Faculty of Health Sciences

A short-form personality measure for military personnel selection

Psychometric investigation and perspectives on usage

Tom Hilding Skoglund

A dissertation for the degree of philosophia doctor (PhD) January 2022

Table of Contents

Acknowledgements

The most important contributors to this thesis are the study participants, to which I am deeply thankful. I wish you all the best in life, whether you still serve in the Armed Forces or have followed your interests elsewhere.

To my two PhD supervisors: Monica Martinussen and Ole Boe. I have received excellent support and guidance from you. Thank you both for prioritizing my project in a professor´s schedule. I have particularly appreciated your availability and quick responses, whether by phone calls, e-mails, or meetings.

I also wish to express gratitude to my former employers, Randi Bjøntegaard and Cia Dunstrøm, at Oslo New University College. Thank you for financing my PhD student start-up while lecturing at your campus. Your flexible leadership style gave me the necessary autonomy for embarking on this project.

Several members of the Norwegian Armed Forces helped me in my doctoral work, whether it be through discussions, data collections, or co-authorship on the studies. I especially thank Morten Nordmo for sharing his knowledge and Ole Christian Lang-Ree for acknowledging and supporting my academic ambitions. I am also grateful for constructive advice on my PhD work from a partner in crime through many years, Patrick Risan, at the Norwegian Police University College.

To my current colleagues and supervisors in the Armed Forces: Thank you for appreciating my academic activities while working for our national security. Chance favors only the prepared mind.

To my family: I have appreciated your kindness and general support through this process, especially from my twin sister (herself an academic). Lastly, I could not have completed this thesis if it weren't for the support from you, dear Annikken. Thank you for being a loving mother to our son and an ally in the journey of life.

Abbreviations

A	Agreeableness
ABLE	Assessment of Background and Life Experiences
ANOVA	Analysis of Variance
BARS	Behaviorally Anchored Rating Scales
C	Conscientiousness
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CTT	Classical Test Theory
E	Extroversion
EFA	Exploratory Factor Analysis
ES	Emotional Stability
FFM	Five-factor model
FSK	SOF department: Forsvarets spesialkommando
GMA	General Mental Ability
IRT	Item Response Theory
MJK	SOF department: Marinejegerkommandoen
N	Neuroticism
NAF	Norwegian Armed Forces
NEO-PI-3	NEO Personality Inventory-3
NMPI	Norwegian Military Personality Inventory
O	Openness to experience
OAR	Overall Assessment Rating
RMSEA	Root Mean Square Error of Approximation
SRMR	Standardized Root Mean Square Residual
SOF	Special Operations Forces
TLI	Tucker-Lewis Index

List of studies

Study I

Skoglund, T. H., Fosse, T. H., Lang-Ree, O. C., Martinsen, Ø. L., & Martinussen, M. (2021). Candidate personality traits associated with ratings in a military officer selection setting. *Scandinavian Journal of Work and Organizational Psychology, 6,* 1-12. http://doi.org/10.16993/sjwop.135

Study II

Skoglund, T. H., Brekke, T-H., Steder, F. B., & Boe, O. (2020). Big Five personality profiles in the Norwegian Special Operations Forces. *Frontiers in Psychology*, 11:747. https://doi.org/10.3389/fpsyg.2020.00747

Study III

Nordmo, M., Skoglund, T. H., Lang-Ree, O. C., Austad, S. K., & Martinussen, M. (2021). The psychometric properties and norm data of the Norwegian military personality inventory (NMPI). *Scandinavian Journal of Psychology, 62*, 596-607. https://doi.org/10.1111/sjop.12719

All studies are reprinted here. Studies I and II are open access publications. Study III is reprinted with permission from the Scandinavian Journal of Psychology / John Wiley and Sons.

Abstract

Immediately after World War II, a non-clinical psychological testing and assessment unit was developed in the Norwegian Armed Forces. Its mission was initially to test conscripts for general mental ability, inspired by soldier placement and selection procedures among allied nations. Later on, this unit engaged in developmental projects related to personality measures. These efforts aligned eventually with the trait approach and the five-factor model (FFM) of personality. In the early 1990s, a Norwegian FFM measure was developed by psychologist and personality researcher Harald Engvik, used operationally to this day.

The Norwegian FFM measure has been used in several military selection processes and has often functioned as an ingredient in team and leadership development courses. However, the language of items, norms, and software solutions has become dated. Additionally, the measure is too lengthy for large-scale selection programs such as the conscript assessment procedure. Therefore, the psychological testing and assessment unit in the Armed Forces recently developed a new short-form measure: the Norwegian Military Personality Inventory (NMPI). This thesis aims to evaluate the psychometric properties of the NMPI in an applied military setting. Three published studies contribute to this end, based on data from three different military arenas: a selection program for officers, the special forces, and the conscript assessment procedure.

Study 1 explores the associations between officer candidate personality traits and mission command competency ratings at a joint selection program for the Army, Navy, and Air Force. The ratings were obtained through interviews ($N = 810$) and a field selection exercise simulating a warlike scenario ($N = 551$). This study preliminarily validated the NMPI-80, the first version of the NMPI, by correlating it with the well-established NEO-PI-3. The results show that extroversion (+) and openness to experience (-) emerged as predictors for both interview and field exercise ratings when controlling for age, sex, and general mental ability in a regression model. Strong correlations between the personality measures support the validity of the NMPI-80.

Study 2 uses the NMPI-80 to investigate personality traits among Special Operations Forces (SOF) personnel ($N = 190$). The analyses of variances showed no differences in mean scores among the different departments within the SOF: the Special Operations Commando (FSK), the Naval Special Operations Commando (MJK), and non-operator personnel (Support). When compared to Study 1 male participants, the operators ($N = 76$, all males) had

significantly lower mean scores on extroversion and agreeableness and a higher mean score on emotional stability.

Study 3 reports the psychometric properties of the NMPI-50, the finalized version of the NMPI, by analyzing responses from an age cohort undergoing the conscript assessment procedure ($N = 54,355$). Factor analyses, graded response models (based on item response theory), and tests of measurement invariance based on sex were used for this purpose. The NMPI-50 scores were evaluated as reliable and valid. However, some measurement challenges were identified, primarily concerning the openness to experience scale and a general risk for socially desirable responses.

The discussion of the results concerns the reliability and validity aspects of the versions of the NMPI and perspectives on using FFM measures in military personnel selections. The study findings inform practical usage when FFM measures are used "clinically" as background material for evaluating candidates undergoing rater-based selection methods. However, the NMPI-50, a short-form measure, is considered suboptimal when rich individual profiles are warranted. In the large-scale conscript assessment procedure, the Armed Forces are advised to evaluate if personality variables from the NMPI-50 can be used in an algorithm "mechanically" combining predictors.

This thesis expands the discipline of military psychology by providing empirical studies that are scarcely obtainable in the published literature (Studies 1 and 2). It also presents comprehensive psychometric analyses that reveal the strengths and weaknesses of self-report personality measures (Study 3). The conclusion of this thesis underlines the importance of a continued scientific approach to specifying and measuring human factors relevant for military activities. The FFM represents a well-validated and pragmatic framework that is eligible for such an approach, although challenges in reporting self-perceived personality are recognized. For strengthening selection practices around non-cognitive factors in the Norwegian Armed Forces, several areas of further NMPI research are identified.

1 Introduction

1.1 Thesis background

People engage naturally in ongoing assessments of each other. Perceptions of individuality and decision making about others permeate the human landscape. Whom to avoid and whom to befriend? Which one is the most suitable for the task at hand? In this way, we are all informal evaluators. Psychological testing represents a more formal evaluation process that aims to improve the accuracy of such assessments. If accurate, the results from psychological testing can provide an elaborate basis for decision making by supplementing and, if necessary, correcting an informal evaluation process.

In the testing realm of individual differences, the measurement of intelligence on one side and personality on the other are considered two basic disciplines (Furnham, 2008). The history of such testing in applied organizational settings lies within the military world, dating back to World War I, when Robert Yerkes and Robert Woodworth respectively developed intelligence tests and personality measures for classifying U.S. recruits (Driskell & Olmstead, 1989; Schultz & Schultz, 1996).

Military organizations became a natural harbor for the further development of intelligence and personality testing (Campbell & Knapp, 2001). The merging of psychological methods and the military is perhaps not surprising when considering that the Armed Forces need to select, educate, and train individuals to participate in military activities. At times, military activities place high psychological demands on service members. Once selected in an armed profession, education and training begin in a context where one must adjust to a military lifestyle – often involving strong discipline, continuous performance measures, and a social environment surrounded by peers. When fully trained and combat-ready, soldiers, specialists, and officers are expected to maintain their capabilities and carry out their duties as ordered. Ultimately, military occupational activity can involve conflicts, crises, and wars. The human demands in such contexts can be extraordinary, including feelings of ambiguity, perceptions of unforeseen consequences, and the taking of enemy lives. Arguably, adequate intelligence levels and adaptive personality functioning amongst military personnel can increase the probability of managing education, training, and the execution of operations. Insufficient management can lead to feelings of incompetency and frustration from the suboptimal service member. At worst, failure in performance increases the risk of accidents and the loss of lives.

To increase the match of human factor demands and military activities, much attention in military psychology has been devoted to personnel assessment and selection (Bowles & Bartone, 2017; Laurence & Matthews, 2012). Two main approaches for evaluating psychological suitability for service are select-in or select-out (King, 2014). A select-out procedure is usually concerned with psychopathological conditions; as such, it is a medical/clinical psychological procedure. For example, it is natural to conclude that an individual with ongoing severe depression or a personality disorder is not fit for participating in military activities. The former may include inadequate cognitive functioning due to symptom production and continuous cooperation problems for the latter. In contrast, the select-in procedure involves the assessment of normal psychological constructs for evaluating if applicants will manage and perform in upcoming service. Intelligence and personality tests are usually used in military select-in procedures as they help assess psychological constructs that are considered important for performance. However, such tests of non-pathological characteristics may not be used solely for personnel selection. Diverse human resources activities may be informed by test results, such as placement decisions, talent screening, and leadership development.

