



## Psychometric Properties of the Swedish Version of the Personality Inventory NEO-PI-3

Håkan Källmen<sup>1\*</sup>, Peter Wennberg<sup>2</sup>, Per Andreasson<sup>3</sup> and Hans Bergman<sup>4</sup>

<sup>1</sup>STAD-Centre for Psychiatry Research Karolinska Institutet/Stockholm County Council, Box 6031, 102 31 Stockholm, Sweden

<sup>2</sup>Centre for Social Research on Alcohol and Drugs, Stockholm University; Department of Clinical Neuroscience, Karolinska Institutet, Sweden; and Centre for Dependency Disorders, Stockholm, Sweden

<sup>3</sup>Hogrefe AB, Gävlegatan 12, Stockholm, Sweden

<sup>4</sup>Professor emeritus, Department of Clinical Neuroscience, Karolinska Institutet, Sweden

\*Corresponding author: Hakan Kallmen, Department of Clinical Neuroscience, Karolinska Institutet, Sweden, Tel: +46-73-7218410, Fax: +46-8- 12349784, E-mail: [Hakan.Kallmen@ki.se](mailto:Hakan.Kallmen@ki.se)

### Abstract

**Background:** The NEO-PI and its successors (NEO-PI-R and the latest NEO-PI-3) are among the most commonly used instruments to assess personality internationally. The American version of NEO-PI-3 is adapted to modern language and is more readable than its predecessors.

**Aims:** The present study describes the development of the Swedish version of the NEO-PI-3 and its psychometric properties.

**Methods:** A web-based questionnaire was sent to a random sample from the Swedish population and 18.5% responded. Reliability, construct validity and congruence with NEO-PI-R were calculated.

**Results:** Compared to NEO-PI-R the responses to NEO-PI-3 were more consistent resulting in a higher internal reliability, but the fit of the Five Factor Model to data was poorer. These results could, however, be due to a biased sample. The congruence between the versions were high.

**Conclusion:** A small adjustment of some items and the re-formulation of 35 items resulted in an improved internal reliability of the factors but a poorer fit to the five-factor model, and, consequently poorer construct validity. The NEO-PI-3 is more readable and easier to understand to young responders.

### Keywords

NEO-PI-3, Swedish population, Reliability, Validity, Congruence

### Background

The five factor model of personality is derived from factor-analyses of trait-descriptive adjectives in the English language, and, according to this model the perception of words describing personal characteristics can be divided into five groups [1]. This personality taxonomy has been called the “Big Five” and consists of factors describing degrees of Neuroticism, Extraversion, Openness to Experience, Agreeableness, and Conscientiousness. The model was considered as a “human universal”, valid in all cultural contexts and invariant over age and gender [2]. If

this holds true, an inventory based on the taxonomy should work well in different languages and cultures, which motivates the present study. Based on the five-factor taxonomy, Costa and McCrae developed a personality inventory called NEO-PI-R consisting of 240 items grouped in six facets for each of the five factors [3]. The test was developed for adult respondents but, since there is a need to be able to use the test for young people of the ages from 12 to 15 years a revision was required. Furthermore, the item wordings have not been adapted to the changes in the language used by youth. To find out if the item wordings should be changed and, if so, which items, a study of the item readability of NEO-PI-R was conducted. In a study of 500 adolescents aged from 14 to 20 years the respondents both rated themselves and were peer-rated. It was found that 37 items were less readable and decreased the validity among youth. These items were reworded to improve the psychometric properties and the new test was called NEO-PI-3 [4,5]. In a European project, research groups in several countries have developed national versions of the NEO-PI-3 based on the American version.

### Aims

The aim of this study was to develop a Swedish version of NEO-PI-3 and to describe its psychometric properties in a random sample from the Swedish general population. More specifically, the Swedish wordings of the 37 items suggested by [4] were to be tested.

### Development of the Swedish Version

The 37 new American items were translated and discussed by the authors until an agreed upon translation was set. In addition, 18 other items of the Swedish NEO-PI-R were slightly adjusted to a more common language. Further, the 37 new items were also tried out for readability among a group of persons in the lower ages (12-14 years). Some of the items were slightly changed on the basis of their responses.

