Abstract

The five-factor structure is a well-established model for personality. The five traits covary with job-performance and work-relevant outcomes. The practical administration of existing big-five measurement scales is however somewhat limited, in a Norwegian setting, as existing scales are impractically large or have unknown psychometric properties. Because of this, a new brief Norwegian personality assessment tool has been developed by the Norwegian Armed Forces. This study aims to uncover the psychometric properties of the 50-item Norwegian military personality inventory (NMPI-50) and establish norm data for practical use. The inventory was administered to the 2002 cohort of Norwegian 17-year old's (N=54,355), and analyzed with factor analysis, graded response models and tests of gender invariance. The five scales of the NMPI-50 showed satisfactory internal consistency, yielded high information across a broad range of the five traits, and conformed to a bi-factor structure with one general factor and five specific factors. The general factor was positively associated with motivation for military service, indicating some measurement bias. The openness scale is less clearly psychometrically defined, compared to the other scales, and both extroversion and openness show some evidence of multidimensionality. The scales also showed scalar invariance between genders except for the openness scale. Overall, the results support the use of NMPI-50 in personnel assessment and research.

Self-perceived personality can be captured using a limited number of factors, and a large body

of lexical and statistical research has converged on the Five-Factor Model (FFM) or the Big

Five traits as a comprehensive taxonomy (Widiger, 2017). The FFM represents a hierarchical

structure of personality where behaviors and very specific traits are described and subsumed

by broader facets of personality that covary and cluster into five factors. Following McCrae

(2010) and Goldberg (1993), the gist of the five main factors or domains can be described as

neuroticism vs. emotional stability, extraversion vs. introversion, openness to experience vs.

closedness to experience, agreeableness vs. antagonism, and conscientiousness vs. casualness.

These broad personality trait dimensions describe differences between individuals, and the

levels of traits together with the combination of them constitute the individual´s personality.

The FFM has been successfully applied in organizational settings (Siebert & DeGeest,

2015), as personality traits tend to predict job performance (Salgado, 1997).

Conscientiousness and neuroticism, in particular, are valid predictors of performance across

jobs, while the effect of extraversion, agreeableness, and openness to experience are

dependent on job type (Barrick, Mount & Judge, 2001). However, the predictive validity of

self-reported personality is often somewhat low, and its use in personnel selection decisions

has therefore been criticized (Morgeson et al., 2007b, 2007a). Nevertheless, research findings

tend to demonstrate an effect of personality traits toward job performance after controlling for

general mental ability (Ones, Dilchert, Viswesvaran & Judge, 2007; Schmidt & Hunter,

1998), thereby supporting the usage of personality measures for selection purposes.

Furthermore, personality traits are often included in job analyses (Ployhart, 2012), underlining

the relevance of personality in many selection decisions.

The results from the comprehensive U.S. Project A have been of importance for

military selection, demonstrating that personality scales predicted military outcomes such as

effort and leadership, personal discipline, and physical military fitness (Campbell & Knapp,

2001). Later meta-analyses using military samples have found results in line with civilian

FFM prediction patterns (Darr, 2011; Salgado, 1998). The Norwegian Armed Forces has used

FFM personality testing in personnel selections and developmental programs for many years

(Eid, Lescreve & Larsson, 2012). However, at the largest selection arena in Norway, the

conscript assessment procedure, personality characteristics are evaluated without the use of

personality measures (Køber, Lang-Ree, Stubberud & Martinussen, 2017), in part because of

a suitable measurement scale is lacking. Available standardized and translated instruments

such as the NEO PI-3 (McCrae & Costa, 2010), NEO PI-R (Costa & McCrae, 1992), and the

5PF military 2.0 (Engvik, 2005) have Norwegian population norms, but a large number of

items limit large scale use. The NEO-FFI, a validated factor-level version consisting of 60

items (McCrae & Costa, 2007), has an optimal length, but the proprietary nature of the

measure deems this scale unsuitable as well. Other brief Norwegian translated scales have

somewhat unsure psychometric properties as they are validated on small samples (Engvik &

Clausen, 2011, Engvik & Føllesdal, 2005). Valid interpretation and practical use of observed

scores is contingent on knowing the psychometric properties of the instrument. The impact of

social desirability and motivation for military service on indicator responses is especially

valuable information, if the instrument is to be used in personnel selection. Research into

response patterns on personality inventories indicates that faking and social desirability

response sets can be viewed as spurious measurement error caused by an interaction between

person and context (Ziegler & Buehner, 2009). Modeling responses, including social

desirability response sets as a general factor, in the same selection arena where the instrument

is implemented helps with interpretation of observed scores within this context. Because of

this, a new brief Norwegian FFM scale is needed for research purposes and large-scale

personnel assessment and selection within the Norwegian Armed Forces. Knowledge of how

indicator responses are influenced by military motivation, and gender measurement

equivalence is of particular interest. The present paper reports on the psychometric properties of a 50-item personality scale developed to meet this need. The Norwegian Armed Forces holds the proprietary rights, and the scale length is optimal considering conscript assessment feasibility.

Method

Scale development

The NMPI-50 was based on translated items from the International Personality Item Pool (Goldberg, 2006), as well as items suggested by experienced military psychologists. An initial pool of 79 items was administered to 850 officer applicants for initial validation purposes (Skoglund, 2017). These preliminary studies demonstrated expected and strong correlations with the NEO PI-3 factors, although the factor structure was suboptimal. The scale reduction from 79 to 50 items was based on results of rudimentary item analyses, internal scale consistency evaluations, and confirmatory factor analysis. The authors also performed a qualitative evaluation of the semantic content of items to ensure that the NMPI-50 scales are sensitive to multiple sub-facets within each of the five factors.

The NMPI-50 was translated to both Norwegian written official languages (Norwegian Bokmål/Nynorsk). A small portion (2.93%) of the present sample identified themselves as "Nynorsk" users and therefore completed the alternate translated version. A preliminary comparison of observed factor means revealed small but potentially confounding differences regarding the two language groups. Thus, we removed the small sample that used the "Nynorsk" version. The results reported in the present paper apply therefore to the "Bokmål" version.

Procedure and participants

The first part of the Norwegian conscription assessment procedure instructs military aged males and females (17 and 18-year-olds) to answer a mandatory online survey

questionnaire. The survey contains questions regarding health status, social functioning, motivation, service preferences, and physical fitness. The NMPI-50 was attached to this questionnaire for the 2002 cohort, which resulted in a sample of $N = 52,760$. The gender ratio was about 1:1, with 50.1% males. The total number of participants equals approximately 94% of the Norwegian cohort of 56,249 individuals born in 2002 (FHI, 2018). The survey was not sent to persons with a criminal record, or those otherwise deemed non-eligible for military service. The survey stated that the answers on the questionnaire, including the NMPI-50, should be responded to as honestly as possible, that answers should not be influenced by thoughts of military fitness or motivation, and that there were no right or wrong answers.

The Norwegian Armed Forces HR and Conscription Centre can collect personal information (including psychological variables) from Norwegian citizens for evaluating mandatory military service suitability (legally regulated in the Norwegian Compulsory Military Service Act). Research use of the data pertaining to the present study was formally approved by the independent Research Commission at the Norwegian Defense University College, and the authors of this study used anonymous data exclusively. Acknowledging ethical concerns considering the omission of voluntary consent, the authors legitimized this research first and foremost by evaluating the NMPI-50 as an unobtrusive measure. Measuring normal personality, we concluded that responding to the NMPI-50 did not have clear foreseeable potentials for any adverse consequences for the participants (e.g., long term negative psychological reactions). Should questions arise when responding to the mandatory survey questionnaire participants can communicate with HR and Conscription personnel by means of telephone or e-mail. Summarized, both an independent Research Commission and the present authors evaluated research use of the present data as ethically acceptable.

Measures

The items of the NMPI-50 are rated on a seven-point Likert scale, where respondents indicate their degree of agreement on 50 items (completely disagree to completely agree), of which 13 items are negatively formulated. The scale takes approximately 10-15 minutes to complete. Ten items measure the personality traits of extraversion, agreeableness, conscientiousness, emotional stability, and openness to experience, respectively. After reversing negatively formulated items, summed scores can be obtained for each of the five scales. Appendix A presents the items, and converted stanine- and T-score norms for use in administration and interpretation of the NMPI-50. Motivations for military service was measured with a single item: *I wish to serve in the Norwegian Armed Forces*. The response options were on a five-point scale ranging from *completely true*, to *completely untrue* and included a neutral *I don't know*.

Data analysis

The full sample ($N = 52,760$) was split into three equal parts to conduct exploratory factor analysis (EFA) ($N = 17586$), confirmatory factor analysis (CFA) ($N = 17587$) and item response analysis ($N = 17587$). A total of 14 respondents without any variance in their responses were interpreted as invalid and removed from the dataset. The data did not contain missing responses. Cronbach's alpha, EFA, CFA, tests of measurement invariance and graded response models (IRT) were calculated using STATA 16 (Statacorp, 2017).