When serving as a military organizational psychologist, I participated in developing a new short-form personality inventory for use in the Norwegian Armed Forces (NAF). Whereas several intelligence tests exist in the NAF (Torjussen & Hansen, 1999) – for example, those connected to the conscript assessment procedure (Køber et al., 2017; Skoglund et al., 2014) and specialized selections such as pilot training (Martinussen & Torjussen, 1998) – we had only the 5PFmil 2.0 as a personality measure, a self-report inventory developed in the early 1990s. While this inventory was designed to capture the established five-factor model (FFM) of personality, thus holding a professionally sound anchor, the wording of some items is dated. Furthermore, the norms are somewhat nonrepresentative, as they rely on answers from United Nations veterans of the Lebanon and Bosnian conflicts and not on a general military population. Lastly, the test is very long and thus time-consuming to administer, making it impractical for military selection programs where many candidates might participate.

These challenges regarding the 5PFmil 2.0 spawned the in-house development of the Norwegian Military Personality Inventory (NMPI). The NMPI project developed a short-form personality measure based on self-reporting using a Likert scale, aiming to measure the five fundamental personality traits as defined by the FFM. Advantages in terms of copyright,

language, length, and, ultimately, adequate norms could therefore be achieved for the Armed Forces.

Research opportunities arose as the NMPI project progressed. The current thesis summarizes my PhD research activities related to this project and is based on three published studies. An initial measure, the NMPI-80, was used in Studies 1 and 2. The items were extracted from the international personality item pool (Goldberg et al., 2006) and discussed and edited in working groups comprised of military psychologists. The NMPI-80 is presented in Appendix A (although the measure contains 79 items, it was named NMPI-80 for convenience). In 2019, the Armed Forces HR and Conscription Centre highlighted a need to assess conscript personality, in addition to the general mental ability tests already in use. The test development project had to shorten the NMPI-80 and thoroughly investigate psychometric properties to meet this need. This effort resulted in Study 3, which examined the NMPI-50, the finalized measure containing 10 items for each FFM scale. The NMPI-50 includes items used in the NMPI-80 in addition to four new ones; it is presented in Appendix B.

By presenting and discussing the NMPI studies, this thesis contributes to the literature concerning the FFM in the Norwegian military context. Three different research strategies using three different samples contribute to this end. In Study 1, the NMPI-80 was correlated with the NEO-PI-3 and used to investigate associations between the FFM traits and officer selection ratings. The NMPI-80 was then used in Study 2 to explore the personality characteristics of special forces personnel, some of the most thoroughly selected and trained in the military. Lastly, Study 3 documents the psychometric investigation of the NMPI-50 based on data from the conscript assessment procedure.

In the following sections, I first present the FFM to define and clarify the personality phenomena with which this thesis is concerned. I share perspectives on how the FFM is related to the military realm. The thesis then provides the research findings concerning officer candidate ratings (Study 1) and special forces personnel (Study 2). A paragraph concerning approaches for investigating psychometrics (primarily relevant for Studies 1 and 3) follows these reviews. Then, I provide a detailed presentation of the study aims, methods, and results. An overall discussion focusing on NMPI measurement quality and usage potential in military selections and a summarizing conclusion concludes this thesis.

1.2 The five-factor model of personality

Among the phenomena that tend to individualize us is personality. Extensive research findings and discussions concerning what personality is and how best to measure it bear witness to the significant position that individual differences occupy in the science of psychology (John & Robins, 2021). For simplicity, Furnham (2008) argues that the main approaches for understanding human personality can be summarized as neo-psychoanalytic, psychophysical, and psychometric – which in turn relates to types, temperaments, and traits. This thesis, using the FFM, is concerned with the latter perspective, understanding personality traits as "relatively enduring patterns of thoughts, ideas, emotions, and behaviors that are generally consistent over situations and time and that distinguish individuals from each other" (Barrick & Mount, 2012, p. 226). The trait approach dates back to ancient times, with its modern scientific birth in the work of Gordon Allport (1931). He wrote that traits, among other things, must be seen as more generalized than a habit and not the same as moral qualities, thus presenting an initial conceptualization of what was believed to be separate psychological entities.

The numbering of traits and their construct understanding has reached considerable consensus with the formulation of the FFM or Big Five taxonomy of personality (Digman, 1990; Widiger, 2015). Although several scholars argue that the FFM and the Big Five taxonomy are essentially equivalent with respect to the concepts used for personality descriptions, at least at the overarching dimension levels (John & Robins, 2021), Simms et al. (2016) remind us that the former belongs to the questionnaire tradition and the latter to the lexical tradition. The latter is older, having its origin in the classical dictionary study by Allport and Odbert (1936), and assumes that personality variations observable in mankind are coded in language. The comprehensive trait approach work by Raymond Cattell is representative of this tradition, although he advocated for 16 factors (e.g., Cattell & Mead, 2010). Several other research efforts factor analyzing responses to personality-descriptive language yielded the five dimensions known as the Big Five (Goldberg, 1992, 1993).

The questionnaire tradition stems from a different process whereby self-report measures and behavior observations have contributed further to categorizing human personality. Such an approach is represented in the seminal work by Hans Eysenck on his personality model, where he argued for including three overarching traits (e.g., Eysenck, 1992). However, the most widespread example of the questionnaire tradition related to the FFM is arguably the NEO inventory system. Here, the NEO-PI-3 (the updated version of the

NEO-PI-R) is considered to be the most recognizable measurement tool (McCrae & Costa, 2010; McCrae et al., 2005).

Although the FFM (or the Big Five) grew into a well-established framework in the 1990s, nuanced criticisms related to methodology and interpretation of research findings were raised in the same period (e.g., Block, 1995). Newer criticism relates to the lack of a deeper theoretical understanding of the origin of traits, including presumed underlying genetic and biological mechanisms (Weiss & Deary, 2020). Some, including Cattell and Eysenck themselves, have also questioned whether five is the correct number for categorizing traits. A more recent example of such questioning is research advocating for the benefits of the HEXACO model, which includes a sixth personality domain (Ashton & Lee, 2020). Another is Hogan (e.g., 2005), who developed seven personality-related scales particularly relevant for the work sphere. Despite these criticisms and alternatives, the FFM is seen by many scholars and practitioners as a valid and fruitful system for describing and investigating human personality (McCrae, 2020; Widiger, 2015).

From a research perspective, the consensus around – and the nature of – the FFM as a widespread taxonomy has several advantages. For example, it has enabled studies across cultures and nations involving a large number of researchers, where results indicate that the five traits are generally universal to all human groups (McCrae et al., 2005). Another significant research advantage concerns the ability to aggregate study findings in the form of meta-analyses. Aggregations of studies have provided extensive knowledge in areas such as the linkage of personality and job performance (Barrick et al., 2001), personality dispositions predictive of occupational attainment (Roberts et al., 2007), and the relationship between personality and health (Strickhouser et al., 2017).

The trait approach advocates that one can be described according to where one scores on a continuum (contrasting typologies, which describe individuals in distinct personality categories). Neuroticism (N), the first trait described here, concerns a general tendency to experience negative affect. In some measures, such as the NMPI versions, N is reversed and named Emotional Stability (ES). According to the NEO-PI-3, the facets underlying N are anxiety, anger, hostility, depression, self-consciousness, impulsiveness, and vulnerability. Although associations toward psychopathological conditions might occur, McCrae and Costa (2010) remind us that N measures a dimension of general personality. They write that "high scorers may be at risk for some kinds of psychiatric problems, but the N scale should not be viewed as a measure of psychopathology" (p. 19). In scrutinizing the literature pertaining to N and its associated outcomes, Tackett and Lahey (2016) underline more strongly that N is

associated with psychopathology, as well as with physical health concerns. Therefore, this personality trait has significant consequences for public health. Accordingly, these authors note that N is associated with quality of life. Low scorers are generally interpreted as being emotionally stable and thus as inhabitants of less negative affect than high scorers.

Extroversion (E) captures, on one side, a tendency toward sociability and liking other people, and a tendency toward assertiveness and talkativeness on the other. The NEO facets are warmth, gregariousness, assertiveness, activity, excitement seeking, and positive emotions. E is a personality tendency included in several models across theoretical orientations since the first introduction of the term by Carl Jung (Wilt & Revelle, 2016). McCrae and Costa (2010) note that the understanding of this trait differs between the NEO system and Jungian psychology, for instance, with regard to the introspection tendency, which is a sign of openness to experience in the former system and low E in the latter. Considering the FFM, low scorers of E are interpreted to have tendencies toward introversion.

Openness to experience (O) is a tendency to experience life as rich, varied, and novel in content, both from the inner and outer worlds. Fantasy, aesthetics, feelings, actions, ideas, and values are the NEO facets. O is associated with creativity and intelligence (McCrae & Costa, 2010). The nature of this personality trait has been the target of more scientific disagreement compared to the other four traits, although it has been widely accepted as part of the FFM (Sutin, 2015). Low scorers on O are interpreted to prefer the familiar and to be more emotionally muted relative to high scorers.

Agreeableness (A) pertains to the "motivation to maintain positive relations with others" (Graziano & Tobin, 2016, p. 1). A is thus a trait directly related to interpersonal tendencies (as is E), where the facets of trust, straightforwardness, altruism, compliance, modesty, and tender-mindedness define the NEO domain. A central theme of discussion for this trait is the domain placement of the facet of warmth, which might theoretically have its natural place within A. In the NEO system, this facet belongs to E, as McCrae and Costa (2010) aimed to weigh E with an affective, interpersonal quality. Low scorers on A are interpreted to be skeptical, competitive, and prone to express anger easily.

Conscientiousness (C), the last of the five traits, is about having self-control, planning, organizing, and enduring. Competence, order, dutifulness, achievement striving, self-discipline, and deliberation are the NEO facets. C is associated with outcomes in life generally seen as positive, such as relationship stability and educational and occupational achievement (Jackson & Roberts, 2015). McCrae and Costa (2010) note that "low scorers are

not necessarily lacking in moral principles, but they are less exacting in applying them, just as they are more lackadaisical in working toward their goals" (p. 21).

In addition to the historical studies and definitions, a growing literature aiming to unravel the genetics of human personality has further contributed to the FFM's construct validity and the perspective that these traits are stable entities (e.g., Sanchez-Roige et al., 2017). Nevertheless, questions of trait stability across time and situations have been debated. One common view is that changes in traits may predominate in the 20–40 age span, whereby changes in ES, A, and C might occur (Atherton et al., 2020; Roberts & Mroczek, 2008). Trait-level changes may be due to a complex array of processes (i.e., biological mechanisms to life experiences). Overall, the formulation that traits show general stability, as emphasized in the definition of Barrick and Mount (2012), may secure a reasonable approach for understanding the complexity of personality.