### Material and Methods

#### Participants

A random sample of 3025 persons in the ages 16-75 yrs. was

drawn from a national register containing all persons with an address registered in Sweden (SPAR). Information regarding the study and an invitation to participate and anonymously respond to the questionnaire was sent by post together with a login code to be used when responding to the web-based questionnaire. If preferred, there was also an offer to respond to a paper version of the questionnaire. Twenty-nine persons did so. After two reminders there were 560 responders, a response rate of 18.5%, but 24 persons were excluded due to more than 41 unanswered items. The responders are probably a selected group which may have an effect for the generalization of the results. The analyses were based on 536 responses (306 women and 230 men). The study was approved by the Regional Ethical Review Board in Stockholm (number 2011/1285-31, 2012/1364-32).

## Questionnaire

Altogether 277 items, 240 from the Swedish NEO-PI-R, including the 18 adjusted items, and the 37 new items, were included in the questionnaire. Two reminders two weeks apart were sent to non-responders. The items of NEO-PI-3 describe eligible feelings and behavior in different situations to which responders are to estimate how well the descriptions fit their actual feelings and behavior. The estimations are made on a 5-point Likert scale coded 0-4. One hundred and five items were reversed when presented.

## Statistics

Descriptives, tests of the effect of gender and calculations of Cronbach alpha based on the 6 facets in each factor were made in SPSS.22. Tests of the significance between differences in alpha coefficients, that are measures of internal consistency, were made by 95% confidence intervals (not shown). If those intervals overlap, there is no statistical difference. Part-whole corrected item-total correlations were computed for each of the new items with its corresponding facet and compared to the corrected part-whole correlation for the old item. Confirmatory Factor Analysis was performed in STATA12.0. The significance of the fit of the "Five Factor Model" with 5 correlated factors each composed of 6 facets to the structure of data was tested with the  $\chi^2$ -test. Root Mean Square Error of Approximation (RMSEA), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) were computed. We also used the SPSS-syntax from the American manual [5], to calculate Procrustes rotated congruence coefficients. The congruence coefficients (Cc) between the Swedish versions of NEO-PI-3 and NEO-PI-R were calculated. To compare the descriptives from the previous study of NEO-PI-R [6] with the results from the present study, 95% confidence intervals for the mean total score in each of the five factors were computed for each sex separately.

**Table 1:** Cronbach's alpha, means and standard deviations for men and women on factors and facets in NEO-PI-3. Congruence coefficients (Cc) between the Swedish versions of NEO-PI-3 and NEO-PI-R.

