, and corrected item–test correlations. Relationships with other constructs were evaluated by calculating Pearson's correlation coefficient. For a correlation to be rated as adequate convergent validity, a coefficient at or above 0.55 was needed. Concerning discriminant validity (i.e., to be able to distinct one construct from another), EFPA does not specify a particular coefficient that must be used to rate the discriminant validity as adequate; overall, the weaker the association the better.

To examine the dimensionality of the METQ and whether the results support the structure of the test (i.e., unidimensionality), the 30 items of the short version of the METQ was subject to maximum likelihood (ML) factor analyses (Costello and Osborne, 2005). First, an eigenvalue Monte Carlo simulation (i.e., parallel analysis) using the O'Connor (2000) SPSS syntax was performed to determine the number of factors that should be retained. Initially, the Kaiser–Meyer–Olkin measure of sampling adequacy (KMO) and Bartlett's test of sphericity were used to examine whether the data were suitable for factor analysis. For the KMO, a value of 0.6 or above was considered to be the minimum, and a significant result (i.e., a p-value of 0.05 or smaller) on Bartlett's test was needed for the factor analysis to be considered appropriate (Field, 2013). The initial data analyses showed that the data were suitable, as the KMO value was 0.88, and Bartlett's test of sphericity displayed a statistically significant result ($p < 0.001$).

RESULTS

Test–Retest Reliability

Test–retest reliability was estimated through a sample of 108 participants, who were asked to re-answer the METQ after a 4-week interval. The results showed excellent test–retest reliability, with an intra-class correlation coefficient of 0.88, $p < 0.001$, an interval of 0.83–0.92, and with 95% confidence.

Item Homogeneity and Item–Total Correlation

Internal consistency was estimated for the entire group ($N = 420$), and for all items in the short version of the METQ, and it was considered to be good (coefficient alpha = 0.88). Further, all items—except one—had an item–total score correlation above 0.30. The item–total score correlations ranged from 0.21 to 0.61, suggesting that the items correlate adequately well, not too high and not too low, with the total test score (European Federation of Psychologists' Associations, 2013).

Dimensionality—Factor Analyses

The parallel analysis extracted five factors ($p < 0.05$). The ML factor analysis showed one clearly dominant factor with an eigenvalue of 7.34, explaining 24.5% of the variance. The second factor had an eigenvalue of 2.05, the third 1.70, the fourth 1.51, and the fifth 1.33. Together, these five factors explained 46.5% of the total variance (see **Table 1** for a presentation of what characterizes the different factors). The rotated solution (oblique) revealed a structure where some of the items loaded on more than one factor, and some items did not load above 0.30 at any factor (**Table 1**). These latter items were items such as whether you would describe yourself as an intellectual or theoretician, whether you prefer assignments

TABLE 1 | Factor loadings for maximum likelihood factor analysis with promax rotation of the Mental Effort Tolerance Questionnaire scale.

Items	Factor 1 (to at all think on problems or other things)	Factor 2 (to learn new things and think of new ways to solve problems)	Factor 3 (to think fast and calculate in my head)	Factor 4 (to think ahead and to ponder about the future)	Factor 5 (to make better decisions you need to think things through)
Item 21	**0.77**	−0.21	−0.14	0.12	0.01
Item 1	**0.66**	0.07	−0.25	−0.03	−0.14
Item 27	**0.60**	0.03	−0.04	0.13	−0.03
Item 7	**0.54**	−0.08	0.08	−0.06	0.02
Item 28[a]	**0.46**	−0.06	−0.14	0.01	**0.45**
Item 3	**0.45**	−0.18	0.06	0.03	0.21
Item 25	**0.44**	0.22	0.06	−0.01	0.01
Item 23	**0.41**	0.04	0.09	0.17	−0.03
Item 16	**0.39**	0.29	0.15	−0.13	0.05
Item 8	**0.31**	0.12	0.05	−0.03	−0.03
Item 5	−0.12	**0.66**	−0.11	0.20	−0.03
Item 24	−0.26	**0.66**	−0.03	0.17	0.06
Item 18	−0.05	**0.56**	−0.03	0.28	−0.08
Item 15	0.07	**0.55**	0.17	−0.15	0.08
Item 11	0.01	**0.53**	−0.13	−0.06	0.10
Item 13	0.19	**0.44**	−0.09	0.14	−0.01
Item 9	0.15	**0.42**	0.07	−0.16	0.06
Item 2	0.22	**0.37**	0.05	−0.03	0.11
Item 10[a]	0.29	0.29	0.24	−0.11	−0.05
Item 17[a]	0.23	0.24	−0.11	0.11	−0.16
Item 4	−0.01	−0.16	**0.83**	0.09	−0.04
Item 29	−0.20	−0.01	**0.66**	−0.06	0.01
Item 30	−0.03	−0.09	**0.60**	0.15	−0.03
Item 6[a]	0.15	0.12	0.27	0.14	−0.02
Item 14[a]	0.21	0.15	0.24	0.00	−0.06
Item 26	−0.04	0.13	−0.02	**0.65**	0.09
Item 19	0.08	−0.07	0.09	**0.62**	0.12
Item 22[a]	**0.31**	0.02	0.11	**0.36**	0.00
Item 12	0.10	0.08	0.15	**0.32**	−0.05
Item 20	−0.14	0.17	−0.01	0.13	**0.82**