The person-situation debate, which dates to a much-cited book by Walter Mischel (1968), is also relevant to the question of stability. He presented a critical viewpoint toward personality situational stability, thus fueling the expanding social psychological skepticism toward the trait approach during those years. Later, however, Mischel (2004) advocated for an interactionist approach (i.e., both person and situation) to understanding behavior. The interactionist approach is arguably the most widespread viewpoint of today's personality (and social) psychology. For example, in personnel psychology, theories of the interplay between situational characteristics and trait activation have been developed to understand nuances in linkages between personality and job performance (Judge & Zapata, 2015; Tett & Burnett, 2003).

1.3 The five-factor model in the military

As the challenges and demands of military personnel are sometimes of an extraordinary nature, psychology has had a long history in military organizations (Driskell & Olmstead, 1989; Maheshwari et al., 2016). Military psychology developed into several branches in the years after Robert Yerkes introduced intelligence testing during World War I, such as clinical services provided by non-psychiatrists (Page, 1996), military leadership theory (Johansen et al., 2019), and aspects of intelligence operations (Staal & Stephenson, 2013). Personality psychology has been relevant in all these branches by providing concepts and nuance for describing how people function. However, the most common military psychology activity involving the FFM is personnel assessment and selection (Bowles & Bartone, 2017; Laurence & Matthews, 2012).

There might be several reasons for implementing the FFM in military selection. First, personal characteristics are often included in military job analyses (Mitchell & Driskill, 1996; Sümer et al., 2001), as in civilian ones (Ployhart, 2012), so personality traits are clearly relevant to many selection decisions. Second, the military must often process a large number of individuals for selection decisions due to a high number of applicants and a high need for personnel. Cost-effective methods, such as psychometric tests based on self-reporting, are thus suitable for large-scale selections. A third reason might be that, as the FFM is thoroughly researched and easily communicated to military commanders and non-psychologists, military psychologists worldwide can effortlessly suggest using tests related to this economical, well-validated personality taxonomy.

1.3.1 Historical perspective

A brief historical review shows an early focus on non-cognitive measures in addition to general mental ability (GMA) in military personnel selection practices. Prior to the arrival of the trait approach and the FFM, non-cognitive measures were first and foremost related to psychiatric screening procedures. For example, Schultz and Schultz (1996) note that the Personal Data Sheet, a self-report inventory for measuring neurotic symptoms among recruits, was developed by Robert Woodworth, also during World War I. Another example pertains to the selection of both European and American military pilots in the interwar years, in which "emotional stability," in addition to GMA and psychomotor abilities, were deemed important (Koonce, 1984; Martinussen, 2005). Following these early efforts, historical sources describe military psychology as expanding markedly after World War II, and more elaborate developments concerning psychological testing were witnessed (Crawford, 1970; Melton, 1957). For instance, a notable development after World War II was a comprehensive project designing measures of characteristics such as "motivational structure," "character integration and emotional stability," and "tolerance for frustration and anxiety" for U.S. combat pilot selection (Sells, 1955).

The initial growth of the FFM in military psychology can be traced to a technical report from the early 1960s by the two U.S. Air Force psychologists Tupes and Christal (1992). They showed that five personality factors were replicated in studies across different military samples. This discovery was an important contribution to general FFM research. However, it was not until the comprehensive research and development effort in the U.S. Army in the 1980s and early 1990s – known as Project A (Campbell, 1990; Campbell & Knapp, 2001) – that the trait perspective and the FFM were fully actualized in military

15

selection procedures. From Project A, the Assessment of Background and Life Experiences (ABLE) was developed, of which several scales had a clear relationship to the FFM (White et al., 2001). Increased research interest in non-cognitive measures and different response systems, such as forced-choice and adaptive testing, followed in the U.S. (e.g., Nye et al., 2020; Stark et al., 2014).

In Norway, an in-house psychological testing and assessment unit was established right after the Nazi occupation (Eid et al., 2012). While the emphasis was on developing cognitive tests, especially for conscript placement and military pilot selection, activities related to personality and interest inventories also took place (Psykologen, 1964; Torjussen & Hansen, 1999). For instance, from the 1960s, the unit developed Norwegian norms for the Minnesota Multiphasic Personality Inventory (MMPI) (Ubostad & Ellertsen, 1999) and had other developmental projects related to measures of "projective anxiety" and "divergent thinking" (Hansen, 2006). Later, the Defence Mechanism Test (DMT) was implemented in pilot selection (Martinussen & Torjussen, 1998), which is believed to measure "precognitive defense organization" (Kragh, 1960). A research project concerning Rorschach testing for selecting naval special forces operators was also performed in the following years (Hartmann & Grønnerød, 2009; Hartmann et al., 2003).

In the late 1980s and early 1990s, the interest in military non-cognitive testing aligned with the emergence of the FFM and culminated in the development of the 5PFmil 2.0 by psychologist and researcher Harald Engvik (Engvik, personal communication, July 30, 2021). This test has since been used as an aid in selection decisions. Today, it is used primarily for selecting officers, pilots, and applicants for different special services.

1.3.2 The five-factor model and military job performance

Studies documenting the predictive validity of personality traits toward military occupational performance are significant for advocating the relevance of the FFM in the military. A seminal study by Hough et al. (1990) analyzed how ABLE scores related to different military performance criteria for around 8,000 U.S. enlisted personnel in 1985. This study documented that several observed correlations were above $r = .20$. Most notable were associations between "energy level" (related to E in the NEO system) and "self-esteem" (related to N) on one side and "effort and leadership" and "physical fitness and military bearing" on the other. C (as it was also named in the ABLE) showed an association with "personal discipline" and "physical fitness and military bearing." Some years later, Salgado (1998) performed a meta-analysis of the relationship between different FFM tests and job performance for European

civilian and military personnel. From a military perspective, the most notable finding was that ES (i.e., reversed N) showed a high uncorrected mean correlation (.30, based on eight studies). However, generalizability is limited, as the military data was comprised of army pilot samples where training proficiency was the criterion. Darr (2011) compiled 11 to 13 studies (depending on which FFM trait) correlating the Self-Descriptive Inventory with supervisor ratings and self-reported performance evaluations. When correcting for predictor and criterion unreliability, the most notable mean correlations were .20 for C, .18 for E, and -.17 for N.

A common perception regarding the linkage between the FFM and civil job performance is that C and N are valid predictors across jobs, while the effect of E, A, and O is dependent on job type (e.g., Furnham, 2008). This perception is based on extensive research findings, culminating in the overarching analysis of 15 meta-analyses performed by Barrick et al. (2001). Darr (2011) concluded that military results mostly align with the civilian. However, upon closer inspection, E may be more related to performance across military jobs relative to civilian ones. The E pattern might be explained through the consideration of trait activation theory, which suggests that situational cues increase the relevancy of a given personality trait (Tett & Burnett, 2003). Arguably, military arenas typically hold several such cues (e.g., working and often living with peers – at least for younger personnel).

The predictive validity findings generally support the use of FFM tests in selection procedures. A related but more detailed argument pertains to incremental validity. Such validity is demonstrated if personality traits increase prediction of job performance in a statistically significant way after controlling for intelligence/GMA due to the latter's established high predictive validity (Schmidt & Hunter, 1998). Regression analyses may be used in such research, where scores from different tests (or selection methods) can be treated as independent variables, while a performance measure serves as the dependent variable. Thus, the unique statistical contribution of each test, while controlling for the other tests, can be documented. For example, Schmidt and Hunter (1998) documented that C provided an 18% increase in validity above that provided by GMA tests. In Salgado's (1998) study, ES provided a 38% increase for military samples. The incremental validity of the ABLE scales was also seen early on in Project A, especially for the criterion of "effort and leadership" (McHenry et al., 1990).

Another argument favoring the use of personality testing relates to expanding the understanding of job performance. Since predictive and incremental validity studies have often used an overall and generalized assessment of (task) job performance as the criterion

(e.g., Schmidt & Hunter, 1998), they may miss the true complexity of occupational performance. The criterion problem, as such a viewpoint has been called, may produce weak correlations between personality tests and performance due to suboptimal measurement of the latter. In their review article on personnel selection, Sackett and Lievens (2008) noted that the criterion constructs used in Project A in the 1980s (personal discipline, effort and leadership, and physical fitness and military bearing) contributed to a more nuanced perspective on job performance. These authors summarized that job performance could be comprised of three parts: counterproductive work behavior, organizational citizenship behavior, and task performance. Gonzalez-Mulé et al. (2014) provided empirical support through a meta-analysis for how the FFM traits predicted counterproductive work behavior better than GMA and showed that the two predictors were equal in predicting organizational citizenship behavior. For task performance, GMA was the strongest predictor.

A general critique of personality test usage in selection concerns the relatively low validities of around .20 that are typically found (Morgeson et al., 2007a, 2007b). However, responses to the Morgeson debate have been more optimistic. For example, Ones and Viswesvaran (2007) highlighted that the FFM is related not only to performance but to other relevant organizational outcomes (e.g., teamwork and leadership) as well. Tett and Christiansen (2007) argued that higher validity could be obtained by a more proper alignment between tests and outcome criteria. Another theme in this debate was that FFM tests suffer from being prone to socially desirable responses and faking. Tett and Christiansen (2007) argued that faking attenuates the predictive validity of personality tests, although enough trait variance remains to help predict job performance. In contrast, Ones and Viswesvaran (2007) argued that response distortion among job applicants is overrated.

Most selection practitioners have probably experienced how self-report measures may be vulnerable to socially desirable responses and faking, which the literature has also documented (e.g., Birkeland et al., 2006). Different response formats, such as a forced choice format instead of Likert rating scales, have been suggested to hinder faking. While such response formats may produce less faking, they do not prevent it altogether, according to a large-scale study by Wetzel et al. (2021). From an applied perspective, it may be a challenge that forced choice tests yield an ipsative (i.e., intraindividual) profile. Such results may not represent the measures of independent FFM traits (Martinussen et al., 2003), making the interpretation of test scores somewhat cumbersome. As for advice to the selection practitioner, authors usually warn them not to ignore faking but to be aware of the challenges that it can produce (e.g., Hughes & Batey, 2017).

Another challenging area for linkages of personality and job performance is related to a temporal perspective. Predictors such as GMA and FFM tests from a selection phase might differ in predictive validity depending on the time of data obtainment regarding a criterion (i.e., job performance measures). For example, A and O were more related to performance at an initial transition stage than at a later maintenance stage in sales occupations – for the latter, C and E were more related (Thoresen et al., 2004). Thus, a predictive validity study may find FFM traits to be weak correlates of job performance due to the timing of measuring the latter. For military occupations, Darr (2011) noted that O seemed to be curbed in initial training phases due to the typical routines characterizing the start-up of service life (e.g., learning instructions and basic weapon use). New ideas and original problem solving may become more relevant later on.