The exploratory factor analysis used principal factors. The solution was rotated using promax oblique rotation that presupposes correlated factors. We determined the number of retained factors using parallel analysis (Horn, 1965), which compares the observed eigenvalues with those obtained from 100 replications (Monte Carlo simulations) of uncorrelated normal variables. A factor is retained when the difference from its associated eigenvalue is bigger than the 95th percentile of the distribution of eigenvalues derived from the random data. When the optimum number of factors was identified, we also investigated

the unique contribution of an additional factor.

The CFA used maximum likelihood estimation to determine model fit. Model fit was tested for each factor as well as for the different factor solutions of an a priori five-factor solution, bi-factor solution and a three-factor solution based on the Eysenck assertion of broad traits (Eysenck, 1975). The three-factor solution contained one factor for neuroticism and extraversion and another for conscientiousness and agreeableness. The final fifth factor of openness is not subsumed by either factor and is therefore analyzed as a third separate factor in both the three- and five-factor solutions. The bi-factor solutions contained five specific factors and a general factor. After fitting the bi-factor model, we correlated the general factor to motivation for military service to assess the relationship between the general factor and an indicator of social desirability. All factor solutions included covariance parameters between latent factors. Modification indices were used to determine the presence of correlated error variance. We assessed model fit with the following indicators: chi-square (χ^2), comparative fit index (CFI), root mean square error (RMSEA), standardized root mean square residual (SRMR) and Tucker-Lewis index (TLI). There are several suggestions for cut-off values for poor, moderate, and satisfactory model fit, we chose to set satisfactory cut-offs for RMSEA at < .08. This was based on recommendations from Maccullum, Brown, and Sugewara (1996) where 0.01, 0.05, and 0.08 correspond to excellent, good, and mediocre fit, respectively. The cut-off value for SRMR was set at < .08 based on recommendations from Hu and Bentler (1999). CFI and TLI cut-off values were set at > .90 (Hooper et al., 2007). Based on the results from both the EFA and the CFA we also determined the presence of multidimensionality in single factors with poor fit, using bi-factor models (Reise, Bonifay & Haviland, 2013). Multidimensionality was assessed with the factor determinacy coefficients using the FSDET module for STATA (Mehmetoglu, 2015). A factor determinacy coefficient > 0.90 indicates a unique factor (Grice, 2001).

The Graded response model (GRM) is based on item response theory, which is an overlapping theoretical framework to Classical Test Theory (Hulin et al., 1983). The primary function of the application of GRMs regarding personality factors is the avoidance of sample sensitive item parameters (Hambleton & Rogers, 1989). In addition, the discrimination parameters and test information curves reveal at what levels of the trait the scale is sensitive. We calculated at what range all five scales have a standard error of measurement of less than 0.5, for determination of range sensitivity. The GRMs were estimated for each of the five factors. We calculated one difficulty parameter (β) for each item threshold (response options), as well as an item discrimination (α) parameter for each item. Item discrimination values from 0.01 to 0.24 are considered very low, 0.25-0.63 low, 0.65-1.34 moderate, 1.35-169 high, and above 1.7 very high (Baker, 2001).

Lastly, we tested the five scales for the presence of both metric and scalar invariance to investigate the presence of measurement equivalence between males and females and evaluate gender differences in observed scores. We first fit a two-group baseline model for each of the five factors. In this baseline model, all parameters are freely estimated for males and females. To achieve model identification, we fixed the first item loading to 1 and intercept at zero for a free estimation of factor means. We further fitted a metric invariance model where coefficients to the latent trait were not allowed to vary across genders, and we compared this model to the baseline model. If the criterion for metric invariance was not met, a partial metric model was fitted. In this model, one item coefficient was allowed to vary freely across genders. We used modification indices to determine what coefficients would increase model fit the most when estimated freely across genders. After assessing metric invariance, we tested for the presence of scalar invariance across genders. A model was fitted with both constrained coefficients and intercepts to the latent trait, across genders. If the constrained model did not achieve invariance, we allowed one intercept parameter to vary

across genders, based on modification indices. Thus, both metric and scalar invariance could be achieved fully, partially, or not at all. We did not estimate strict measurement invariance, which includes equal residual variances, as strict invariance is rarely achieved in an applied context (Van De Schoot et al., 2015). The very large sample size renders significance testing with likelihood ratio (Chi-square) tests not practical, as small and trivial differences result in significant differences between models (Brannick, 1995; Kelloway, 1995). We instead reject the null hypothesis of gender invariance if the CFI is smaller than or equal to -0.01, based on Cheung and Rensvold´s (2002) recommendations. If the CFI difference was found to be larger than -0.01, the scale was deemed to not achieve measurement invariance. A lack of metric invariance makes it difficult to interpret relationships with other external across genders (Putnick & Bornstein, 2016). Lack of scalar invariance indicates that differences in observed gender scores are due to measurement issues, and not gender difference in personality and that scores on the trait cannot be directly compared across genders (Marsh et al., 2018).

Results

Means, standard deviations, Cronbach's alphas, and bivariate correlations are presented in Table 1. The alpha coefficients are all above an acceptable threshold, but the openness scale demonstrates lower interitem reliability compared to the other scales. The correlations between factors are comparable to other self-reports (DeYoung, 2006), except for a notably low correlation between openness and emotional stability. The high mean of agreeableness indicates a slight ceiling effect and some resulting lack of discrimination in the upper ranges of the scale. The notable associations between motivation for military service and all five scales illustrates how personality may inform selection decisions. The nature of these relationships is however best understood in a bi-factor model, presented below.

--INSERT TABLE 1 HERE--

Exploratory- and confirmatory factor analysis

The results of the exploratory factor analysis using parallel analysis revealed five factors with an eigenvalue > 95th percentile of the eigenvalues from random data. The parallel analysis plot can be viewed in appendix B (Figure I). The first five factors explained 59.5%, 18.6%, 8.9%, 6.7% and 5.3% of the variance, respectively. Adding a sixth factor in the unrotated solution explains an additional 2.6% of the variance. The results from the rotated factor solution are presented in Table 2. Most of the items load on their corresponding factors to a satisfactory degree, with the exemption of two items intended for the openness scale. Item o2(*desire for self-development*) and item o7(*familiarity with words and concepts*) cross-loads on conscientiousness. In addition, three items from the openness scale: o3(*taking the time to reflect*), o1(*notices beauty*) and o10(*curious about other culture*) do not robustly load on the openness factor. The sixth factor divides the openness scale into two separate factors, suggesting multidimensionality. Items o9(*Thinks creatively*), o6(*Generates ideas easily*), and o5(*Vivid imagination*) loads on the sixth rather than the fifth factor. The shared semantic content of these items is related to self-described intellect/mental ability.

--INSERT TABLE 2 HERE--

Correlations from latent traits after CFA are shown in Table1. They showed a notable increase in association between openness and extroversion, as well as openness and emotional stability, compared with correlations of observed openness scores. This discrepancy indicates the presence of measurement error in the openness factor. Presented in Table 4 are the fit indicators for confirmatory factor analyses for all five factors, as well as the bi-factor, five-factor and three-factor solutions. The bi-factor model with one general factor and five specific factors achieved the best model fit. Based on both RMSEA and SRMR indices, the bi-factor model showed an overall good fit with the data. CFI and TLI did not reach the threshold. This is likely due to the presumed low average size of the correlations between different factor indicators and especially due to measurement error and cross-loadings in the openness factor.

Fitting a bi-factor solution without indicators for openness yields satisfactory results in all fit

indices (RMSEA = .05, SRMR = .05, CFI = .91 & TLI = .90). The general factor is likely

comprised of a response set linked to social desirability. The results showed a notable positive

association between the general factor and motivations for military service (r = .44, p < .001,

95% CI = .43 - .45). Examining what indicators load strongest on the general factor reveals

that the highest loading indicators contain lexically evaluative language: e10 (*Enjoys*

leadership), e15 (*Likes to make decisions*) and e1(*View myself as an outgoing person*). Especially

in a military/leadership selection setting, where social dominance and extroversion may

consciously or non-conscientiously be viewed as favorable attributes. Openness indicators

have the lowest average loading on the general factor. Latent bivariate associations between

the five domains and motivation for military service change in important ways with the

inclusion of a general factor, as shown in table 3. The strength of the relationship increases for

emotional stability, and decreases the remaining domains. The associations between the latent

personality domains themselves were substantially altered after the exclusion of a general

factor as the average correlation was substantially reduced.

--INSERT TABLE 3 HERE--

The five-factor solution showed a better overall fit compared to the three-factor

solution, but did not reach the accepted cut-off for SRMR, as seen in Table 4. The

combination of a satisfactory RMSEA and non-satisfactory SRMR may be due to the large

number of variables modeled and the degrees of freedom (Savalei, 2012). The five individual

factors showed varying degrees of the goodness of fit. The traits of neuroticism,

conscientiousness, and agreeableness show moderate to good model fit, while extroversion

and openness are below the acceptable cut-off.