[a]*Items that did not load on any factor or cross-loaded.*
Factor loadings >0.30 are in boldface.

that do not require much thought, and whether you prefer to work in the same way as you always have.

The dominant first factor made it interesting to further examine the expected unidimensionality of the METQ. An exploratory factor analysis (ML) using latent variable software, Mplus (Muthén and Muthén, 2015), with the purpose of examining the model fit indices of the 1-factor model, and also to compare models, was therefore performed. With Mplus it was possible to compare models containing one to five factors, and to identify the model with the best fit. Multiple fit indices were calculated to provide information for model fit and model comparisons. The model fit indices used were Chi-square (χ^2); expected to be low and non-significant ($p > 0.05$), CFI; the minimum of acceptable fit is a value of 0.90, RMSEA; a value of 0.05–0.08 indicates reasonable error of approximation, below 0.05 good fit, and SRMR; a value below 0.08 is considered acceptable [see, e.g., Hu and Bentler (1999)]. **Table 2** shows that the models are significantly different ($p_s < 0.001$), and that the unidimensional model had poor model fit. However, the analysis supported the results from the parallel analysis and showed that a five factor model seems to be the most appropriate with regards to the model fit indices used. Note that none of the models compared had a non-significant Chi-square result.

Relationships with Other Measurements of Individual Differences

Raven was used to examine construct validity and whether the validity of the results of the short version of the METQ might be adequate. The correlation analysis showed that the relationship between the results of the METQ and general fluid intelligence, as measured by Raven, was statistically significant, albeit this relationship was rather weak, $r = 0.25$ ($n = 122$), $p < 0.01$. This result reflects discriminant validity rather than convergent validity according to the EFPA guidelines, supporting a conceptual autonomy.

O-span was used to examine whether there is a relation between the results of METQ and WMC. The correlation analyses showed a statistically significant but relatively weak relationship between the results of the METQ and WMC, as measured by O-span, $r = 0.22$ ($n = 164$), $p < 0.01$, thus reflecting discriminant validity.

GRIT-S was also used to examine the construct validity of the results of the short version of the METQ. The correlation analysis showed a statistically significant relation between the result of METQ and GRIT-S scores, $r = 0.26$ ($n = 169$), $p < 0.01$, however not strong enough to suggest convergent validity (European Federation of Psychologists' Associations, 2013).

Criterion-Related Validity

Grades were used to examine (concurrent) criterion-related validity of the results from the short version of the METQ. The results from the correlation analysis showed that the relationship between the results of the METQ and students' grades was statistically significant and very close to 0.35 [$r = 0.34$ ($n = 125$), $p < 0.01$], which is rated as good when examining criterion validity (European Federation of Psychologists' Associations, 2013).

DISCUSSION AND CONCLUSION

The focus of the present study was to examine the reliability and validity of the results of the short version of the METQ, a Swedish adaptation of the *NFC Scale*, in a group of young students. Overall, the evaluation showed that the results of the METQ in this group have excellent reliability, good criterion-related validity, and reasonable construct validity, when related to the EFPA's criteria.