The use of FFM tests in personnel selection is thus not without its critics. In addition to the relatively low validities found (also for military samples) and potential for faking, challenges with self-reporting naturally arise from different degrees of self-understanding among applicants. This acknowledgment is probably the main argument for using a self-report FFM measure as background material for applicant evaluations in interviews or other rater-based selection methods. Of course, the quality of such a practice is contingent on the test user's knowledge of the constructs measured and the psychometric properties of test scores. However, it is generally recommended to combine predictors in a "mechanical" way for achieving optimal predictive validity in personnel selection instead of leaving a predictor (e.g., a personality test) open for "clinical" interpretation (Kuncel et al., 2013). For example, a mathematical weighting principle can combine interview ratings and personality scale scores to reach a final score that differentiates the applicants. Weighting rules can be derived from incremental validity findings. In practice, though, this is not easy to comply with (Highhouse, 2008; Meijer et al., 2020) – perhaps especially so for personality tests, as the manuals themselves advocate for a qualitative interpretation of results rather than a mathematically derived profile based on score cut-offs (e.g., McCrae & Costa, 2010).

1.3.3 Summary

To summarize, there is a longstanding interest in personality testing in military contexts, which has culminated in the use of FFM or similar measures in assessment and selection practices. The content of military job analyses, predictive validity studies, and findings of incremental validities legitimize the use of such tests. However, there are challenges related to the relatively low validity and vulnerabilities in self-report measures. Using FFM measures as

background material in rater-based selection methods by knowledgeable may alleviate challenges related to faking and differences in self-understanding. However, such a practice is generally not recommended in the literature, as a mechanic combination of predictors yields the highest predictive validity.

The following section addresses FFM research findings and perspectives relevant to Study 1 (officer selection ratings) and Study 2 (special forces personnel). There is scant relevant published research for these studies compared to the predictive validity literature. Therefore, this thesis expands on FFM personality discussions within military psychology beyond predictive validity. Still, while the data in Studies 1 and 2 did not include criteria of occupational performance, the findings are relevant for discussing military selection procedures. Study 3 – and to a certain degree, Study 1 – are concerned with psychometric investigations, the very foundation for interpreting how test scores are related to the FFM.

1.4 Officer selection ratings

In the two-week-long Norwegian military officer selection program, candidates are individually rated for their potential to develop into military leaders. The ratings rest on competency formulations believed to operationalize a mission command leadership style. In the first week, candidates who pass medical examinations and physical tests are interviewed by a selection officer. Those who pass the interview attend a seven-day selection exercise simulating a war-like scenario outdoors (i.e., in the field), where they are assessed again by a selection officer. The leadership competencies are rated by employing behaviorally anchored rating scales (BARS), which the method section of this thesis will explain in more detail.

In terms of measurement methodology, these selection methods correspond to civilian employment interviews and assessment center (AC) exercises (e.g., job simulations). Several studies relying on civilian samples have examined the extent to which the assessment of applicants in such selection methods overlap with their FFM personality traits (Collins et al., 2003; Hoffman et al., 2015; Roth et al., 2005; Salgado & Moscoso, 2002). By examining whether personality constructs are embedded in the ratings, these studies increase our understanding of predictors relying on rater-based methods. There are few studies on this topic within military psychology, however. In this thesis, Study 1 investigates how officer candidate personality traits relate to assessment ratings; the findings contribute to the literature on this matter. Additionally, Study 1 incorporates a psychometric investigation of the NMPI-80, as the measure was correlated with the NEO-PI-3. It is also used in parallel

with the NEO for investigating the associations between traits and ratings. Study 1's findings can help provide advice concerning FFM test usage in this selection program.

With regard to military selection interviews, while there are studies of predictive validity (e.g., Darr & Catano, 2016; Køber et al., 2017), published military studies regarding FFM overlap with interview ratings are lacking, to my knowledge. When inspecting meta-analyses using civilian samples, E and C show mean correlations of around .15 with employment interview ratings (Roth et al., 2005; Salgado & Moscoso, 2002). Wiersma and Kappe (2016) speculate that those traits may fuel such characteristics as assertiveness (E) and intrinsic motivation (C), which is often interpreted as advantageous in an interview setting.

Studies concerning the FFM and ratings in military AC exercises exist, although few in number. Two studies have focused exclusively on the association between E and leadership ratings obtained in a five-week AC (Thomas et al., 2001) and a 15-week training course (Darr et al., 2018). These studies found small positive correlations of .14 and .16, respectively. Sørlie et al. (2020) used samples from the Norwegian officer selection program to investigate the predictive validity of a person-organization fit measure. Based on their regression model, the authors reported minor predictive impacts of E and O ($\beta = .17$ and $-.10$) toward the field selection exercise's overall rating.

Inspecting meta-analyses based on civilian samples, the results seem somewhat contradictory. One example is Collins et al. (2003), who reported that the FFM traits except for C (which they did not include) correlated with AC ratings (mean $r = .16 - .47$), where E was the strongest correlate. Those results were based on five to 13 studies, where the mean correlations were corrected for range restriction and unreliability in the criteria. Another example is Hoffman et al. (2015), who performed a meta-analysis investigating all FFM traits for different forms of validity for different AC exercise categories (in-basket, leaderless group discussions, roleplay, case analysis, and oral presentation). Regarding the overlap between the FFM and AC ratings, several non-existent and weak mean correlations corrected for predictor and criterion unreliability were reported (based on two to 13 studies, depending on which trait and which AC exercise category). The most notable finding was that O had a mean correlation of .20 toward case analysis, which was categorized as a writing exercise. Such different findings could point to the complexity, and perhaps also to local variations, in the AC method.

The findings from studies that investigate the degree of construct overlap, whether personality or GMA, with assessments from rater-based methods, can be of practical value for evaluating cost-effectiveness in a selection program. As Collins et al. (2003) suggested based

on their findings of notable correlations, costly selection methods using rater-based assessments (e.g., AC exercises) may be replaced by cost-friendly psychometric tests. However, given low construct saturation, the argument could be that psychometric tests may provide incremental validity to the predictive validity offered by interviews or AC exercises. Of course, costly rater-based selection methods may hold other benefits, such as giving applicants realistic job previews.

1.5 Special Operations Forces

Special Operations Forces (SOFs) are considered to be one of the most demanding military specialties. Candidates for such a service must participate in particularly challenging assessment procedures for final selection. Once employed, operators engage in physically harsh and advanced tactical training regimes. Examples include survival training, parachute jumping and combat diving exercises, and handling complex equipment in high-stress contexts. Furthermore, operators are expected to perform missions that require an above-average skill set and a risk-taking attitude relative to those serving in conventional forces. An example is counterterrorism operations with a high-risk profile.

The term "warrior-diplomat" describes the ideal profile of SOF personnel (Berg-Knutsen & Roberts, 2015). By weighting the diplomacy as well as the warrior aspect, this description emphasizes that operators must manage to combine knowledge of societal perspectives with a tactical level skill set. Thus, operators can perform optimally in, for example, international combat operations where they can function as both liaisons between different organizations and operational assets if needed. Due to the "specialness" of operators, what Spulak (2009) underlined when naming such soldiers "elite warriors," studies of SOF personnel personality characteristics is of interest to the field of military psychology. Knowledge of the FFM characteristics of employed SOF personnel can contribute to an increased understanding of exceptionally talented and particularly high-performing persons in a military setting.

Only a few publications concerning SOF personnel characteristics exist, likely due to confidentiality issues and the secrecy surrounding such units. A NATO technical report (2012) highlights personnel requirements such as versatility, speed and surprise, and being able to work both independently and in direct support of others. Further, Picano and Roland (2012) introduced the concept of "high-risk operational personnel" to encompass demanding service roles such as SOF operators, intelligence operatives, and astronauts. They report that six attribute dimensions are commonly required for successful performance in such

occupations: emotional stability, adaptability, teamwork abilities, physical stamina and fitness, sound judgment and decision making, and intrinsic motivation. While these publications provide insight into which personality functions are deemed essential in demanding services, they do not report empirical findings.

However, Braun et al. (1994) provided empirical data by collecting the NEO-PI-R scores of 139 Navy SEALs. Relative to the norms for adult American males, this sample of SEALs had lower scores on N and A, the same to lower score on O, and higher scores on C and E. At the time of writing Study 2 (2019–2020), there were no other publications presenting empirical findings on the FFM traits of SOF operators, though a later Danish study reported that upcoming operators in a naval special force (called Frogmen) scored higher on C and A relative to university students (Bech et al., 2021). Thus, findings from different countries do not necessarily align.

Of particular interest in the Norwegian SOF environment is the organization of two separate – but equal in terms of special forces approvement – units. These are the Special Operations Commando (FSK in Norwegian), a department formerly affiliated with the Army, and the Naval Special Operations Commando (MJK in Norwegian), a department formerly affiliated with the Navy. The former has historical roots in paratrooper operations, the latter in frogman operations (Olsen & Thormodsen, 2014). The operator selection processes to the FSK and the MJK are separate, and both selection systems are considered to be highly demanding in terms of physical and psychological requirements. A third personnel category in the Norwegian SOF environment may be called SOF support. While this personnel category is diverse, including personnel from several military specialties and backgrounds, they share the common characteristic of not having been through an operator selection and training process.

As Study 2 used the NMPI-80 for measuring personnel across the different categories, intradepartmental comparisons can made. Further, by comparing the FFM scores of operators with those of the officer selection candidates (from Study 1), we can investigate if the "specialness" of the operators was reflected in their personalities. From an applied perspective, insights into the personalities of employees can be used to strengthen recruitment and personnel selection processes. However, with regard to personnel selection, the findings from such a study design are not as informative as those from a predictive validity study in which test results are correlated with performance measures. Nevertheless, the FFM profiles of successful employees may serve as a guiding framework for evaluating personality strengths among candidates for SOF selection.

1.6 Psychometrics

A fundamental assumption in psychological science is that constructs relate to observable behaviors and that measurement, therefore, can involve a translation of the construct in question into behavioral terms (Cronbach & Meehl, 1955; Murphy & Davidshofer, 2001). In line with this reasoning, the most common measurement method for the FFM involves registering individuals' responses to items operationalizing the five traits. The items are often presented on a Likert scale, where one must decide the degree to which the descriptions in the items relate to self-perception. An aggregation principle is applied when calculating test scores, which means that the items that are somewhat similar – albeit containing slightly different operationalizations – are merged to reach an overall trait score. Hundreds of example items that operationalize different constructs related to the FFM can be found in the open web resource the "International Personality Item Pool" (Goldberg et al., 2006).