--INSERT TABLE 4 HERE--

Using modification indices, we assessed the effects of correlating indicators of error variances on the overall goodness of fit on the extroversion- and openness models. The presence of correlated error variance in pairs of indicators means the latent variable does not adequately explain some portion of shared variance within the pair. In the context of the five-factor personality taxonomy, a likely explanation is that a sub-facet of the personality trait is not adequately explained by the latent trait, but measured by a pair of indicators. Another possible explanation is that the content of one item in a pair is redundant (Byrne et al., 1993). Modification indices suggested correlating the error variances between items e5 (*Likes to make decisions*), e10 (*Enjoys leadership*), and e2 (*Becomes a leader frequently*). This model has an acceptable SRMR fit (0.056) but still has RMSEA above the threshold (0.112). The semantic content of these three items indicates the inclination towards assertive behavior and social dominance, a sub-facet of extroversion (Wilt & Revelle, 2015). The results from the modification indices of the openness factor suggested correlating error terms of the item pair o4 (*Fascinated by shapes and colors*) and o8 (*Appreciates art*), as well as the item pair o6 (*Generates ideas easily*) and o10 (*Curious about other cultures*). This model reaches acceptable SRMR fit (0.080) but still has RMSEA above the threshold (0.134). The former item pair likely reflects that the openness factor does not adequately explain the variance of the sub-facet aesthetic sensitivity. The latter pair is not easily interpretable but may reflect a sub-facet indicating a preference for variety or intellectual curiosity.

Based on the poor model fit found with CFA, we tested for the presence of multidimensionality on both the extroversion- and openness scale using bi-factor models. Adding another latent factor to the extroversion scale resulted in a notably improved model fit (χ^2= 11822(24), RMSEA= .097, SRMR= .031, TLI= .924 and CFI= .959). The added latent factor loaded strongly on the three leadership items (e2, e5, and e10) and the correlation

between the two latent factors was moderate (r = .55). The added latent factor showed a factor

determinacy coefficient of 0.908. The remaining factor, had a lower factor determinacy

coefficient of 0.894. Adding a second latent factor to the openness scale also improved model

fit, suggesting multidimensionality (χ^2= 16150(24), RMSEA= .097, SRMR= .031, TLI= .924

and CFI= .959). The second latent factor loaded strongly on items o9(*Thinks creatively*),

o6(*Generates ideas easily*), and o5(*Vivid imagination*), mirroring the EFA results. The added

latent factor (o5, o6 & o9) showed a factor determinacy coefficient of 0.901, the other latent

factor showed a factor determinacy coefficient of 0.885. The size of the correlation between

the two latent factors (B = .49) indicates that they reflect the same trait to a moderate degree.

Graded response model

Item parameters are displayed in Table 5. Mean item discriminations (α) varied across

the five factors. Agreeableness had the highest mean discrimination at 2.20 (logistic scale).

Openness showed the poorest mean discrimination at 1.57, which is in line with the factor

analysis results. Neuroticism also showed a high mean discrimination parameter at 2.13, as

did conscientiousness at 2.01. Finally, extroversion had the second poorest mean

discrimination at 1.91. Overall, the values suggest high to very high discrimination (Baker,

2001). A psychometrically robust scale also has high discrimination across a broad range of

the latent trait. Thus, the discrimination parameter values cannot be interpreted without

consideration of the latent trait range within which the items discriminate. This is indicated by

evenly distributed thresholds across the latent trait. Threshold parameters for response options

(β) varied across items and scales, as seen in Table 4. Several items exhibited a negative skew

indicating that most individuals are unlikely to endorse lower response options, especially in

the agreeableness sub-scale. The most difficult items -i.e. requiring high theta to endorse

positive response options - of the agreeableness scale were items a5(*interested in other people*)

and a3(*something nice to say about everyone*), whereas e10(*enjoys leadership*) and e3(*likes being*)

the center of attention) were the most difficult items for the extroversion scale. Items 3n(*fear*

embarrassment) and n2(*worry a lot*) were the most difficult from the neuroticism scale, and

items c6(*completes duties diligently*) and c7(*generally prepared*) for the conscientiousness scale.

Lastly, the most difficult items for the openness scale were o8(*appreciates art*) and o9(*thinks*

creatively). Figures IIa-e shows individual item characteristic curves and can be viewed in

appendix B.

--INSERT TABLE 5 HERE--

Graphs of the overall item information can be seen in Figure 1. All five scales yield

more information at lower levels of theta (θ) values. The emotional stability scale achieves

SE > 0.5 from θ = -3.1 to 2.3. The openness scale demonstrates the narrowest range of

acceptable test information, from θ = -2.9 to 1.7. In contrast, the agreeableness shows the

largest range of acceptable test information from θ = -3.9 to 2.0. The conscientiousness scale

achieves SE > 0.5 from θ = -3.6 to 2.1, and the extroversion scale from θ = -2.9 to 2.4. The

total information is consistently two standard deviations above and below the mean theta

value, except for the openness scale that is more restricted in range. The discrimination ability

of the agreeableness scale also drops off steeply at two standard deviations above the mean,

corresponding to the ceiling effect in observed scores. The results show that the accuracy of

scores positioned above and below two standard deviations on all five scales should be

viewed with caution as they have large confidence intervals.

--INSERT FIGURE 1 HERE--

Observed, metric and scalar gender differences

The gender differences in observed scores can be viewed in table 6 and figure 2. We

found that the mean female agreeableness score (M = 55.13, SD = 9.91) was slightly higher

than the mean male scores (M = 52.26, SD = 10.12). Males scored notably higher on

emotional stability (M = 46.98, SD = 11.54), compared to females (M = 37.42, SD = 12.37).

A slight gender difference was also found using observed extroversion scores, where the

males (M = 43.47, SD = 11.65) were found to be higher, compared to females (M = 41.84, SD

= 12.88). A very small mean difference in observed scores was also found in openness, where

females (M = 46.45, SD = 11.11) scored higher than males (M = 45.75, SD = 10.34). The

mean observed scores on conscientiousness for males (M = 49.11, SD = 10.50) and females

(M = 49.04, SD = 11.01) were found to be very similar.

--INSERT TABLE 6 HERE--

The means, standard deviations, and data distributions (as seen in Figure 2) reveal that

the difference in observed scores on emotional stability also applies to the distribution of

observed scores. Females have a higher spread of scores across a broader range of all traits

except for agreeableness. The ceiling effect of the agreeableness score restricts the range of

higher female scores more than males and likely contributes to a lower *SD* for females

compared to males. The results of the current study show that females have higher personality

variance on self-report measures confirming previous findings (Borkenau et al., 2013).

Research using non-self-report measures- i.e. other raters- suggest that this finding is due to

differences in self-rating, as males are found to be more variable when using non-self-report,

and females have more variability in rating others as well (Borkenau, McCrae & Terracciano,

2013).

--INSERT FIGURE 2 HERE--

Table 7 reports the findings for tests of measurement invariances between males and

females. All subscales achieved full metric invariance, meaning that the relationship between

scores on all five factors of the NMPI-50 and other variables are comparable across genders.

The results from the scalar invariance testing vary between the five personality scales.

Emotional stability, conscientiousness, and extroversion showed full scalar invariance

between genders. The openness subscale did not achieve either full or partial scalar

invariance. Items o8(*Appreciates art*) and o6(*Generates ideas easily*) in particular showed very high bias. Lastly, the agreeableness scale achieved partial scalar invariance, when the intercept for item a10(*Trusts others*) was allowed to vary freely across genders.

---INSERT TABLE 7 HERE---

The results suggest that there exists a systematic response bias in item a10 that contributes to a partially differing latent agreeableness scale for men and women. The same bias occurs in multiple openness items rendering the latent scale incomparable between genders. A summary of the NMPI-50 scale results is presented in Table 8.

---INSERT TABLE 8 HERE---

Discussion

The purpose of this study is to report on the psychometric evaluation of the NMPI-50. Individual differences in the five personality factors constitute potentially informative variables in military personnel selection and research, and several scales measure the traits of the FFM with robust psychometric properties. However, there exist no brief scales translated in Norwegian with thoroughly estimated properties for purposes of large-scale conscript assessment. The purpose of the NMPI-50 is to fill this gap and to secure that the Norwegian Armed Forces holds the proprietary rights. The present study also serves as a replication of the FFM of personality with a very large sample, by using both classical test theory and item response theory (graded response models), and as an investigation of the nature of the bifactor solution for FFM.

Internal consistency and multifactorial structure.

Overall, the results indicate that the NMPI-50 is a reliable scale when measuring the FFM of personality in a Norwegian setting, in which the 50 indicators show an acceptable fit to a bifactor solution. The bi-factor solution has important practical implications for

interpretations of observed scores. Indicator responses are based on two processes: The individuals personality traits and a general response pattern. The general response pattern is likely influenced by social desirability and is positively associated with the motivation to complete military service. This information should be taken into account when interpreting any individual summed observed scores that are the product of both a specific and the general factor. The relationship between the general factor and motivation for military services highlight the on-going debate on whether the general factor consists of method or substance (Chen et al., 2016). One interpretation of this finding is that the notable association between the general factor and motivation for military service is evidence of bias that systematically influences the overall response pattern via social desirability. Another is that the general factor does not necessarily affect social desirability, but rather reflects a source of important information that covary with motivations for military service, not captured by the five specific factors. The general factor has been positively associated with the belief that efforts pay off, engagement in socially sanctioned behaviors (Chen et al., 2016) as well as adaptive social functioning and lack of mal-adjustment (Watson et al., 1989). The results from the current study also imply that the general factor is partly substantive. There remains a very small correlation between latent extroversion, and motivation for military service in the bi-factor solution. An unlikely interpretation of these results is that the extroversion domain is not correlated with motivation for military service, and that the substantial correlation using observed scores is an artifact of measurement. Rather, it is likely that some of the variance in the general factor is substantive, and reflects important information and much of this variance comes from evaluative items on the extroversion scale.