Factor/facet	Total sample		Men		Women		P	Cc	
	Cronbach $\alpha$	Mean	Standard deviation	Mean	Standard deviation	Mean			Standard deviation
Neuroticism	0.87	75.1	24.7	69.8	23.3	79.0	25.1	< 0.001	0.99
Extraversion	0.78	110.5	22.9	108.6	22.1	112.1	23.5	0.197	0.98
Openness	0.74	111.8	21.8	107.1	23.1	115.5	20.0	< 0.001	0.99
Agreeableness	0.76	128.3	19.5	121.7	18.3	133.3	18.9	< 0.001	0.99
Conscientiousness	0.82	126.3	20.0	125.6	19.9	126.8	20.1	0.530	1.00
<b>N-facets</b>									
Anxiety	0.86	13.1	6.2	11.5	5.7	14.4	6.3	< 0.001	1.00
Angry hostility	0.74	11.5	4.8	10.9	5.0	12.0	4.6	0.021	0.99
Depression	0.85	13.7	6.0	12.4	5.8	14.6	6.1	< 0.001	1.00
Self-consciousness	0.77	12.4	5.3	11.7	5.2	12.9	5.4	0.006	0.99
Impulsiveness	0.72	14.4	4.7	13.9	4.7	14.8	4.7	0.035	0.99
Vulnerability	0.76	10.4	4.4	9.6	4.3	11.1	4.4	< 0.001	1.00
<b>E-facets</b>									
Warmth	0.80	22.2	4.9	21.2	4.5	23.0	5.0	< 0.001	0.99
Gregariousness	0.81	18.2	5.8	17.8	5.9	18.6	5.8	0.099	0.97
Assertiveness	0.82	16.2	5.6	16.9	5.6	15.7	5.6	0.007	0.96
Activity	0.70	16.6	4.6	15.9	4.4	17.1	4.7	0.003	0.98
Excitement-seeking	0.69	14.6	5.7	16.1	5.4	13.5	5.7	< 0.001	0.95
Positive Emotions	0.84	22.6	5.8	21.0	5.8	23.8	5.5	< 0.001	0.99
<b>o-facets</b>									
Fantasy	0.79	17.6	5.6	16.8	5.5	18.2	5.6	0.001	0.99
Aesthetics	0.84	15.1	7.0	13.8	7.1	16.1	6.8	< 0.001	0.99
Feelings	0.79	21.4	4.9	19.3	5.0	23.0	4.2	< 0.001	0.99
Actions	0.78	16.5	5.2	16.4	5.3	16.5	5.0	0.710	0.98
Ideas	0.78	18.1	5.9	18.3	6.1	17.9	5.8	0.422	10.0
Values	0.59	23.0	3.7	22.5	7.3	23.4	3.6	0.017	0.94
<b>A-facets</b>									
Trust	0.84	20.9	5.0	20.2	4.9	21.6	5.0	0.003	0.99
Straight-forwardness	0.75	21.6	5.1	20.4	5.2	22.6	4.9	< 0.001	0.97
Altruism	0.74	25.1	3.8	24.0	3.7	25.9	3.7	< 0.001	0.99
Compliance	0.72	18.2	4.8	17.5	4.9	18.7	4.6	0.004	0.99
Modesty	0.79	20.3	5.1	19.3	5.3	21.1	4.9	< 0.001	0.98
Tendermindedness	0.71	22.3	4.5	20.6	4.5	23.6	4.0	< 0.001	0.97
<b>C-facets</b>									
Competence	0.71	23.0	3.6	23.2	3.4	22.9	3.8	0.494	0.99
Order	0.73	19.8	4.8	19.1	4.6	20.3	4.8	0.002	0.99
Dutifulness	0.68	25.1	3.6	24.5	3.6	25.6	3.5	< 0.001	0.97
Achievement-striving	0.75	18.0	4.9	18.4	4.6	17.6	5.1	0.074	0.97
Self-Discipline	0.84	21.0	5.2	21.1	5.1	21.0	5.2	0.808	0.99
Deliberation	0.81	19.3	5.0	19.5	5.1	19.2	4.9	0.426	0.98

**Table 2:** Confirmatory factor analysis of NEO-PI-3 testing the hypothesis of facets loading in five factors. Chi-square, p, degrees of freedom, Root mean square error of approximation, Comparative Fit Index, Tucker-Lewis Index

Factor	Chi-square	Significance level	Df	RMSEA	CFI	TLI
N	119.01	P < 0.001	9	0.15	0.94	0.89
E	202.51	P < 0.001	9	0.20	0.81	0.68
O	66.24	P < 0.001	9	0.11	0.91	0.85
A	129.92	P < 0.001	9	0.16	0.84	0.73
C	121.77	P < 0.001	9	0.15	0.90	0.84

## Results

Means and standard deviations for men and women on the factors and facets of NEO-PI-3 based on the Swedish population sample are presented in table 1. Thirty-five of the 37 new items suggested by [4] had a higher item-total correlation in the Swedish version than its corresponding old item. Women scored higher than men did on the Neuroticism, Openness to Experience and Agreeableness factors. There were no gender differences in Extraversion and Conscientiousness. On facet level men scored higher than women only on “assertiveness” and “excitement-seeking” belonging to Extraversion. In the facets “gregariousness”, “actions”, “ideas”, “competence”, “achievement-striving”, “self-discipline” and “deliberation” there were no significant gender differences. Women scored significantly higher than men on the 21 other facets.

Foremost Neuroticism (-0.29) and Extraversion (-0.15) were negatively associated to age but Agreeableness (0.23) and to some degree Conscientiousness (0.09) were positively associated to age. This shows that the association between age and personality is weak or moderate.

A test of the construct validity of the Swedish NEO-PI-3 was made using a confirmatory factor analysis. The construct validity of the factors of NEO-PI-3 was low (Table 2) and lower than that of NEO-PI-R [6]. According to [7] the Root Mean Square Error of Approximation (RMSEA) should be below 0.08 and both CFI and TLI should be equal to or above 0.90. RMSEA indicates that the Openness factor was best fitted to the data. However, the CFI and TLI indicate that the Neuroticism factor was best fitted. An explorative principal axis factoring with promax rotation showed an overlap between Openness to Experience and Extraversion and between Agreeableness and Extraversion. According to 95% Confidence Intervals (not shown) there was a non-significant tendency towards higher internal reliabilities for four of the five factors and twenty-one of the 30 facets in NEO-PI-3 as compared to NEO-PI-R. The equivalence between NEO-PI-R and NEO-PI-3 was found to be high since all congruence coefficients except one was above 95.