Psychometrics concerns theories and techniques for measuring mental constructs; the concepts of validity and reliability are of primary interest. Information about recommended thresholds for validity and reliability estimates are, for example, obtainable in a publication by the European Federation of Psychologists' Association (EFPA, 2013). A more elaborate source for guidance on testing and psychometrics is the American "Standards for Educational and Psychological Testing" (2014). While reliability is necessary for validity, it does not guarantee the latter. These concepts are therefore treated separately.

The Standards describe validity as "the degree to which evidence and theory support the interpretations of test scores for proposed uses of the test" (p. 11). Thus, it is not the test itself that is validated but, rather, the very use of the scores. For example, when interpreting how scores on an FFM measure reflect the personality trait levels of the test taker, it is this interpretation that needs to be validated. The newest version of the Standards (2014) emphasizes that validity is a unitary construct that should not be divided into the classical types of validities – content-related, criterion-related, and construct-related. Instead, a validity argument should be made, where the goal is to reduce the two major threats to validity, which are often termed construct underrepresentation and construct-irrelevant variance (Reynolds et al., 2021). A validity argument identifying the strengths and weaknesses of the intended test score interpretation can be based on five different evidence sources or categories (Standards, 2014).

In the first category, *evidence based on test content*, test developers focus on analyzing the relationship between the content of test materials (for example, the wording of items) and the construct in question. The second category concerns *evidence based on response*

processes, which involves inspecting if test takers' responses are relevant for the intended construct. One approach for investigating such evidence may be an inspection of whether test items and scales are vulnerable to socially desirable responses. Third, when analyzing if the relationships between test items and components reflect expectations (i.e., as predicted by the construct), *evidence based on internal structure* is investigated. The widespread use of factor analyses in personality research may be categorized here. Often, test scores are investigated for their ability to predict behavior in other areas, such as job performance. Such predictor-criterion investigations, formerly known as concurrent or predictive validity studies, illustrate one of the approaches subsumed by the fourth validity evidence category, *evidence based on relations to other variables*. The last category, *evidence based on consequences of testing*, is first and foremost related to inspecting if the assumed benefits of testing are obtained. Arguments related to the soundness of score interpretations, including consequences such as adverse impact and the user experience, can be incorporated here (Standards, 2014).

Reliability concerns the consistency or precision of test scores (Kaplan & Saccuzo, 2009; Standards, 2014). In classical test theory (CTT), every score on a psychological test (X_i) is seen as being composed of a true score (T_i) and an error component (E_i), yielding the equation: $X_i = T_i + E_i$. The error component symbolizes factors that can either decrease or increase an individual's observed score, while the true score represents the score obtained if the measure in question is flawless. In addition, it is assumed that the mean error is zero ($M_e = 0$), that error components are unrelated ($r_{ee} = 0$), and that the error and true score are also unrelated ($r_{te} = 0$) (Magnusson, 1961). Error components can, for example, be related to a changing mood, training effects, or the physical environment in which the test is administered; they are collectively referred to as the random measurement error. There are also measurement errors related to item sampling, which is the extent to which the test items represent the domain (i.e., the construct).

Reliability estimates are concerned with such random errors (E_i) that can produce different observed scores (X_i) at different administrations of a test and thus contaminate the true score (T_i) of an individual. Using the CTT equation for thinking in terms of variances, the total score variance is the sum of the true score variance and the error variance. Mathematically, a reliability coefficient can be calculated from the ratio of the true score variance to the total score variance. A reliability estimate is thus the proportion of test score variance due to true score differences (Reynolds et al., 2021). There are several traditional ways of investigating reliability. When a test can be administered on more than one occasion, the scores can be correlated to yield a test-retest reliability estimate. If parallel forms of a test

exist, the correlation between the two forms gives a parallel form reliability estimate. A single test administration can achieve internal consistency reliability estimates and concerns the stability in responding to items intended to measure the same construct.

Regarding internal consistency reliability estimates, perhaps the most widespread approach is the coefficient alpha (Cho & Seonghoon, 2015), which produces an estimate for the average inter-item correlations. This procedure provides an answer to the extent to which responses within a measure, or a scale, are reliable and consistent. A high alpha indicates high item interrelatedness (presumably high true score variance), while a low alpha indicates low item interrelatedness (presumably low true score variance). However, a very high alpha may indicate item redundancy (Streiner, 2003). A common interpretation of the alpha coefficient is that it reveals information about the degree to which items seem to measure the same underlying construct (i.e., high inter-item correlation, high alpha) or if too many construct-unrelated characteristics are measured (i.e., low inter-item correlation, low alpha).

CTT has weaknesses with regard to psychological measurement, especially for the concept of error (Murphy & Davidshofer, 2001). For instance, while internal consistency reliability estimates do not consider potential errors due to temporal factors (e.g., carryover effects), test-retest strategies are not eligible for investigating the influence of systematic errors (e.g., repetitive inaccuracy due to faulty administration).

In addition to the traditional reliability approaches described above, rooted in CTT, there are also more mathematically advanced approaches. Two examples used in Study 3 of this thesis are graded response models based on item response theory (IRT) and approaches investigating measurement invariance. Graded response models can demonstrate that responses to an item may be more or less reliable depending on the latent trait level that is supposed to be measured. Such an analysis can contribute to a more nuanced reliability estimate relative to traditional approaches (Hambleton & Rogers, 1989; Reynolds et al., 2021). While CTT is concerned with the scale/test level, IRT is first and foremost concerned with the item level. For inspection of personality test items, a difficulty parameter (β) can reveal information about the probability of endorsing positive response options based on the latent trait level (θ). Items deemed difficult require a high latent trait level for a test taker to respond positively (i.e., agreeing with the item), whereas a less difficult item does not. A discrimination parameter (α) for a personality test item concerns the item's ability to differentiate the latent trait (θ). Mean item discriminations can be used to inspect scale

properties, revealing the thresholds at which scale scores become less reliable for measuring the trait in question.

Measurement invariance can be affected if different groups of individuals interpret test items differently. The consequence of this can be that different constructs are being measured, which are not based on the test content but rather on the group characteristics of those being tested (Cheung & Rensvold, 2002). As such, evidence of measurement invariance can be important for ensuring that a test measures the same construct across different groups.

Although it is concerned primarily with reliability, IRT and measurement invariance analyses can reasonably be included in the second category of validity evidence formulated in the Standards (2014), aiming to inspect if test takers' responses are relevant for the intended construct. Responses to items that reveal difficulty and low discrimination align poorly with the construct meaning, as do scales that show measurement variance across groups. In addition to estimates of socially desirable responses, such analyses may inform personality response processes quite broadly.

To summarize, one can apply different strategies regarding validity and reliability for evaluating psychometrics. The goal is to present a validity argument and a reliability estimate aligned with the intended use of test scores (Standards, 2014). Nuanced psychometric evidence can be obtained by using additional techniques than those based on CTT, such as analyses based on IRT and measurement invariance. Study 3 especially takes on this goal by analyzing NMPI-50 responses from a large sample using selected statistical approaches described later.

1.7 Thesis aims

This thesis aims to present and discuss NMPI findings related to officer selection ratings, special forces personnel, and psychometrics. Three empirical studies with specific purposes were published for this thesis. Combined, the three studies illustrate perspectives on the linkage between personality and military from a Norwegian context and present different approaches for investigating the psychometrics of a short-form FFM measure. Together, these studies examine whether the NMPI versions soundly measure the FFM and reveal whether it is worthwhile to use the NMPI in military selections.

1.7.1 Study 1

Study 1's primary aim is to explore the associations between candidate personality traits and mission command competency ratings. The findings allow for discussing the degree of personality saturation in the widespread rater-based military selection methods of interviews and field selection exercises. The secondary aim of Study 1 is to preliminarily validate the NMPI-80 by correlating scale scores with NEO-PI-3 domain scores. In addition, by inspecting similarities – or lack thereof – in the NMPI and the NEO associative patterns toward ratings, the NMPI-80 psychometric properties can be further interpreted.

1.7.2 Study 2

Study 2 aims to investigate the personality traits of SOF employed personnel using the NMPI-80 preliminarily validated in Study 1. In addition, data concerning personnel category, age, number of deployments, and rank were obtained. Comparing the NMPI scores of SOF operators with those of the participants in Study 1 allows for the investigation of whether operators' "specialness" was reflected in their FFM profiles. Norwegian operators were assumed to be higher on ES, C, and O but lower on E and A. Study 2 includes internal consistency estimates (alpha) regarding the psychometric investigation. Inspecting the expected differences between operators and officer candidates also provides information on the test's ability to discriminate between different military groups.

1.7.3 Study 3

After the data collection for Studies 1 and 2, the Armed Forces HR and Conscription Centre advocated for assessing conscript personality. The NMPI-80 was deemed to be too long for this purpose; thus, the NMPI-50 with fewer items was developed. Using an age cohort undergoing conscript assessment, Study 3 aims to uncover NMPI-50 psychometrics by employing factor analysis, item response analysis, and measurement invariance analysis.

Additionally, to evaluate social desirability effects, Study 3 inspects how responses to the NMPI-50 were related to self-reported military motivation.

2 Methods

2.1 Samples, procedures, and ethical considerations

For Study 1, we asked all 1,287 candidates attending the first two days of the joint selection program for basic officer schools (Army, Navy, and Air force) to participate. The final sample consenting to participate was $N = 901$, which yielded a response rate of 70%. There were 207 women (23%) and 694 men (77%) in the final sample, and the age range was 18–34 years ($M = 19.6$, $SD = 1.86$). A paper-and-pencil version of the NMPI-80 and the NEO-PI-3 was used. Due to the attrition of candidates as the selection program advanced, $N = 810$ obtained interview ratings and $N = 551$ obtained field ratings. There was no differential attrition concerning sex or age. We informed candidates of the study's purpose orally and in writing, while acknowledging ethical challenges such as obedience to authority and pressure to conform when obtaining informed consent in a military arena. Ethical challenges were also reduced by considering the NMPI and the NEO as unobtrusive measures – that is, as not having clear potential to challenge the well-being of the participants. As for the formalities, Study 1 was part of a more extensive research program at the Norwegian Defence University College, "The Leadership Candidate Study" (NAF, 2021), which the Norwegian Centre for Research Data had approved.