 Nevertheless, previous research on bi-factor personality solutions in a selection setting indicates that the general factor also contains systematic social desirability bias (Ziegler & Buehner, 2009). For example, job applicants scored 1.31 standard deviations

higher on the general factor, compared to non-applicants (Anglim et al., 2017), indicating that

the general factor is largely comprised by a socially desirable response set. The finding that

the general factor explains less variance in indicators of openness is in line with previous

findings (Ziegler & Buehner, 2009). A likely explanation for this is that high scores on

openness indicators are not necessarily viewed as desirable traits. The measurement bias

represented by the general factor should dissuade the use of NMPI as the only selection tool

in high stakes selection arenas. Observed scores from low-stakes testing are also likely to be

somewhat lower, compared to high stakes testing, especially in the extroversion domain. The

substantial reduction in correlations from observed scores to latent scores in a bi-factor model

converge with previous research indicating that higher order factor correlations are largely

caused by item-level social desirability effects (Bäckström et al., 2009).

Not all goodness of fit indicators reached acceptable thresholds, although this is a

common finding for brief FFM scales. Indeed, even in longer scales, examinations of RMSEA

typically yield values ranging from 0.09 to 0.13, CFI from 0.61 to 0.79, and TLI values from

0.52 to 0.70 (Hopwood & Donnellan, 2010). Arguments have been made that the CFA

framework is too restrictive to evaluate personality scales (Marsh et al., 2010), and brief

scales have an inherent reliability disadvantage relative to longer personality scales (Mullins-

Sweatt & Widiger, 2006). It has also been argued that the inherent complexity of personality

often results in global evaluations of multiscale inventories not reaching the conventional cut-

offs for CFA (Hopwood & Donnellan, 2010). The CFA results bi-factor and five factor

solutions achieve better fit compared with other brief FFM scales (Baldasaro et al., 2013),

except for the openness scale. The EFA results revealed that most items loaded on their

primary factor, whereas a few items had cross-loadings – similar to previous research (Cooper

et al., 2010). The cross loadings were found on the openness scale, suggesting that this factor

is not as clearly defined as the other four. This finding mirrors other results with a

Scandinavian sample (Källmen et al., 2011).

Individual scale properties

The individual scales have varying properties. In general, brief scales balance between

representing a narrow range of semantic trait content and achieving higher reliability, with a

broad range and lower reliability. The scales for extroversion and openness were less well

defined by their indicators. The extroversion scale has robust internal consistency as

measured by Cronbach's alpha, no cross-loading items, a high mean discrimination

parameter, and yields satisfactory test information two standard deviations above and below

the mean theta values ($\theta = 0$). In contrast to this, the goodness of fit indicators that

emphasizes model parsimony (RMSEA), and unidimensionality (CFI and TLI), are above the

suggested cut-off values. The CFA results with a bi-factor solution substantially improved the

model fit, but only one of the factors achieved an adequate factor determinacy. Combined, the

results show some evidence of multidimensionality on two moderately correlated factors of

extroversion and leadership. The theoretical implication of this finding is that extroversion

and the tendency towards leadership are less clear in a Norwegian military aged sample.

Based on the moderate correlations between the two sub-factors, Cronbach's alfa, and the

theoretical link between them (Costa & McCrae, 1985), the extroversion scale is considered

to be a reliable, but multidimensional, measure. However, the lack of two factors achieving

factor determinacy, should dissuade the use of two extroversion factors in place of observed

scores.

The openness scale also demonstrated evidence of multidimensionality, where items

corresponding to semantic content describing intellectual curiosity was moderately correlated

with openness. Intellectual curiosity is not particularly evident in the openness factor in our

sample, and the overall moderate psychometric properties of the openness scale make this

distinction difficult to evaluate. In general, openness is a more complex factor and thus is not easily measured with brief scales. The exact nature and structure of the openness factor have been debated since its first description (McCrae & Costa, 1997), and it has been argued to be more closely described as an associate of the ability domain, rather than the personality domain (Ferguson & Patterson, 1998). The multidimensionality found in the present study suggests that openness and intellectual curiosity may not be equivalent in our sample, and only the intellectual curiosity factor achieved satisfactory factor determinacy. In general, cultural differences may produce different interpretations of the openness factor. This sentiment is echoed in other translated versions of FFM measures of openness (Barrio et al., 2004). The results in the current study suggest that the openness factor has a decent internal consistency and yields high test information. The use and application of the results from the openness scale of the NMPI-50 should be viewed with some caution however, as observed scores may measure intellectual curiosity, which is only moderately related to openness. Both agreeableness and conscientiousness scales demonstrate robust psychometric properties, although the range of the agreeableness scale is restricted in higher ranges of the trait, especially among female test-takers.

Gender differences in observed scores and measurement invariance

The multigroup analysis of gender invariance yielded encouraging results overall. Some differences were found in factor loadings, but these were not practically significant. The results from the scalar invariance tests were mixed. Agreeableness achieved partial scalar invariance and openness did not achieve scalar invariance. In comparison, extroversion, conscientiousness, and emotional stability achieved full scalar invariance. Scalar differences suggest that potential group differences are the product of measurement differences and not personality differences (Vandenberg & Lance, 2000). For the application of the NMPI-50, this means that scores on agreeableness scale should be evaluated with caution across

genders. The lack of any scalar invariance in the openness scale means that scores on this trait cannot be confidently compared across genders. Scores on extroversion, emotional stability, and Conscientiousness can be directly compared between genders. The presence of gender metric invariance in all five scales means that relationships between NMPI-50 scores and other variables are can be assessed for both genders and that the factor structure holds for both genders. The results add to the growing body of research demonstrating gender invariance for items from the MINI-IPIP database (Laverdière et al., 2013). The mean differences in personality factors between men and women were similar to results from comparable cultural samples (Källmen et al., 2011; Martinsen, 2005). Women were found to be somewhat more agreeable but this difference could partly be due to measurement differences. Women were also found to be slightly more open to new experience, compared to men, but this difference is strongly influenced by measurement differences.

Limitations, further research, and conclusion

The results of the current study confirm that NMPI-50 is a robust measure of the FFM. The bi-factor structure of the five traits and a general trait was confirmed in a sample of Norwegian military aged males and females. The emotional stability and conscientiousness scales have robust unidimensional psychometric properties across a broad range of the traits and observed scores can be compared across genders. The agreeableness scale also has robust unidimensional psychometric properties, but discrimination is limited in the upper ranges of the scale. The scale shows partial gender scalar invariance and comparisons of scores between genders should be done with caution. The extroversion scale has decent internal consistency and wide discrimination range, but some evidence of multidimensionality. Observed scores may measure the tendency towards leadership which is only moderately correlated with extroversion. The openness scale should be used with some caution. The results indicate that the underlying personality trait of openness is not fully captured with this brief scale, or that

openness is not as distinct of a trait in the Norwegian population, compared to samples from the United States (Costa & McCrae, 1985). Its observed score can be used across genders when examining the relationship with other variables, but direct comparisons of scores between men and women are not possible.

The main strength of this study is the large sample, while the main limitation is age representativeness in the sample. The NMPI-50 is currently ideal for young adults, but its psychometric properties are not demonstrated at younger and older ages. Efforts to test the properties of the scale and evaluate differential item functioning or invariance in older samples and non-military test administrations situations, as well as test for invariance across ethnic groups would improve the usability of the NMPI-50. Classical test-retest reliability and multilevel within and between effects (Geldhof, Preacher, & Zyphur, 2014) would also inform the reliability of the measure. Further research should aim to uncover the criterion validity of the NMPI-50 for selection and placements decisions in the military. Reliability is a necessary but not sufficient criteria for any valuable measure of personality. These efforts should include tests of agreement with other measures of the FFM as well as using NMPI-50 to make predictions regarding personality relevant measurable behavioral outcomes.

References

Anglim, J., Morse, G., de Vries, R. E., MacCann, C., & Marty, A. (2017). Comparing Job Applicants to Non-Applicants Using an Item-Level Bifactor Model on the HEXACO Personality Inventory. *European journal of personality*, 31(6), 669-684 doi:10.31219/osf.io/8j54g

Bäckström, M., Björklund, F. & Larsson, M.R., 2009. Five-factor inventories have a major general factor related to social desirability which can be reduced by framing items neutrally. *Journal of Research in Personality*, 43(3), 335–344.doi: 10.1016/j.jrp.2008.12.013.

Baker, F. B. (2001). *The basics of item response theory*. ERIC publications.