The results of the short version of the METQ seem to be stable over time and exhibit internal stability, which is in accordance with previous studies. For example, Fleischhauer et al. (2015) found a test–retest correlation of 0.83, and studies examining internal stability with Cronbach's alpha have found that the NFC-scale typically exceeds 0.80 [see, e.g., Venkatraman and Price (1990) and Fleischhauer et al. (2010)]. Further, the result of the METQ had a significant, but rather weak, relationship with general fluid intelligence ($r = 0.25$) and working memory ($r = 0.22$). The weak relationship between the METQ and fluid intelligence is in line with earlier studies and theories about NFC [see, e.g., Fleischhauer et al. (2010) and Hill et al. (2013)]. Fleischhauer et al. (2010) reported a correlation of 0.25 between NFC and general fluid intelligence, and Hill et al. (2013) reported an even higher correlation of 0.38.

TABLE 2 | Model comparison and model fit indices for one to five factor models.

Models compared	$\Delta\chi^2$	df	*p*-Value	
1-factor against 2-factor	306.59	29	0.000	
2-factor against 3-factor	231.77	28	0.000	
3-factor against 4-factor	193.20	27	0.000	
4-factor against 5-factor	133.98	26	0.000	
Model fit	**χ^2 (df, *p*-Value)**	**CFI**	**RMSEA (CI)**	**SRMR**
1-factor	1,491.00 (405, $p < 0.001$)	0.68	0.079 (0.074–0.083)	0.071
2-factor	1,184.41 (376, $p < 0.001$)	0.76	0.070 (0.066–0.075)	0.059
3-factor	952.64 (348, $p < 0.001$)	0.82	0.063 (0.058–0.068)	0.049
4-factor	759.44 (321, $p < 0.001$)	0.87	0.056 (0.051–0.061)	0.042
5-factor	625.46 (295, $p < 0.001$)	0.90	0.051 (0.045–0.056)	0.036

CI is 90% confidence interval.

Regarding working memory, the relation found is somewhat stronger than in earlier studies (Hill et al., 2013, 2016). Hill et al. (2016) reported a correlation coefficient of 0.13 between NFC and working memory (O-span). The relationship found between METQ scores and the results of GRIT-S ($r = 0.26$) also supports construct validity. Thus, these two measures of non-cognitive personality traits, both found to be important for educational success, are somewhat related, but they do not measure the same construct. Fleischhauer et al. (2010) reported $r = 0.30$ between NFC and the Tridimensional Personality Questionnaire dimension Persistence, which is in line with the result in the present study. With respect to grades, the association with the results of the METQ in this study was significant and stronger compared to general fluid intelligence ($r = 0.35$). Earlier studies have examined the relationship with grades and found that NFC explains, for example, math grades over and above intelligence (Preckel, 2014).

These results strengthen the evidence supporting the idea that the short version of the METQ features reliable and valid results. However, the examination of the dimensionality of the results of the METQ does not support the theory of unidimensionality especially well. Even though a dominant factor accounting for about 25% of the total scale variance was found, which is in line with earlier studies of the NFC-scale [see, e.g., Cacioppo et al. (1984), Verplanken et al. (1992), Sadowski and Gulgoz (1992), and Bless et al. (1994)], neither the parallel analysis nor the model comparison supported a unidimensional solution. The parallel analysis extracted five factors (see **Table 2** for a description of the different factors), suggesting multidimensionality rather than unidimensionality in the results of the Swedish measure of NFC. Possibly indicating that some caution might be needed when interpreting the results of the METQ, as the METQ might measure something more than the original NFC-scale. Especially, the third factor (characterized by how much you like to calculate in your head and think fast) seems to be a dimension not found in the short version of the original NFC-scale (Cacioppo et al., 1996). This might also be true for the fifth factor only including one item about whether you make better decision when you think things through. When examining the rotated factor solution (**Table 1**) it also becomes clear that the short version of the METQ might benefit from being even more shortened. For example, some of the items loaded above 0.3 on several factors, only one item loaded above 0.3 at the fifth factor, and some items did not load above 0.3 at any factor. If dropping these seven items, and possibly the three items in the third factor, the dimensional structure (and construct validity) of the METQ might be improved. It would also make the scale more economical, and still capture the essence of NFC.