For Study 2, an invitation letter was sent by a military mail system to all employees in the SOF. The study was approved by the Norwegian Social Science Data Service, the research commission at the Norwegian Defence University College, and the SOF command. Those who consented to participate were asked to respond to a digitalized version of the NMPI-80 and answer questions concerning demographic information. The final sample was $N = 190$ (almost all males), of which 76 were operators (all males) and the remaining were support personnel. The response rate is unknown, as the actual size of this department is classified information. The ages of the participants ranged from younger than 30 to older than 40; they were coded in categories in order to ensure anonymity. Most participants were in their 30s. We evaluated ethical challenges related to this study as being less prominent relative to Studies 1 and 3.

In Study 3, we obtained data from the conscript assessment procedure, where all Norwegian males and females are instructed to answer an online survey questionnaire in the year that they turn 17. The NMPI-50 was attached to this questionnaire in 2019 (the 2002 cohort), which resulted in a sample of $N = 52,760$. Approximately half were males (50.1%) and half were females (49.9%). The survey instructs respondents to be as honest as possible,

that there are no right and wrong answers, and that responses should not be influenced by how they view military fitness or motivation. Particular ethical challenges were actualized when using the data obtained with the omission of consent from the participants. We concluded that three aspects legitimized the use of these data. First, the research team had exclusive access to anonymized data. Second, we concluded that the NMPI-50, which measures normal personality, did not have clear foreseeable adverse consequences for the participants (e.g., long-term negative psychological reactions). Third, the independent research commission at the Norwegian Defence University College approved the use of these data. We also concluded that we had an ethical and professional responsibility to investigate the psychometric properties of a test that the Armed Forces HR and Conscription Centre used operationally.

2.2 Study variables

2.2.1 Study 1

In addition to the sex and age of Study 1 participants, the following variables were used:

Norwegian Military Personality Inventory - 80 (NMPI-80)

The NMPI-80 is a self-report seven-point Likert scale test consisting of 79 items. Most items were extracted from the International Personality Item Pool database (Goldberg et al., 2006) and translated into Norwegian. Military psychologists provided the remaining items. All items are reported in Appendix A. There are no facet scores for this test. The five scales (number of items in parentheses) of the inventory are ES (15), E (17), O (17), A (16), and C (14). Test completion time was approximately 15 minutes, and the raw scores were used in this study.

NEO Personality Inventory-3 (NEO-PI-3)

The NEO-PI-3 is a 240-item self-report five-point Likert scale test that includes six facets for each of the five domain scales (N, E, O, A, and C) (McCrae & Costa, 2010). Study 1 used the Norwegian version of the NEO-PI-3 (Martinsen, 2007). Test completion time was approximately 45 minutes, and the raw scores were used in this study.

General mental ability (GMA)

The GMA measure used was a total score based on three tests that are part of the conscript assessment procedure, which measure reasoning, numerical abilities, and verbal abilities (Køber et al., 2017). The administration time is one hour. A previous study (Skoglund et al.,

2014) documented adequate parallel form reliability by correlating the total score from a paper version and a computerized version of the GMA test ($r = .85$). The present study used the mean GMA stanine score.

Interview ratings

The interview lasted one hour and was performed by an experienced selection officer. The results from the NMPI and NEO were not available. Three leadership competencies were rated by the interviewer: (1) *role model* – acts in line with the NAF's core values, is open to feedback, and shows integrity; (2) *mental robustness* – can cope with high demands and stressful life events, is emotionally stable, and adapts to uncertain circumstances; and (3) *development* – stimulates autonomy in others and encourages reflection, original thinking, and self-development in others. Interviewers scored each competency on a nine-point scale (1 indicated the weakest score and 9 indicated the strongest) in alignment with a scoring system that operationalized each competency into example answers along with the scales. An overall assessment rating (OAR) was calculated as the mean rating.

Field selection exercise ratings

The exercise lasted seven days in the field in a simulated warlike scenario. The results from the NMPI and the NEO were not available here. The candidates took turns solving ongoing work sample cases as leaders within teams of five to seven, where a selection officer followed the team and rated the performance of the candidate holding the leadership role. The competencies of *role model*, *mental robustness*, and *development* were rated, in addition to (1) *task focus* – takes initiative, works systematically toward goals, and prioritizes adequately and (2) *cooperation* – gains trust from others, communicates efficiently, delegates, and supports others. A total of five competencies were thus rated. The selection officers used a nine-point scale, operationalized in behaviorally anchored rating scales (BARS) adapted to the different work sample cases. An overall assessment rating (OAR) was calculated as the mean rating.

2.2.2 Study 2

Study 2 used the NMPI-80 as described for Study 1. The other variables recorded were:

Background

The background variable was registered as either FSK operator, MJK operator, or other. Personnel in both operator categories are approved special forces operatives. The other category was collectively named SOF support, which was comprised of personnel contributing to the operational capacity in diverse ways. Examples include participating in training for operations, providing contributions in various missions and combat deployments, and executing leadership from different organizational hierarchies.

Age

Age was registered in three categories: 29 and under, 30 to 40, and 41 and above.

Number of deployments

Deployments was registered in four categories: 0, 1–5, 6–10, and above 10. Only three participants reported above 10 deployments; this variable was thus recoded into three groups (0, 1–5, and above 5). In the last decade, the Norwegian SOF has usually been deployed four to six months at a time in international conflict areas to participate in combat operations.

Rank

Rank was registered in three categories: Other Ranks, Officer Rank, and Civilian. Those with officer ranks have graduated from a three-year-long education program at a military academy, resulting in a bachelor's degree.

2.2.3 Study 3

In addition to the sex of Study 3 participants, the following variables were used:

Norwegian Military Personality Inventory - 50 (NMPI-50)

This measure consists of 50 items rated on a seven-point Likert scale. The NMPI-50 is thus somewhat shorter than the version used in Studies 1 and 2, and it also contains four new items: "View myself as an outgoing person" (E), "Frequently becomes the leader of groups" (E), "Trust others" (A), "Desire for self-development" (O); and "Takes time to reflect" (O). The scale takes 10–15 minutes to complete. Appendix B reports all items.

This variable was measured by a five-point scale (completely true to completely untrue) on a single item: I wish to serve in the Norwegian Armed Forces.

2.3 Statistical analyses

IBM SPSS was used for all statistical analyses in Study 1 (version 26.0) and Study 2 (version 25.0). STATA (Statacorp, 2017) was used for all statistical analyses in Study 3. Cronbach's alphas were computed for the NMPI-80 and NMPI-50 scales for all three samples.

2.3.1 Correlation and regression analyses (Study 1)

In addition to correlating all study variables, hierarchical regression analyses were performed in Study 1 for testing hypotheses related to associations between personality variables and ratings.

The Pearson bivariate correlations between the NMPI and the NEO scales on one side, and the competency ratings obtained respectively in the interview and the field exercise on the other, were calculated for investigating expected associations.

Hierarchical regression analyses were applied to test hypotheses related to expected linkages between personality traits and the OARs. A total of four analyses were performed. The two candidate OARs (as obtained in the interview and the field exercise) were regressed separately on the NMPI and NEO scales. In all four analyses, the control variables of sex, age, and GMA were entered in the first analytic step, followed by the personality variables in the second. First, the explained variance in the OARs was inspected for changes when entering the personality variables in the four models. Second, in the final models where all variables were entered, the personality variables' standardized regression coefficients (β) were evaluated for possible unique statistical contributions.

Besides the statistical analyses related to hypotheses testing, the NMPI and NEO scales were correlated, thus producing a validity estimate of the NMPI-80. Lastly, the NEO facets were correlated with the ratings (both OARs and single competencies), which were presented in a supplementary file in the publication.

2.3.2 *t*-tests and analyses of variances (Study 2)

Study 2 applied statistical techniques to compare groups due to the category-coding of demographic variables. Firstly, histograms and Q-Q plots were inspected to evaluate parametric assumptions, and they were further assessed by using Levene's homogeneity tests

for the specific analyses. Secondly, independent samples *t*-tests and one-way between-groups analyses of variances (ANOVAs) were used to compare the NMPI-80 scale mean scores. For the ANOVAs, the Tukey HSD test performed post-hoc comparisons of the groups (if parametric assumptions were met). The effect sizes for mean differences were calculated as Cohen's *d* for the *t*-tests and eta squared (η^2) for the ANOVAs.

2.3.3 Factor, item response, and measurement invariance analyses (Study 3)

Study 3 calculated the means and standard deviations for the NMPI-50 scales and the motivation variable (indicator of social desirability) and correlated the observed scores on all study variables.

The sample was split into three equal parts for conducting an exploratory factor analysis (EFA) ($N = 17{,}586$), a confirmatory factor analysis (CFA) ($N = 17{,}587$), and an item response analysis ($N = 17{,}587$).

The EFA used a promax oblique rotation presupposing correlated factors. A parallel analysis determined the number of retained factors. If the difference from a factor's associated eigenvalue was larger relative to the 95th percentile of the distribution of the randomly derived eigenvalues, it was retained. We also inspected the unique contribution of an additional factor.

The CFA used maximum likelihood estimation for determining model fit, based on the following indicators: chi-square (χ^2), comparative fit index (CFI), root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), and the Tucker–Lewis index (TLI). We chose to set satisfactory cut-offs at < 0.08 for RMSEA (MacCallum et al., 1996) and SRMR (Hu & Bentler, 1999) and at > 0.90 for CFI and TLI (Hooper et al., 2007).

After the CFA, possible differences in the correlations between observed scores and latent traits were inspected. Model fit was then tested for each factor, in addition to a five-factor solution, a three-factor solution (N+E, C+A, and O), and a bi-factor solution (five specific factors and a general factor). Covariance parameters between latent factors were included from all factor solutions, and modification indices were used for determining the presence of correlated error variance. The general factor from a fitted bi-factor solution was correlated with the motivation variable.

Based on the results from the EFA and the CFA, the multidimensionality of factors with poor fit was assessed with the factor determinacy coefficients using the FSDET module of STATA (Mehmetoglu, 2019).

Graded response models (which is based on IRT) were estimated for each of the five factors. For determining the range sensitivity of the scales, a standard error of measurement of less than 0.5 was used as a threshold. Difficulty (β) and discrimination (α) parameters for each item were calculated.