Baldasaro, R. E., Shanahan, M. J., & Bauer, D. J. (2013). Psychometric properties of the Mini-IPIP in a large, nationally representative sample of young adults. *Journal of Personality Assessment*, *95*(1), 74–84. doi:10.1080/00223891.2012.700466

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: a meta-analysis. *Personnel Psychology, 44(1)*, 1-26. doi:10.1111/j.1744-6570.1991.tb00688.x

Barrick, M. R., Mount, M. K., & Judge, T. A. (2001). Personality and performance at the beginning of the new millennium: What do we know and where do we go next? *International Journal of Selection and Assessment, 9*, 9-30. doi:10.1111/1468-2389.00160

Barrio, V. del, Aluja, A., & García, L. F. (2004). Relationship between empathy and the Big Five personality traits in a sample of Spanish adolescents. *Social Behavior and Personality: An International Journal*, *32*(7), 677–681. doi:10.2224/sbp.2004.32.7.677

Borkenau, P., McCrae, R. R., & Terracciano, A. (2013). Do men vary more than women in personality? A study in 51 cultures. *Journal of research in personality, 47(2)*, 135-144. doi:10.1016/j.jrp.2012.12.001

Borkenau P, Hřebíčková M, Kuppens P, Realo A, Allik J. (2013). Sex differences in variability in personality: a study in four samples. *Journal of Personality. 81(1)*, 49-60. doi:10.1111/j.1467-6494.2012.00784.x

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior*, *16*(3), 201–213. doi:10.1002/job.4030160303

Byrne, B. M., Baron, P., & Campbell, T. L. (1993). Measuring adolescent depression: Factorial validity and invariance of the Beck Depression Inventory across gender. *Journal of Research on Adolescence*, *3*(2), 127–143. doi:10.1207/s15327795jra0302_2

Campbell, J. P., & Knapp, D. J. (Eds.). (2001). *Exploring the Limits in Personnel Selection and Classification*. Lawrence Erlbaum Associates.

Chen, Z., Watson, P. J., Biderman, M., & Ghorbani, N. (2015). Investigating the Properties of the General Factor (M) in Bifactor Models Applied to Big Five or HEXACO Data in Terms of Method or Meaning. Imagination, Cognition and Personality, 35(3), 216–243. doi:10.1177/0276236615590587

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. doi:10.1207/s15328007sem0902_5

Cooper, A. J., Smillie, L. D., & Corr, P. J. (2010). A confirmatory factor analysis of the Mini-IPIP five-factor model personality scale. *Personality and Individual Differences*, *48*(5), 688–691. doi:10.1016/j.paid.2010.01.004

Costa, Paul T., & McCrae, R. R. (1985). *The NEO personality inventory*. Odessa, FL. Psychological Assessment Resources

Costa, P. T., & McCrae, R. R. (1992). *NEO PI-R Professional Manual*. Psychological Assessment Resources.

Darr, W. (2011). Military personality research: a meta-analysis of the Self Descriptive Inventory. *Military Psychology, 23(3)*, 272-296. doi:10.1080/08995605.2011.570583

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of Personality and Social Psychology, 91(6)*, 1138–1151. doi:10.1037/0022-3514.91.6.1138.

Eid, J., Lescreve, F., & Larsson, G. (2012). An international perspective on military psychology. In J. H. Laurence & M. D. Matthews (Eds.), *The Oxford Handbook of Military Psychology*. Oxford University Press.

Engvik, H. (1993). *5PF mil 2.0*. Forsvaret Forskningsinstitutt: Intern rapport: unpublished

Engvik, H., & Clausen, S. E. (2011). Norsk kortversjon av Big Five Inventory (BFI-20). *Tidsskrift for Norsk psykologforening, 48(9)*, 869-872.

Engvik, H., & Føllesdal, H. (2005). The Big Five Inventory på norsk. *Tidsskrift for Norsk psykologforening, 42(2)*, 128-129.

Eysenck, H. J., & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire (junior and adult)*. Hodder and Stoughton.

Ferguson, E., & Patterson, F. (1998). The five factor model of personality: Openness a distinct but related construct. *Personality and Individual Differences*, *24*(6), 789–796. doi:10.1016/s0191-8869(97)00241-9

FHI. (2018). *Medisinsk fødselsregister—Statistikkbank*. http://statistikkbank.fhi.no/mfr/

Geldhof, G. J., Preacher, K. J., & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods*, 19(1), 72-91. doi:10.1037/a0032138

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48(1)*, 26-34. doi:10.1037/0003-066X.48.1.26

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R., & Gough, H. C. (2006). The International Personality Item Pool and the future of public-domain personality measures. *Journal of Research in Personality, 40*, 84-96.

Grice, J. W. (2001). Computing and evaluating factor scores. Psychological Methods, 6(4), 430–450. doi:10.1037/1082-989x.6.4.430

Hambleton, R. K., & Rogers, H. J. (1989). Solving criterion-referenced measurement problems with item response models. *International Journal of Educational Research*, *13*(2), 145–160. doi:10.1016/0883-0355(89)90003-7

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. Psychometrica, 30(2), 179–185. doi:10.1007/bf02289447

Hooper, Daire & Coughlan, Joseph & Mullen, Michael. (2007). Structural Equation Modeling: Guidelines for Determining Model Fit. *The Electronic Journal of Business Research Methods*. 6(1). doi:10.21427/D7CF7R

Hopwood, C. J., & Donnellan, M. B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, *14*(3), 332–346. doi:10.1177/1088868310361240

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. doi:10.1080/10705519909540118

Hulin, C. L., Drasgov, F. & Parsons, C. (1983). *Item Response Theory Application to Psychological Measurement*. The Dorsey Professional Series.

Källmen, H., Wennberg, P., & Bergman, H. (2011). Psychometric properties and norm data of the Swedish version of the NEO-PI-R. *Nordic Journal of Psychiatry*, *65*(5), 311–314. https://doi.org/10.3109/08039488.2010.545433

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior*, *16*(3), 215–224. doi:10.1002/job.4030160304

Kører, P. K., Lang-Ree, O. C., Stubberud, K. V., & Martinussen, M. (2017). Predicting basic military performance for conscripts in the Norwegian Armed Forces. *Military Psychology, 29(6)*, 560-569. doi:doi.org/10.1037/mil0000192

Laverdière, O., Morin, A. J. S., & St-Hilaire, F. (2013). Factor structure and measurement invariance of a short measure of the Big Five personality traits. *Personality and Individual Differences*, *55*(7), 739–743. doi:10.1016/j.paid.2013.06.008

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1(2)*, 130–149. doi:10.1037/1082-989x.1.2.130

Marsh, H. W., Guo, J., Parker, P. D., Nagengast, B., Asparouhov, T., Muthén, B., & Dicke, T. (2018). What to do when scalar invariance fails: The extended alignment method for multi-group factor analysis comparison of latent means across many groups. *Psychological Methods*, *23*(3), 524–545. https://doi.org/10.1037/met0000113

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22(3)*, 471–491. doi:10.1037/a0019227

McCrae, Robert R., & Costa Jr, P. T. (1997). Conceptions and correlates of openness to experience. In *Handbook of personality psychology* (pp. 825–847). Elsevier.

McCrae, Robert R., & Terracciano, A. (2005). Universal features of personality traits from the observer's perspective: Data from 50 cultures. *Journal of Personality and Social Psychology*, *88*(3), 547-561. doi:10.1037/0022-3514.88.3.547

McCrae, R. R., & Costa, P. T. (2007). Brief versions of the NEO PI-3. *Journal of Individual Differences, 28(3)*, 116-128. doi:10.1027/1614-0001.28.3.116

McCrae, R. R., & Costa, P. T. (2010). *NEO inventories professional manual*. Psychological Assessment Resources.

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007a). Are we getting fooled again? Coming to terms with limitations in the use of personality tests for personnel selection. *Personnel Psychology, 60(4)*, 1029-1049. doi:10.1111/j.1744-6570.2007.00100.x

Morgeson, F. P., Campion, M. A., Dipboye, R. L., Hollenbeck, J. R., Murphy, K., & Schmitt, N. (2007b). Reconsidering the use of personality tests in personnel selection contexts. *Personnel Psychology, 60(4)*, 683-729. doi:doi.org/10.1111/j.1744-6570.2007.00089.x

Mullins-Sweatt, S. N., & Widiger, T. A. (2006). The Five-Factor Model of Personality Disorder: A Translation across Science and Practice. In R. F. Krueger & J. L. Tackett (Eds.), *Personality and psychopathology* (p. 39–70). Guilford Press.

Ones, D. S., Dilchert, S., Viswesvaran, C., & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60(4)*, 995-1027. doi:10.1111/j.1744-6570.2007.00099.x

Ployhart, R. E. (2012). Personnel Selection: Ensuring Sustainable Organizational Effectiveness Through the Acquisition of Human Capital In S. W. J. Kozlowski (Ed.),

The Oxford Handbook of Organizational Psychology, Volume 1. Oxford University Press.

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review, 41*, 71–90. doi:10.1016/j.dr.2016.06.004

Reise, S. P., Bonifay, W. E., and Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. J. Pers. Assess. 95, 129–140. doi: 10.1080/00223891.2012.725437

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology, 82(1)*, 30-43. doi:10.1037/0021-9010.82.1.30

Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civilian occupations: a European perspective. *Human Performance, 11(2)*, 271-288. doi:10.1207/s15327043hup1102&3_8

Savalei, V. (2012). The Relationship Between Root Mean Square Error of Approximation and Model Misspecification in Confirmatory Factor Analysis Models. *Educational and Psychological Measurement, 72(6)*, 910–932. doi:10.1177/0013164412452564

Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. Psychological Bulletin, 124(2), 262–274. doi:10.1037/0033-2909.124.2.262

Siebert, S. E., & DeGeest, D. S. (2015). The Five Factor Model in Business and Industry. In T. A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model*. Oxford University Press.