Still, these findings should be interpreted in the perspective of methodical limitations. First, although the sample size in the present study was above the limit recommended by the European Federation of Psychologists' Associations (2013), the sample was—from an ethnicity perspective—rather homogenous, and there were somewhat more females (about 55%) than males, which has to be considered when external validity is interpreted. Second, in all studies, it is always a risk that construct–irrelevant variance becomes part of the analyses. For instance, filling in the questionnaire too hastily could reflect respondents' low motivation. However, when the instrument was distributed, the impression was that the students took their time to fill in the questionnaire, and they did so in a serious manner. Third, Cronbach's alpha has been criticized as a measure of internal consistency, particularly when instruments utilize many items, when scales exceed 40 items, or when these scales feature highly inter-correlated items (Field, 2013). In the present study, the METQ instrument was based on 30 items that were not highly inter-correlated; therefore, Cronbach's alpha was considered to be a reasonable estimate of internal consistency. Finally, the psychometric evaluation of the METQ has convincingly shown evidence of discriminant validity, that is the result of METQ is weakly related to cognitive ability, such as fluid intelligence and WMC, and to the personality trait of Grit (or persistence). However, there is a lack of evidence regarding convergent validity of the results of the METQ. Measures of, for example, other motivational traits would have been needed to give some insight in the convergent validity of the METQ. Further, the criterion used as indicator of criterion-related validity (i.e., combined grades from the final (ninth) year in compulsory school) might be questioned as it portrays post-dictive validity rather than concurrent or predictive validity. The grades were attained about a year or two before the students participated in this study. However, post-dictive studies are not unusual according to the EFPA guidelines (European Federation of Psychologists' Associations, 2013), and given that the combined grades are valid, the relation found between the results of METQ and grades is relevant.

Conclusion

The overall conclusion is that the evaluation of the quality of the short version of the METQ was found to be acceptable and therefore applicable for use in an educational context. These results offer opportunities to carry out further research on the NFC, using the METQ, and this concept's importance for educational attainments in Sweden. With regard to the present study, and previous studies, it seems reasonable to assume that individuals with higher NFC will outperform individuals with lower NFC on school tasks requiring a lot of thinking, such as mathematical problem solving, reading comprehension, and natural science issues, but not on tasks that are less effortful (Leone and Dalton, 1988). In a recent publication (Luong et al., 2017) it was shown that the association between NFC and academic achievement increases from grade 3 to grade 9. To the extent that NFC is important already at elementary school level (or even earlier) and that NFC is regarded as a trait-based characteristic a question arises; is it possible to encourage and improve NFC in an early age and maintain this disposition over the school years? Cacioppo et al. (1996) showed in their review that individuals with high NFC experience more cognitive effort and engaged in more thoughts than those with low NFC. However, when situational factors required considerable mental effort, the individuals with low NFC reported equal high or even higher effort. Given the validity of these finding the answer to the question is yes—NFC can be improved. However, it will require

an understanding among educators that NFC is an individual difference characteristic, and that the situational factors has to be explicit in terms of how and when mental engagement and effort is expected, otherwise some of the students will choose not to engage in cognitive elaboration and thinking. Ultimately, measures of NFC can provide unique information that can be complementary to other measures or information, and thus, be useful for understanding student's progress. It is however important to note that there is a need to replicate the results found in the present study in other age groups and in more representative samples.

ETHICS STATEMENT

This study was carried out in accordance with the recommendations of the Swedish Central Ethical Review Board with written informed consent from all subjects. All subjects gave written informed consent in accordance with the Declaration of Helsinki; the protocol was approved by the Swedish Central Ethical Review Board (2015/328-31Ö).

AUTHOR CONTRIBUTIONS

Both authors, TS and BJ, contributed to the design of the study. They wrote the introduction and discussion in the manuscript together. TS wrote materials and methods and results, and made the statistical analyses and interpretations, with support from the second author (BJ). Both authors also contributed with revising and improving the manuscript and approved the final version of the manuscript. Both authors are accountable for all aspects of the work.

ACKNOWLEDGMENTS

We would like to thank the students for participating in and assisting with the present research.

FUNDING

This research was funded by the Swedish Research Council, grant number 721-2014-2099.