Finally, using the full sample, the NMPI-50 scales were tested for metric and scalar invariance in order to investigate if the same constructs were measured across the sexes. In this multigroup CFA approach, the CFI values were inspected as obtained in a metric model, a scalar model, and a two-group baseline model for each factor. As the large sample deemed significance testing with likelihood ratio to be suboptimal, the null hypothesis of sex (group) invariance was instead rejected if the CFI difference was smaller than or equal to -0.01 (Cheung & Rensvold, 2002). The scales were thus interpreted as not achieving measurement invariance if the CFI differences were larger than -0.01.

3 Results

3.1 NMPI-80 scale reliability and correlations with the NEO

The psychometric properties of the NMPI-80 used in Studies 1 and 2 were inspected through Cronbach's alpha values and obtained correlations with the NEO-PI-3 domains.

Based on the samples in Studies 1 and 2, the Cronbach's alpha values were respectively .89/.87 for ES, .89/.88 for E, .78/.75 for O, .90/.88 for A, and, lastly, .87/85 for C. Streiner (2003) recommended an alpha above .80 for research tools. He also emphasized that an alpha above .90 might point to the redundancy of items within a scale. Therefore, we evaluated the scale reliability of the NMPI-80 as being good to excellent, which is also in agreement with the recommendations put forth by EFPA (2013).

The correlations between the NMPI scales and the NEO domains were as follows: ES (reversed N), $r = -.82$; E, $r = .80$; O, $r = .80$; A, $r = .62$; C, $r = .82$. Additionally, a large correlation was observed for NMPI E and NEO A ($r = .61$). Therefore, NMPI A held less similarity to NEO A than the remaining scales with overlapping construct meanings. Otherwise, there were no apparent correlations across the other scales in the two inventories. The correlations obtained are considered good (A) and excellent (ES, E, O, and C) according to EFPA (2013).

Overall, alpha reliability and validity estimates as obtained by correlations with the NEO gave rudimentary support for interpreting that the NMPI-80 scale scores adequately measured the FFM for research purposes.

3.2 Officer candidate personality traits and selection ratings

In the regression analysis controlling for sex, age, and GMA, the inclusion of the NMPI-80 scales provided a statistically significant contribution for explaining the competency rating variance in both the interview (7%) and field selection exercise (3%). Replacing the NMPI with the NEO in the same analytic procedure provided only a statistically significant contribution for the interview (5%). Thus, personality variables contributed to an overall marginal increment in explained variance above that provided by the control variables, somewhat higher for the interview ratings than for the field selection exercise ratings.

Inspecting the standardized regression coefficients from the analyses, the NMPI E and O were significant predictors toward both selection methods ($\beta = .25/.17$ and $-.14/-.16$). These results aligned with the NEO, although NEO O did not predict the interview ratings. Therefore, when controlling for age, sex, and GMA, E was positively related to the ratings to

some extent, while O was somewhat negatively related. The individual predictors of N, A, and C were not associated with the ratings in a statistically significant way.

Inspection of the correlations between the personality variables and the interview competency ratings (role model, mental robustness, and development) indicated no clear competency-dependent associative patterns. Regarding the field selection exercise, a marginal tendency of negative associations was found between O and the competencies of role model, task focus, and mental robustness, but not for cooperation and development.

The NMPI and the NEO coincided in associations toward the ratings. An exception concerned the NMPI A and NEO A, which had different correlations toward the overall interview rating – $r = .16$ and $.06$, respectively. The moderate correlation between the A scales may explain this difference.

3.3 Personality characteristics of special forces personnel

The ANOVAs showed the following statistically significant differences of the NMPI-80 scale scores across demographic variables: (1) ES across the three age groups: $F (2, 187) = 5.17$, $p < 0.01$; (2) A across the three groups of number of combat deployments: Welch's $F (2, 41.61) = 5.33$, $p < 0.01$; and (3) ES across the groups of number of combat deployments: $F (2, 186) = 3.22$, $p < .05$. These group differences resulted respectively in 5%, 4%, and 3% explained variances in the personality variables (η^2). As summarized by post hoc comparisons, older personnel had a higher mean ES score than younger personnel did. Furthermore, those with more combat deployments had higher ES mean scores but lower A scores than those with few or no deployments. For rank, a t-test demonstrated that officers ($M = 4.93$, $SD = 0.81$) had a significantly higher E mean score than non-officers ($M = 4.54$, $SD = 0.79$): $t (187) = -3.41$, $p < .001$). Cohen's d was 0.50. While we expected that older employees were higher on A, C, and ES than their younger colleagues, we were unsure of our expectations concerning deployments and rank. Thus, expectations were met with regard to ES, but not with A and C.

There were no statistically significant findings indicating differences in NMPI scale scores between FSK operators, MJK operators, and SOF supports. When comparing the NMPI scale scores of the SOF support group with those of a merged SOF operator group, t-tests did not produce any statistically significant results. SOF personnel were thus quite similar in their mean personality trait scores across the background variables.

When comparing the NMPI scale scores of the merged SOF operator group with males from Study 1, three findings reached statistical significance. Firstly, SOF operators had lower E scores ($M = 4.67$, $SD = 0.83$) than the officer candidates ($M = 5.16$, $SD = 0.76$): $t (722) = -$

5.37, $p < .001$. Secondly, SOF operators had lower A scores ($M = 5.26$, $SD = 0.68$) than the officer candidates ($M = 5.65$, $SD = 0.68$): $t(736) = -4.72$, $p < .001$. Thirdly, the mean score of ES was higher for the SOF operators ($M = 5.54$, $SD = 0.78$) than for the officer candidates ($M = 5.33$, $SD = 0.82$): $t(729) = 2.07$, $p < .05$. Cohen's d was 0.62 (E), 0.57 (A), and 0.26 (ES). Thus, our expectations of higher C and O among operators did not manifest. However, the findings of higher ES and lower E and A were in alignment with our expectations.

3.4 Psychometric properties of the NMPI-50

For the NMPI-50 scales, the Cronbach's alpha values were .91 for ES, .90 for E, .85 for O, .91 for A, and .90 for C. The scale reliabilities were therefore excellent (EFPA, 2013; Streiner, 2003).

3.4.1 Exploratory and confirmatory factor analyses

Based on the EFA using parallel analysis, five factors achieved eigenvalues above the threshold, explaining 59.5%, 18.6%, 8.9%, 6.7%, and 5.3% of the variance. Adding a sixth factor explained an additional 2.6%. While most items loaded on their corresponding factors in the rotated solution, two items of the O scale had cross-loads toward C: o2 (wish to engage in self-development) and o7 (know many words and concepts). Additionally, three O items had low factor loadings toward the intended factor: o3 (spend time reflecting on life), o1 (notices beautiful objects), and o10 (am curious about other cultures). When including the sixth factor, the O scale was divided into two separate factors. Three items loaded on the sixth, instead of the fifth, factor: o9 (am good at creative thinking), o6 (generates many good ideas), and o5 (have a vivid imagination).

After the CFA, when comparing the intercorrelations among observed scores and latent traits, notable increases were observed for O and E and for O and ES. These differences in associations indicated the presence of measurement error in the O scale. The CFA showed that the bi-factor model with one general factor and five specific factors achieved the best model fit. The RMSEA and SRMR reached the threshold for a good fit, while the CI and TLI did not (a bi-factor model excluding the O indicators produced satisfactory results in all fit indices). The five-factor solution had a better overall fit than the three-factor solution, although it did not reach the cut-off for SRMR. With regard to the five factors, ES, C, and A had moderate to good model fit, while E and O did not reach the recommended cut-offs.

The E and O scales were thus investigated further. Firstly, the effects of correlating indicators of error variances on the overall goodness of fit of the models were assessed. A

correlation between the error variance in pairs of indicators was interpreted as a weakness of the trait in question for encapsulating a sub-facet within the trait – or that the content in one item in a pair was redundant. Based on the modification indices, the error variances of the items e5 (likes to decide), e10 (likes to lead others), and e2 (frequently becomes the leader of groups) were correlated. This model had a satisfactory SRMR and RMSEA, indicating a manifestation of a sub-facet of E pertaining to leadership. For O, the error variances of the item pairs o4 (am fascinated by patterns and colors) and o8 (thinks art is important) on the one hand, and o6 (generates many good ideas) and o10 (am curious about other cultures) on the other, were correlated. The results indicated that aesthetics was not particularly reflected in the O scale. However, intellectual curiosity seemed to be reflected. Secondly, the presence of multidimensionality for the two scales was tested using bi-factor models. For E, introducing a second latent factor resulted in an improved model fit, where the added factor loaded strongly on e2 (frequently becomes the leader of groups), e5 (likes to decide), and e10 (likes to lead others). The factor determinacy coefficient for the added latent factor was 0.908 and was 0.894 for the remaining factor. For O, improved model fit was also witnessed. The added latent factor loaded strongly on o9 (am good at creative thinking), o6 (generates many good ideas), and o5 (have a vivid imagination), which was in alignment with the EFA results. The factor determinacy coefficient for the added latent factor was 0.901 and was 0.885 for the remaining factor.

The general factor, likely related to socially desirable responses, had a clear positive association toward motivation for service ($r = .44$). While the indicators e10 (likes to lead others), e5 (likes to decide), and e1 (view myself as an outgoing person) had the highest loadings on the general factor, the O indicators had the lowest.

3.4.2 Item response analyses

The graded response models (based on IRT) showed that all scales had high to very high mean discrimination (α): A 2.20, ES 2.13, C 2.01, E 1.91, and O 1.57. However, the threshold parameters for response options (β) varied across items and scales. Especially for A, a negative skew was witnessed, in which the most difficult items (i.e., requires high latent trait – θ – for endorsing positive response options) were a5 (am interested in other people) and a3 (has something nice to say about everyone). The most difficult items were, for E, e10 (likes to lead others) and e3 (like to be the center of attention); for ES, n3 (am often afraid of making a fool of myself) and n2 (worry a lot); for C, c6 (completes my duties right away) and c7

(always meet prepared); and, finally, for O, o8 (thinks art is important) and o9 (am good at creative thinking). Item characteristic curves are provided in the Study 3 publication.

With regard to the scales, the acceptable test information was between two standard deviations above and below the mean theta value, except for the O scale, which was more restricted in range. The discriminatory ability for the A scale aligned with the ceiling tendency witnessed in the observed scores. An overall finding was that the accuracy of scores above or below two standard deviations on all five scales have large confidence intervals, meaning that such scores should be viewed with caution.