Skoglund, T. H. (2017). *The Norwegian Military Personality Inventory (NMPI): conscript assessment*. Paper presented at the The International Military Testing Association, Bern.

Statacorp. (2017). *StataCorp. 2019. Stata Statistical Software: Release 16. College Station, TX: StataCorp LLC*.

Van De Schoot, R., Schmidt, P., De Beuckelaer, A., Lek, K., & Zondervan-Zwijnenburg, M. (2015). Editorial: Measurement Invariance. *Frontiers in Psychology, 6*. doi:10.3389/fpsyg.2015.01064

Vandenberg, R. J., & Lance, C. E. (2000). A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods, 3(1)*, 4–70. doi:10.1177/109442810031002

Watson, P. J., Morris, R. J., & Hood, R. W. (1989). Interactional Factor Correlations with Means and End Religiousness. Journal for the Scientific Study of Religion, 28(3), 337. doi:10.2307/1386744

Widiger, T. A. (Ed.) (2015). *The Oxford Handbook of the Five Factor Model*. Oxford University Press.

Wilt, J., & Revelle, W. (2015). Affect, Behaviour, Cognition and Desire in the Big Five: An Analysis of Item Content and Structure. *European Journal of Personality, 29(4)*, 478-497. doi:10.1002/per.2002

Ziegler, M., & Buehner, M. (2009). Modeling socially desirable responding and its effects. *Educational and Psychological Measurement*, 69, 548–565. doi:10.1177/0013164408324469

Figure captions

Figure 1. Total test information curves over theta after the graded response models across five traits. Total information is the sum of information across all trait levels and response options for each item.

Figure 2. Histogram of percentage of un-adjusted observed scores on personality factors over gender as measured by the NMPI-50.

Figure 1
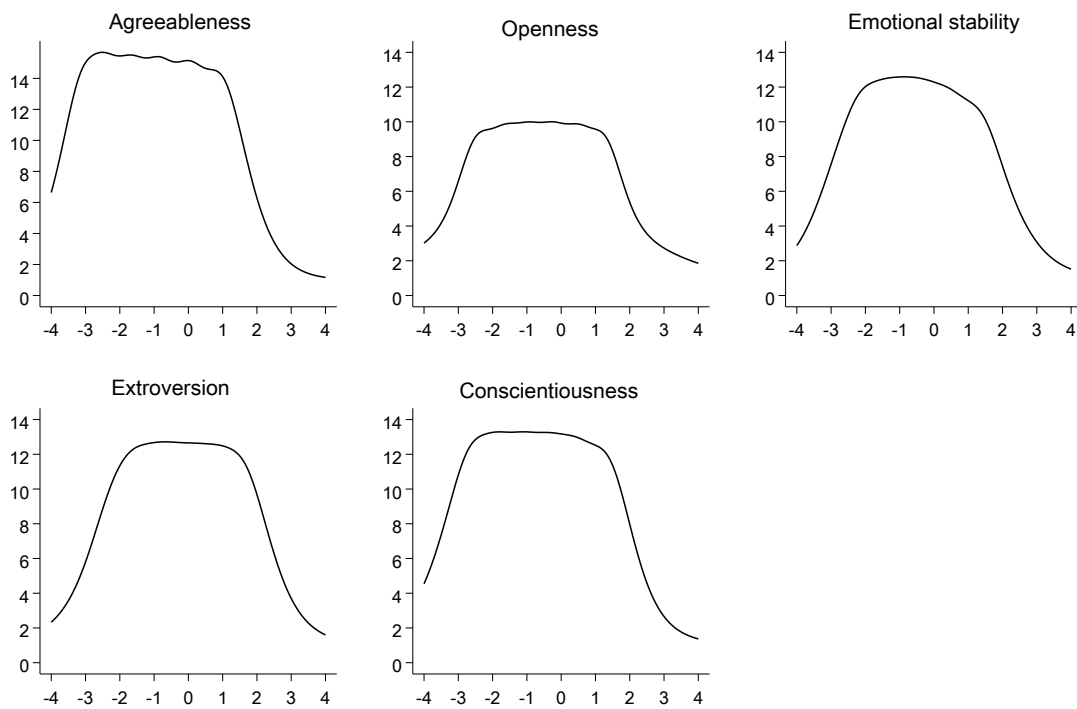
Figure 2
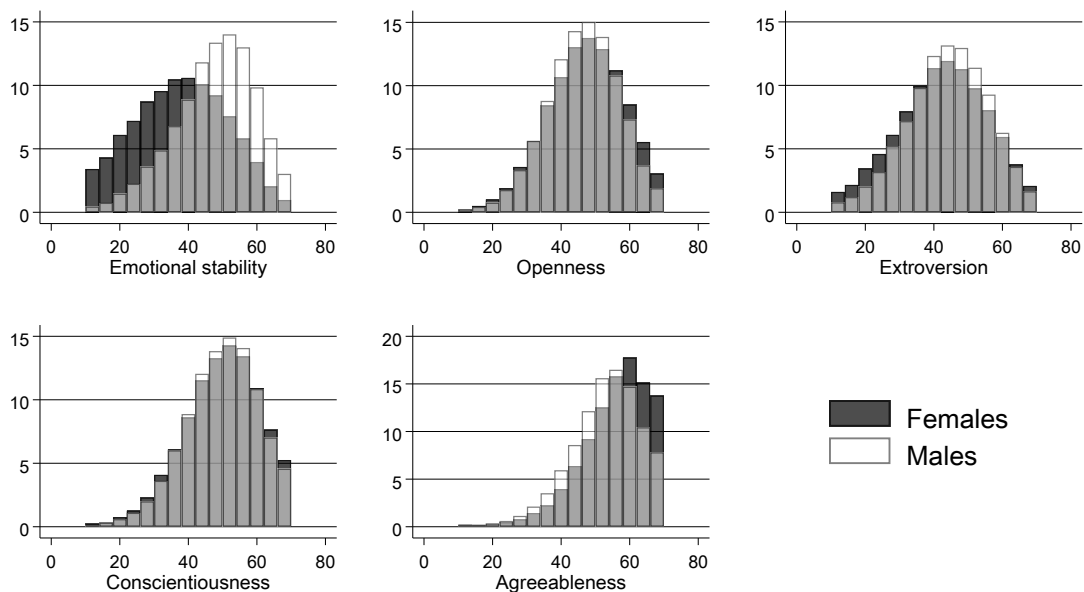
Tables

Table 1. Means, Standard Deviations (SD), Alpha Coefficients, Observed and Latent Trait Correlations. Correlations Above the Diagonal Are From Latent Traits After CFA. Correlations Below The Diagonal Are Observed Scores.

Variable	Mean (SD)	1	2	3	4	5	6
1.Emotional stability	42.21 (13.37)	**.91**	.52*	.23*	.31*	.44*	-
2. Extroversion	42.66 (12.31)	.47*	**.90**	.63*	.62*	.61*	-
3. Openness to experience	46.10 (10.74)	.11*	.42*	**.85**	.68*	.62*	-
4. Agreeableness	53.69 (10.12)	.18*	.48*	.56*	**.91**	.72*	-
5. Conscientiousness	49.07 (10.75)	.35*	.47*	.48*	.61*	**.90**	-
6. Motivation for military service	2.70 (1.52)	.46*	.41*	.30*	.38*	.44*	**-**

Note: Chronbach´s alfa in bold along the diagonal. * $p < .01$(two-tailed).

Table 2

Exploratory Factor Analysis Matrix After Promax Rotation with a Five-Factor Solution