REFERENCES

Bertrams, A., and Dickhäuser, O. (2010). University and school students` motivation for effortful thinking. Factor structure, reliability, and validity of the German need for cognition scale. *Eur. J. Psychol. Assess.* 26, 263–268. doi:10.1027/1015-5759/a000035

Bless, H., Wänke, M., Bohner, G., Fellhauer, R. F., and Schwarz, N. (1994). Need for cognition: Eine skala zur erfassung von engagement und freude bei denkaufgaben. *Zeitschrift fur Sozialpsychologie* 25, 147–154.

Bråten, I., Anmarkrud, O., Brandmo, C., and Stromso, H. I. (2014). Developing and testing a model of direct and indirect relationships between individual differences, processing, and multiple-text comprehension. *Learn. Instruct.* 30, 9–24. doi:10.1016/j.learninstruc.2013.11.002

Cacioppo, J. T., and Petty, R. E. (1982). The need for cognition. *J. Pers. Soc. Psychol.* 42, 116–131. doi:10.1037/0022-3514.42.1.116

Cacioppo, J. T., Petty, R. E., Feinstein, J. A., Blair, W., and Jarvis, G. (1996). Dispositional differences in cognitive motivation: the life and times of individuals varying in need for cognition. *Psychol. Bull.* 119, 197–253. doi:10.1037/0033-2909.119.2.197

Cacioppo, J. T., Petty, R. E., and Kao, C. F. (1984). The efficient assessment of need for cognition. *J. Pers. Assess.* 48, 306–307. doi:10.1207/s15327752jpa4803_13

Cohen, A. R., Stotland, E., and Wolfe, D. M. (1955). An experimental investigation of need for cognition. *J. Abnorm. Soc. Psychol.* 51, 291–294. doi:10.1037/h0042761

Costello, A. B., and Osborne, J. W. (2005). Best practice in exploratory factor analysis: four recommendations for getting the most from your analysis. *Pract. Assess. Res. Eval.* 10, 1–9.

Curseu, P. L. (2011). Need for cognition and active information search in small student groups. *Learn. Individ. Differ.* 21, 415–418. doi:10.1016/j.lindif.2011.02.005

Dai, D. Y., and Wang, X. (2007). The role of need for cognition and reader beliefs in text comprehension and interest development. *Contemp. Educ. Psychol.* 32, 332–347. doi:10.1016/j.cedpsych.2006.05.002

Dornic, S., Ekehammar, B., and Laaksonen, T. (1991). Tolerance for mental effort: self-ratings related to perception, performance and personality. *Pers. Individ. Dif.* 12, 313–319. doi:10.1016/0191-8869(91)90118-U

Duckworth, A. L., Peterson, C., Matthews, M. D., and Kelly, D. R. (2007). Grit: perseverance and passion for long-term goals. *J. Pers. Soc. Psychol.* 92, 1087–1101. doi:10.1037/0022-3514.92.6.1087

Duckworth, A. L., and Quinn, P. D. (2009). Development and validation of the short grit scale (Grit-S). *J. Pers. Assess.* 91, 166–174. doi:10.1080/00223890802634290

European Federation of Psychologists' Associations. (2013). *EFPA Review Model for the Description and Evaluation of Psychological and Educational Tests (Version 4.2.6.)*. EFPA Board of Assessment.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*, 4 Edn. London: SAGE.

Fleischhauer, M., Enge, S., Brocke, B., Ullrich, J., Strobel, A., and Strobel, A. (2010). Same or different? Clarifying the relationship of need of cognition to personality and intelligence. *Pers. Soc. Psychol. Bull.* 36, 82–96. doi:10.1177/0146167209351886

Fleischhauer, M., Strobel, A., and Strobel, A. (2015). Directly and indirectly assessed need for cognition differentially predict spontaneous and reflective information processing behavior. *J. Individ. Differ.* 36, 101–109. doi:10.1027/1614-0001/a000161

Fosterlee, R., and Ho, R. (1999). An examination of the short form of the need for cognition scale applied in an Australian sample. *Educ. Psychol. Meas.* 59, 471–480. doi:10.1177/00131649921969983

Hill, B. D., Foster, J. D., Elliot, E. M., Shelton, J. T., McCain, J., and Gouvier, W. D. (2013). Need for cognition is related to higher general intelligence, fluid intelligence, and crystallized intelligence, but not working memory. *J. Res. Pers.* 47, 22–25. doi:10.1016/j.jrp.2012.11.001