Ninety-five percent confidence intervals showed that both women and men scored significantly higher on Openness to Experience and men also on Conscientiousness in the NEO-PI-3 sample as compared to the NEO-PI-R sample. A reversed trend was found on the Neuroticism factor.

## Discussion

It was found that the Cronbach alpha coefficients for four of the factors were somewhat but non-significantly higher except for Openness when the Swedish versions of NEO-PI-3 and NEO-PI-R were compared. However, the construct validity was lower for NEO-PI-3. This implies that the respondents perceived the items of a specific factor and its facets as more related and responded more consistently after the change of 35 items in NEO-PI-3 and a slight rewording of 18 items. The congruence coefficients were high and indicated a high similarity between both Swedish versions. However, Confirmatory Factor Analysis indicated a lower fit of the five-factor model of personality to the covariance structure of the data generated from NEO-PI-3 as compared to NEO-PI-R.

When interpreting the higher reliability of NEO-PI-3 as compared to NEO-PI-R the low response rate must be considered (18.5%). The previous study of NEO-PI-R was carried out on a paper version about

15 years ago with a response rate of 61%. Both studies were carried out on samples of the general Swedish population. Non-responding is a great problem nowadays in surveys, particularly when internet-based. The difference in response rate due to type of administration was evident in a study of the Alcohol Use Disorder Identification Test (AUDIT) [8]. The response rate was 54% for the paper version and only 27% for the internet version. The present sample responding to NEO-PI-3 was more biased in the direction of less Neuroticism and more Openness to Experience and Conscientiousness as compared to the previous study on NEO-PI-R. The much lower response rate in the present study may partly explain the higher internal-consistency reliability of NEO-PI-3. Thus, we cannot conclude that NEO-PI-3 should be more reliable than NEO-PI-R. The lower fit of NEO-PI-3 to the five factor model of personality should also be considered. However, the high congruence between the Swedish NEO-PI-3 and NEO-PI-R indicates that the difference in factor structure is negligible and the updated wording speaks in favour of using the NEO-PI-3.

## Acknowledgements

The authors are grateful to the Swedish Research Council for Health, Working life and Welfare for supporting this study, grant nr 2009-1705.

## Disclosure of Interest

Hogrefe AB who is the publisher of the Swedish NEO-PI-3 was responsible for the collection of data and the third author who is employed by the publisher made the most of the calculations for table 1. The other authors have no interest to declare.

## References

1. Paul D Trapnell, Jerry S Wiggins (1990) Extension of the Interpersonal Adjective Scales to include the Big Five dimensions of personality. *Journal of Personality and Social Psychology* 59: 781-790.
2. McCrae RR, Costa PT (1997) Personality trait structure as a human universal. *Am Psychol* 52: 509-516.
3. Costa PT, McCrae RR (1992) Revised NEO Personality Inventory (NEO-PI-R) and NEO Five Factor Inventory (NEO-FFI) Professional manual. Odessa, FL: Psychological Assessment Resources.
4. McCrae RR, Costa PT, Martin TA (2005) NEO-PI-3: A more readable revised NEO personality inventory. *J Pers Assess* 84: 261-270.
5. McCrae RR, Costa PT (2010) NEO Inventories for the NEO personality Inventory-3 (NEO-PI-3), NEO Five Factor Inventory-3 (NEO-FFI-3), NEO-Personality Inventory-Revised (NEO-PI-R). Professional Manual: PAR.
6. Kallmen H, Wennberg P, Bergman H (2011) Psychometric properties and norm data of the Swedish version of the NEO-PI-R. *Nord J Psychiatry* 65: 311-314.
7. Schermelleh-Engel K, Moosbrugger H, Müller H (2003) Evaluating the fit of structural equation models: test of significance and descriptive goodness-of-fit measures. *Methods of Psychological Research Online* 8: 23-74.
8. Kallmen H, Sinadinovic K, Berman AH, Wennberg P (2011) Risky drinking of alcohol in Sweden: A randomized population survey comparing web-and paper based self-reports. *Nordic Studies of Alcohol and Drugs* 28: 123-130.