Item	Brief description	Factor1	Factor2	Factor3	Factor4	Factor5
o1	Notices beauty	.05	.25	.05	-.03	**.42**
o2	Desire for self-development	**.31**	.30	.15	.02	.11
o3	Takes time to reflect	.10	.09	.03	-.31	**.40**
o4	Fascinated by shapes and colors	-.06	.08	-.13	-.06	**.65**
o5	Vivid imagination	-.12	.06	-.02	.11	**.74**
o6	Generates ideas easily	.14	-.01	.11	.12	**.61**
o7	Familiar with many concepts	**.31**	.03	.15	-.01	.25
o8	Appreciates art	-.16	.05	-.11	-.03	**.70**
o9	Thinks creatively	.06	.03	.00	.16	**.71**
o10	Curious about other cultures	-.01	.20	.04	.00	**.44**
e1	View myself as outgoing	-.01	.31	**.63**	.03	-.14
e2	Frequently becomes a leader	.29	-.12	**.64**	-.06	.07
e3	Likes being the center of attention	-.09	.14	**.76**	-.09	-.02
e4	Do not like attention (r)	-.23	.09	**.63**	.14	-.03
e5	Likes to make decisions	.30	-.13	**.62**	-.06	.12
e6	Starts conversations	-.00	.28	**.59**	.01	.04
e7	Keeps in the background (r)	-.15	.07	**.63**	.23	-.14
e8	Dominant and self-confident	.14	-.12	**.59**	.08	.10
e9	Is introverted (r)	-.19	.18	**.56**	.17	-.23
e10	Enjoys leadership	.28	-.16	**.66**	-.06	.15
n1	Easily stressed (r)	.10	.04	-.05	**.65**	.09
n2	Worries a lot (r)	.00	.07	-.05	**.76**	.01
n3	Fear embarrassment (r)	-.10	-.02	.19	**.61**	.11
n4	Thinks about other´s view of me (r)	-.09	-.10	-.03	**.68**	.05
n5	Feelings of guilt (r)	.06	-.14	-.04	**.64**	-.03
n6	Frequently sad (r)	.13	.11	-.03	**.64**	-.11
n7	Prone to nervousness (r)	-.00	-.03	.15	**.71**	.08
n8	Feelings of inferiority (r)	.10	-.15	.15	**.57**	.01
n9	Feel tense (r)	.02	.03	.00	**.69**	-.05
n10	Frequently scared (r)	.09	.08	.06	**.62**	.02
c1	True to appointments	**.61**	.20	-.12	.05	-.11
c2	Reaches goals	**.57**	.08	.17	.09	-.03
c3	Attention to details	**.58**	-.05	-.01	-.12	.18
c4	Punctual person	**.72**	.14	-.07	-.02	-.07
c5	Systemizes frequently	**.72**	.01	.00	-.04	.09
c6	Completes duties diligently	**.71**	.12	-.09	.08	-.09
c7	Generally prepared	**.73**	.15	-.05	.04	-.05
c8	Very industrious	**.70**	.12	-.03	.01	-.07
c9	Surrounded by clutter (r)	**.50**	.02	-.06	.23	-.15
c10	Sets goals and works to reach them	**.70**	-.01	.09	.02	.02
a1	Understands others' needs	.23	**.44**	.09	.05	.10
a2	Attention to the wellbeing of others	.13	**.63**	.02	.01	-.04
a3	Something nice to say about everyone	.11	**.61**	-.02	.02	.07
a4	Takes the time to listen to others	.14	**.66**	.08	-.01	.02
a5	Interested in other people	.02	**.58**	.16	-.03	.11
a6	Easily cares for others	-.06	**.64**	.20	-.15	.03
a7	Very helpful	.27	**.61**	-.04	.02	.07
a8	Compassionate	-.01	**.71**	-.04	-.13	.12
a9	Likes to help others	.11	**.69**	-.05	.01	.12
a10	Trusts others	.09	**.60**	-.05	.11	.01

Note: Boldface denotes the highest factor loading, *r* indicates reversed items.

Table 3. Bivariate Correlations Among The Five Personality Domains, The General Factor And Motivation For Military Service

Variable	1	2	3	4	5	6
1. Emotional stability	-					
2. Extroversion	.19*	-				
3. Openness to experience	-.07*	-.10*	-			
4. Agreeableness	-.08*	.04*	.17*	-		
5. Conscientiousness	.12*	-.11*	.13*	.27*	-	
6. Motivation for military service	.54*	.03*	.09*	.16*	.28*	-

* $p < .01$(two-tailed).

Table 4. Chi-squares, Degrees of Freedom, P-values and Fit Indices For Each Factor, Bi-factor, Five- and Three-Factor Solutions For The NMPI-50.

Factor	χ^2	df	p	RMSEA	SRMR	CFI	TLI
Five-factor solution	145170	1165	.00	.067	.087	.819	.809
Three-factor solution	269951	1172	.00	.092	.121	.661	.646
Bi-factor solution	69018	1115	.00	.058	.062	.875	.862
Extroversion	23727	35	.00	.158	.076	.834	.786
Openness	22444	35	.00	.153	.088	.788	.728
Neuroticism	8400	35	.00	.094	.040	.941	.924
Conscientiousness	7393	35	.00	.087	.038	.944	.922
Agreeableness	4005	35	.00	.065	.027	.975	.961

Note. NMPI-50, Norwegian military personality inventory 50.

Table 5. Discrimination (α) and Difficulty (β) Parameter Estimates for the NMPI-50 based on Five Graded Response Models

Item	α	β1	β2	β3	β4	β5	β6
o1	1.32	-3.71	-2.46	-1.50	-0.52	0.43	1.48
o2	1.17	-4.51	-3.60	-2.72	-1.75	-0.77	0.31
o3	0.98	-3.88	-2.58	-1.43	-0.38	0.73	1.99
o4	1.10	-2.27	-1.16	-0.28	0.63	1.60	2.64
o5	2.27	-2.50	-1.73	-1.09	-0.38	0.33	1.12
o6	2.80	-2.60	-1.87	-1.11	-0.34	0.42	1.20
o7	1.07	-3.75	-2.52	-1.29	-0.05	1.02	2.42
o8	0.96	-1.46	-0.32	0.53	1.35	2.14	3.07
o9	2.96	-2.48	-1.65	-0.93	-0.22	0.53	1.29
o10	1.08	-2.65	-1.54	-0.79	0.02	0.94	2.02
e1	1.84	-2.39	-1.70	-1.01	-0.19	0.55	1.39
e2	2.31	-1.83	-1.04	-0.37	0.30	1.00	1.82
e3	1.97	-1.90	-1.12	-0.40	0.37	1.12	1.86
e4	1.23	-2.72	-1.84	-1.03	-0.10	0.83	2.05
e5	2.34	-2.12	-1.46	-0.73	-0.01	0.75	1.54
e6	2.12	-2.36	-1.52	-0.78	0.02	0.81	1.70
e7	1.39	-2.84	-1.71	-0.78	0.11	0.99	2.39
e8	1.94	-2-24	-1.46	-0.67	0.24	1.18	2.17
e9	1.00	-3.57	-2.36	-1.37	-0.24	0.55	1.79
e10	2.44	-1.72	-0.98	-0.38	0.32	0.98	1.64
n1	1.85	-2.58	-1.75	-1.08	0.42	0.31	1.56
n2	2.31	-2.19	-1.44	-0.89	-0.34	0.28	1.31
n3	1.60	-1.99	-1.18	-0.49	0.15	0.85	1.88
n4	1.37	-2.34	-1.27	-0.43	0.35	1.19	2.36
n5	1.47	-2.53	-1.60	-0.82	-0.05	0.74	2.00
n6	1.89	-3.00	-2.35	-1.75	-1.13	-0.47	0.70
n7	2.54	-2.19	-1.45	-0.79	-0.23	0.45	1.51
n8	1.51	-2.38	-1.50	-0.68	0.10	0.95	2.12
n9	1.90	-2.57	-1.83	-1.08	-0.24	0.53	1.64
n10	1.92	-2.98	-2.30	-1.68	-1.06	-0.31	0.85
c1	1.66	-3.59	-2.65	-1.90	-1.09	-0.26	1.02
c2	1.95	-3.32	-2.60	-1.80	-0.85	0.13	1.36
c3	1.36	-3.28	-2.29	-1.30	-0.16	0.84	2.07
c4	2.06	-2.92	-2.18	-1.43	-0.55	0.27	1.22
c5	2.22	-2.51	-1.83	-1.08	-0.25	0.59	1.45
c6	2.25	-2.31	-1.38	-0.63	0.12	0.93	1.84
c7	2.65	-2.62	-1.79	-1.06	-0.30	0.50	1.44
c8	2.27	-2.80	-1.92	-1.13	-0.34	0.40	1.24
c9	1.16	-3.25	-2.17	-1.23	-0.28	0.72	2.23
c10	2.23	-2.69	-1.7<u>6</u>	-0.93	-0.06	0.70	1.57
a1	1.81	-3.46	-2.71	-1.72	-0.78	0.29	1.47
a2	1.94	-3.43	-2.75	-1.96	-1.10	-0.24	0.83
a3	1.88	-2.98	-2.24	-1.50	-0.67	0.23	1.22
a4	2.68	-3.03	-2.29	-1.<u>52</u>	-0.67	0.20	1.20
a5	2.05	-2.97	-2.26	-1.59	-0.72	0.20	1.27
a6	1.83	-3.22	-2.31	-1.45	-0.55	0.33	1.28
a7	2.89	-3.22	-2.47	-1.70	-0.86	0.03	0.96
a8	2.10	-3.19	-2.40	-1.62	-0.78	0.15	1.18
a9	2.88	-3.27	2.42	-1.69	-0.87	-0.05	0.79
a10	1.75	-3.35	-2.48	-1.67	-0.75	0.29	1.59

Note: NMPI: Norwegian Military Personality Inventory

Table 6. Male and Female Observed NMPI-50 Factor Means.

	Males	Females
Openness	45.75 (10.34)	46.45 (11.11)
Extroversion	43.47 (11.65)	41.84 (12.88)
Emotional stability	46.98 (11.54)	37.42 (13.37)
Conscientiousness	49.11 (10.50)	49.04 (11.01)
Agreeableness	52.26 (10.12)	55.13 (9.91)

Note: SD in parenthesis

Table 7: Goodness of Fit Statistics on Tests of Measurement Equivalence of Male and Female Metric and Scalar Invariance on the NMPI-50.