Hill, B. D., Foster, J. D., Sofko, C., Elliot, E. M., and Shelton, J. T. (2016). The interaction of ability and motivation: average working memory is required for need for cognition to positively benefit intelligence and the effect increases with ability. *Pers. Individ. Dif.* 98, 225–228. doi:10.1016/j.paid.2016.04.043

Hu, L., and Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: conventional criteria versus new alternatives. *Struct. Equ. Modeling* 6, 1–55. doi:10.1080/10705519909540118

Leone, C., and Dalton, C. H. (1988). Some effects of the need for cognition on course grades. *Percept. Mot. Skills* 67, 175–178. doi:10.2466/pms.1988.67.1.175

Luong, C., Strobel, A., Wollschläger, R., Greiff, S., Vainikainen, M.-P., and Preckel, F. (2017). Need for cognition in children and adolescents: behavioral correlates and relations to academic achievement and potential. *Learn. Individ. Differ.* 53, 103–113. doi:10.1016/j.lindif.2016.10.019

Maslow, A. H. (1943). Dynamics of personality organization, I & II. *Psychol. Rev.* 50, 514–558. doi:10.1037/h0062222

Meier, E., Vogl, K., and Preckel, F. (2014). Motivational characteristics of students in gifted classes: the pivotal role of need for cognition. *Learn. Individ. Differ.* 33, 39–46. doi:10.1016/j.lindif.2014.04.006

Muthén, L. K., and Muthén, B. O. (1998–2015). *Mplus User's Guide*, Seventh Edn. Los Angeles, CA: Muthén & Muthén.

O'Connor, B. P. (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. *Behav. Res. Methods Instrum. Comput.* 32, 396–402. doi:10.3758/BF03200807

Preckel, F. (2014). Assessing need for cognition in early adolescence validation of a German adaption of the Cacioppo/Petty scale. *Eur. J. Psychol. Assess.* 30, 65–72. doi:10.1027/1015-5759/a000170

Raven, J. C. (1990). *Advanced Progressive Matrices: Sets I, II*. Oxford: Oxford Univ Press.

Richardson, M., Abraham, C., and Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychol. Bull.* 138, 353–387. doi:10.1037/a0026838

Sadowski, C. J., and Gulgoz, S. (1992). Internal consistency and test-retest reliability of the need for cognition scale. *Percept. Mot. Skills* 74, 610. doi:10.2466/pms.1992.74.2.610

Sicilia, M., Ruiz, S., and Munuera, J. L. (2005). Effects of interactivity in a web site: the moderating effect of need for cognition. *J. Advert.* 34, 31–44. doi:10.1080/00913367.2005.10639202

Stenlund, T., Jönsson, F. U., and Jonsson, B. (2016). Group discussions and test-enhanced learning: individual learning outcomes and personality characteristics. *Educ. Psychol.* 1–15. doi:10.1080/01443410.2016.1143087

Tanaka, J. S., Panter, A. T., and Winborne, W. C. (1988). Dimensions of the need for cognition: subscales and gender differences. *Multivariate Behav. Res.* 23, 35–50. doi:10.1207/s15327906mbr2301_2

Unsworth, N., Heitz, R., Schrock, J., and Engle, R. (2005). An automated version of the operation span task. *Behav. Res. Methods* 37, 498–505. doi:10.3758/BF03192720

Venkatraman, M. P., and Price, L. L. (1990). Differentiating between cognitive and sensory innovativeness: concepts, measurement, and implications. *J. Bus. Res.* 20, 293–315. doi:10.1016/0148-2963(90)90008-2

Verplanken, B., Hazenberg, P. T., and Palenéwen, G. R. (1992). Need for cognition and external information search effort. *J. Res. Pers.* 26, 128–136. doi:10.1016/0092-6566(92)90049-A

von Stumm, S., and Ackerman, P. L. (2013). Investment and intellect: a review and meta-analysis. *Psychol. Bull.* 139, 841–869. doi:10.1037/a0030746

Waters, L. K., and Zakrajsek, T. (1990). Correlates of need for cognition total subscale scores. *Educ. Psychol. Meas.* 50, 213–217. doi:10.1177/0013164490501026

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.