Personality trait	Model	χ^2	df	CFI	TLI	RMSEA	Invariant*
Emotional stability	Baseline	16708	70	.936	.918	.095	-
	Full metric	17393	79	.934	.924	.091	Yes
	Full scalar	19476	88	.926	.924	.091	Yes
Conscientiousness	Baseline	14775	70	.944	.928	.089	-
	Full metric	14883	79	.943	.935	.084	Yes
	Full scalar	17275	88	.934	.933	.086	Yes
Openness	Baseline	39516	70	.809	.754	.146	-
	Full metric	40452	79	.804	.777	.139	Yes
	Full scalar	52776	88	.745	.739	.151	No
	Partial scalar	46183	86	.777	.766	.143	No
Extroversion	Baseline	49488	70	.830	.782	.164	-
	Full metric	50076	79	.828	.804	.155	Yes
	Full scalar	52522	88	.820	.816	.150	Yes
Agreeableness	Baseline	7969	70	.972	.964	.068	-
	Full metric	8009	79	.972	.968	.062	Yes
	Full scalar	11678	88	.959	.959	.071	No
	Partial scalar	9799	87	.967	.966	.064	Yes

Note. NMPI-50, Norwegian military personality inventory 50. *CFI difference from baseline model =< - 0.01

Table 8. Summary of NMPI-50 Scale Results

Scale	Dimensionality	θ Information range*	Gender measurement invariance [a]	
			Assess relationships with other variables	Compare group scores
Openness	Not unidimensional	-2.9 to 1.7	Yes, full metric	No
Extroversion	Not unidimensional	-2.9 to 2.4	Yes, full metric	Yes
Emotional stability	Unidimensional	-3.1 to 2-3	Yes, full metric	Yes
Conscientiousness	Unidimensional	-3.6 to 2.1	Yes, full metric	Yes
Agreeableness	Unidimensional	-3.9 to 2.0	Yes, full metric	Yes, partial scalar

Note. [a]Based on Cheung & Rensvold (2002) CFI criteria. *SE < 0.5

Appendix A Items and Standardized Scores†

	Åpenhet/Openness		Emosjonell stabilitet/Emotional stability		Ekstroversion/Extroversion
o1	Legger merke til vakre ting *Notice beautiful objects*	n1	Blir lett stressa* *Am easily stressed*	e1	Ser på meg selv som en utadvendt person *View myself as an outgoing person*
o2	Har ønske om å utvikle meg *Wish to engage in self-development*	n2	Bekymrer meg mye* *Worry a lot*	e2	Blir ofte leder av grupper *Frequently becomes the leader of groups*
o3	Bruker tid på å reflektere over ting i livet *Spend time reflecting on life*	n3	Er ofte redd for å dumme meg ut* *Am often afraid of making a fool of myself*	e3	Liker å være sosialt midtpunkt *Like to be the center of social attention*
o4	Lar meg fascinere av mønstre og farger *Am fascinated by patterns and colors*	n4	Tenker ofte på hva andre mener om meg* *Often think about what others think of me*	e4	Liker ikke oppmerksomhet* *Do not like attention*
o5	Har en god fantasi *Have a vivid imagination*	n5	Har ofte skyldfølelse* *Often feel guilty*	e5	Bestemmer gjerne *Likes to decide*
o6	Får mange gode ideer *Generates many good ideas*	n6	Er ofte lei meg* *Am often sad*	e6	Starter ofte samtaler *Often start conversations*
o7	Kjenner mange ord og begreper *Know many words and concepts*	n7	Blir fort nervøs* *Become nervous easily*	e7	Holder meg ofte litt i bakgrunnen* *Often keep in the background*
o8	Synes kunst er viktig *Think art is important*	n8	Føler ofte at andre er bedre enn meg* *Often feel that others are better than me*	e8	Er dominerende og selvsikker *Am dominant and self-confident*
o9	Er flink til å tenke kreativt *Am good at creative thinking*	n9	Føler meg ofte anspent* *Often feel tense*	e9	Er innadvendt* *Am introverted*
o10	Er nysgjerrig på andre kulturer *Am curious about other cultures*	n10	Er redd for mye* *Am afraid of many things*	e10	Liker å lede andre *Like to lead others*

*Reversed †Non-commercial use only

Planmessighet/Conscientiousness		Omgjengelighet/Agreeableness		Alternativer/anchors	
c1	Holder alltid avtaler *Always keep appointments*	a1	Er flink til å forstå andres behov *Am good at understanding the needs of others*	1.	Helt uenig – *Strongly disagree*
c2	Når mine mål *Reach my goals*	a2	Oppfattes som varm og vennlig *Am perceived as warm and friendly*	2.	
c3	Er detaljorientert *Am detail oriented*	a3	Har noe godt å si om alle *Have something nice to say about everyone*	3.	
c4	Er en punktlig person *Am a punctual person*	a4	Tar meg tid til andre *Take time for others*	4.	
c5	Er flink til å sette ting i system *Am good at putting things in order*	a5	Er interessert i andre mennesker *Am interested in other people*	5.	
c6	Gjør mine plikter med en gang *Completes my duties right away*	a6	Blir lett glad i andre mennesker *Easily grow fond of other people*	6.	
c7	Stiller alltid forberedt *Always meet prepared*	a7	Er hjelpsom overfor andre *Am helpful to others*	7.	Helt enig – *Strongly agree*
c8	Er arbeidsom og liker å få tingene unna *Am hardworking and like to get things done*	a8	Har ofte medfølelse for andre *Often feel compassion for others*		
c9	Har mye rot rundt meg* *Have a lot of clutter around me*	a9	Liker å hjelpe andre *Like helping others*		
c10	Har klare mål og arbeider systematisk for å nå dem *Have clear goals and work systematically to achieve them*	a10	Har tillit til andre *Trust others*		

*Reversed †Non-commercial use only

Appendix A – *Continued. Standardized scores*

T-scores: Åpenhet/openness*

Sum score males	25.1	30.2	35.4	40.6	45.8	50.9	56.2	61.3	66.5
Sum score females	24.2	29.7	35.3	40.9	46.5	52.1	57.6	63.2	68.7
T-score	30	35	40	45	50	55	60	65	70

*Sum scores rounded to 1 decimal

Stanine scaled scores: Åpenhet/openness*

Sum score males	<=27	28-33	34-38	39-43	44-49	50-54	55-58	59-62	>=63
Sum score females	<=26	27-32	33-38	39-43	44-50	51-55	56-60	61-64	>=65
Stanine score*	1	2	3	4	5	6	7	8	9

*Sumscore 44 manually adjusted from stanine 4 to 5 in both groups to obtain better distribution

T-scores: Emosjonell stabilitet/emotional stability*

Sum score males	23.8	29.6	35.4	41.2	47.0	52.7	58.6	64.3**	
Sum score females	10.5	17.2	24.0	30.7	37.4	44.1	50.8	57.5	64.2
T-score	30	35	40	45	50	55	60	65	70

*Sum scores rounded to 1 decimal ** range restricted (10-70)

Stanine scaled scores: Emosjonell stabilitet/emotional stability

Sum score males	<=24	25-32	33-39	40-45	46-51	52-56	57-60	61-64	>=65
Sum score females	<=14	15-20	21-26	27-34	35-41	42-48	49-55	56-59	>=60
Stanine score*	1	2	3	4	5	6	7	8	9

T-scores Ekstroversjon/Extroversion*

Sum score males	20.2	26.0	31.8	37.7	43.5	49.2	55.2	61.1	66.9
Sum score females	16.1	22.5	28.9	35.4	41.9	48.3	54.8	61.2	67.7
T-score	30	35	40	45	50	55	60	65	70

*Sum scores rounded to 1 decimal

Stanine scaled scores: Ekstroversjon/Extroversion

Sum score males	<=21	22-29	30-35	36-41	42-47	48-53	54-58	59-62	>=63
Sum score females	<=18	19-25	26-32	33-39	40-46	47-52	53-58	59-62	>=63
Stanine score*	1	2	3	4	5	6	7	8	9

T-scores: Planmessighet/Conscientiousness*

Sum score males	28.1	33.2	38.5	43.8	49.1	54.3	59.6	64.8**
Sum score females	27.0	32.5	38.0	43.5	49.0	54.5	60.1	65.5**
T-score	30	35	40	45	50	55	60	65

*Sum scores rounded to 1 decimal. ** range restricted (10-70)

Stanine scaled scores: Planmessighet/Conscientiousness

Sum score males	<=29	30-36	37-42	43-47	48-53	54-57	58-62	63-66	>=67
Sum score females	<=28	29-35	36-41	42-47	48-53	54-58	59-62	63-66	>=67
Stanine score*	1	2	3	4	5	6	7	8	9

T-scores: Omgjengelighet/Agreeableness*

Sum score males	32.0	37.1	42.1	47.2	52.2	57.3	62.4	67.4**
Sum score females	35.2	40.2	45.2	50.1	55.1	60.0	65.1	70.0**
T-score	30	35	40	45	50	55	60	65

*Sum scores rounded to 1 decimal. ** range restricted (10-70)

Stanine scaled scores: Omgjengelighet/Agreeableness

Sum score males	<=32	33-40	40-46	47-51	52-56	57-60	61-64	65-67	>=68
Sum score females	<=35	36-43	44-49	50-54	55-59	60-63	64-66	67-68	>=69
Stanine score*	1	2	3	4	5	6	7	8	9