3.4.3 Measurement invariance

Inspecting the distribution of the observed scores, it became apparent that females had a higher spread of scores across a broader range of all traits except for A. The ceiling effect of A restricted the range of higher female scores more than for male scores. The tests of measurement invariances showed that all scales achieved full metric invariance, meaning that the items load onto the latent traits in a similar way across sexes. However, the scalar invariance testing showed varying results. ES, C, and E showed full scalar invariance. A showed partial scalar variance when the intercept for item a10 (trusts others) was allowed to vary freely across the sexes. O did not achieve either full or partial scalar invariance, where items o8 (thinks art is important) and o6 (generates many good ideas) were substantial contributors to measurement non-invariance. The comparisons of the scores of males and females on A and O were therefore not optimally valid.

4 Discussion

The main findings from the studies of officer selection ratings, SOF personnel, and the conscript assessment procedure inform different aspects of FFM–military linkages and psychometric evidence for the NMPI versions. The findings regarding the overlap between personality variables and officer candidate ratings were twofold. First, E and O had a slight positive association and a minor negative association, respectively, toward OARs obtained in both the interview and the field selection exercise. The NMPI and the NEO correlated highly and provided similar results. Second, neither the NMPI nor the NEO scales had a correlational pattern pointing to differences in single competencies' personality overlap. The rating of single competencies correlated highly.

With regard to the SOF personnel, operators from the FSK and MJK – whose scores aligned – were generally somewhat less extroverted and agreeable and slightly more emotionally stable than male officer candidates. Furthermore, older employees and those with deployment experience scored higher on ES than their younger colleagues and those without or with only a few deployment experiences. Those with deployment experience also scored lower on A than those with only a few or no experiences. With regard to rank, officers (OF) had a higher mean E than specialists (OR) did.

The psychometric findings of the NMPI-50 scores were promising, demonstrating that a bi-factor solution with a general factor and the five factors was the most suitable model. However, the O scale witnessed construct underrepresentation, as intellectual curiosity was the most dominant characteristic. Also, all scales except for O were vulnerable to socially desirable response sets, as indicated by the pattern of loadings toward the general factor, which correlated notably with self-reported motivation for military service. Scale scores between two standard deviations above or below the latent trait level generally showed sound reliability, which means that the scales mostly yielded accurate FFM information across broad ranges. Lastly, cautiousness was warranted when comparing scores between males and females on A and O.

While the three studies differed in research aims and methodologies, their common ground can be organized into two overarching discussion themes. The first theme concerns whether the versions of the NMPI soundly measure the FFM; it is thus devoted to the psychometric findings. The second theme concerns the evaluation of whether it is worthy to use the NMPI in military personnel selection. Considering that predictive validity toward

performance is unknown, such an evaluation will take into account other findings that are deemed to be relevant.

4.1 Does the NMPI soundly measure the FFM?

The FFM traits are psychological constructs assumed to manifest in observable behaviors. For the versions of the NMPI, these observable behaviors are the test takers' graded agreement with items operationalizing the five traits. The items on each of the five scales are then aggregated to calculate a mean score or total item score that purport to reflect the level of the trait construct. The reliability and validity of this measurement procedure rely on several mechanisms to be discussed in light of the psychometric findings reported in the studies.

Reliability estimates based on the Cronbach's alpha values were reported for the three study samples. According to recommendations from different sources (e.g., Cho & Seonghoon, 2015; EFPA, 2013; Streiner, 2003), the alpha values obtained showed that all five scale scores of both NMPI versions had a sound internal consistency. There were no clear indications of item redundancy for either version (i.e., alphas were not too high). The graded response models based on IRT showed that very low or very high scores had large confidence intervals, thus indicating the reduced reliability of such scores. The A scale was particularly affected. Overall, based on the data obtained, the reliability results were promising.

The validities of NMPI-80 and the NMPI-50 are best summarized with the sources of evidence formulated in the Standards (2014). While the current thesis does not contain evidence based on the consequences of testing (the fifth source), validity evidence based on the remaining four was obtained. Evidence based on test content (the first source) relates to the inspection of item content. In the NMPI development project, military psychologists, serving as subject matter experts, provided and evaluated an initial item pool. The finalized item formulations can be seen in Appendices A and B. Although the items, at face value, reveal a proper alignment toward the content domain of the FFM, weaknesses necessarily relate to short-form measures.

Whereas evidence based on response processes (the second source) was not obtained for the NMPI-80, the graded response models, measurement invariance analyses, and the social desirability findings on the NMPI-50 may represent such a validation strategy. Firstly, responses to several items showed that test takers were somewhat unlikely to report lower response options. The items most susceptible for receiving higher (i.e., positive) response options were two in number for each scale, where the A scale was the most severely affected of the five scales. Secondly, the A and O scales did not show scalar invariance, meaning that

males and females may interpret some of the items in those scales differently. Lastly, all scales except for O were vulnerable to socially desirable response sets, as indicated by the pattern of loadings toward the general factor. The E scale was particularly affected.

The factor analyses and tests for measurement invariance for the NMPI-50 represent evidence based on internal structure (the third source). The results demonstrate that this test replicated the FFM in a Norwegian age cohort, where items load onto the latent traits similarly across sexes due to full metric invariance for all scales. However, three findings are of particular importance in nuancing this main finding. First, as a bi-factor solution was the most suitable model, a general factor should be considered in addition to the five scales. Not all modification indices for this solution reached satisfactory cut-offs. However, it is not uncommon to find less than satisfactory cut-offs for CFAs of personality inventories (Hopwood & Donnellan, 2010), even though an EFA is interpreted as satisfactory (Cooper et al., 2010). Second, the O scale had apparent weaknesses regarding construct underrepresentation (Reynolds et al., 2021), where intellectual curiosity seemed to be measured first and foremost, rather than the trait more broadly defined. The findings of O suboptimality (e.g., clear multidimensionality) agreed with other empirical findings (e.g., Källmen et al., 2010) and theoretical perspectives (Sutin, 2015); as such, they were not surprising. E scale multidimensionality should also be acknowledged (primarily due to the leadership items), although this was not as challenging as for the O scale. Third, cautiousness is warranted when comparing scores across males and females on A, and comparisons are not valid for O. This is due to inadequate scalar invariance.

With regard to evidence based on relations to other variables (the fourth source), the NMPI-80 correlations toward the NEO-PI-3 are representative. Strong and meaningful correlations were found, except for the NMPI A scale relating equally to the NEO A and E. An explanation for this might concern the NEO facet of warmth, which is subsumed by the NEO E domain (McCrae & Costa, 2010). Items relating to warmth are intended for the A scale in the NMPI, thus somewhat contrasting the NEO framework. Further evidence of validity relates to the parallel results of the NMPI and the NEO in associative patterns toward competency ratings and the NMPI's ability to discriminate between personnel categories and levels of demographic variables in the SOF, although not all expectations were met.

Overall, the psychometric properties of the measure scores were deemed adequate for interpreting whether the research findings were representative of the FFM. The NMPI-50 is naturally the version for use in further research and practice, as it does indeed soundly measure the FFM. This is not to say that limitations of the measure do not exist. When using

the measure, it is advised to consider a general factor that partly represents socially desirable responses. The O scale's limitation should also be acknowledged. Furthermore, comparisons between the sexes on A and O must be interpreted with caution. Due to few items for each scale (10), the possibility of obtaining rich individual profiles is limited. However, an advantage may be that it is feasible to qualitatively process the gist of the domain content chosen for each FFM trait. Nevertheless, other FFM measures, which include facets such as the NEO-PI-3 (consisting of 240 items), are more suitable for nuanced personality profiling.

Future psychometric research of the NMPI should inspect test-retest reliability, as this would estimate the proportion of random errors in the scores, in alignment with classical test theory, given that a change in the FFM constructs from test to retest is not expected. Such a reliability analysis should be feasible in the conscript assessment arena and is particularly valuable, as the reliability estimates provided in this thesis do not consider time-related errors.

Furthermore, evidence based on the consequences of testing (the fifth validity source lacking in this thesis) should be obtained. A particularly interesting point relates to conscript personality testing and the Armed Forces' general goal of recruiting more females in the mandatory soldier service. A point of concern for reaching such a goal is that the observed scores of the NMPI witness certain differences across sexes – especially for ES, where males ($M = 46.98$, $SD = 11.54$) scored notably higher than females ($M = 37.42$, $SD = 12.37$). The consequences of using a measure showing such a difference for female recruitment might be scrutinized in the future. I also note that a point of optimism for female recruitment relates to the observation that mean scores on C were very similar. A last strategy to inform the consequences of testing might be future measurement invariance analyses based on ethnicity and culture variables, as well as sex.

Socially desirable responses are a challenge in selection settings (Birkeland et al., 2006), as was also witnessed for the NMPI due to the bi-factor solution and its correlation with self-reported military service motivation. Of interest is the ongoing discussion of what a general personality factor represents. It may be interpreted as spurious measurement error caused by test-takers' exaggerated responses toward trait items they interpret as desirable, governed by the context in which the test is taken (Ziegler & Buehner, 2008). Therefore, the general factor represents socially desirable responses threatening construct validity. However, the general personality factor may also be substantive, representing social effectiveness or emotional intelligence (Van der Linden et al., 2021). Thus, it may capture a construct, or combination of constructs, that the Big Five scales do not represent directly. Future research

may attend to this matter for exploring how a bi-factor solution might be more thoroughly understood.

Concrete psychometric approaches that can alleviate socially desirable responding and faking include validity scales and forced choice response formats (Hughes & Batey, 2017). Concerning the first approach, the NAF has a longstanding experience, as the current FFM test (the 5PFmil 2.0) includes such a scale. Therefore, it may be feasible to develop a parallel scale for the NMPI. However, at the time of writing, forced choice formats have never been actualized in Norwegian military personality inventories.

Other future research directions may relate to investigating whether the psychometric properties of scores replicate with other age groups, considering the possibility of trait changes over the life span (Atherton et al., 2020; Roberts & Mroczek, 2008). Finally, if item revisions will be actualized, the item characteristic curves provided in the Study 3 publication are informative.

4.2 Is NMPI usage in military personnel selection worthwhile?

In general, arguments for using personality measures in selection can be anchored in findings of predictive validity studies. While several trait-based personality measures exist (e.g., Ashton & Lee, 2020; Hogan, 2005), FFM tests have received a prominent place in research documenting linkages between personality and job performance (Barrick et al., 2001). In military studies, results have documented that C, E, and ES relate positively to training and occupational performance and provide incremental validity to GMA (Darr, 2011; Hough et al., 1990; McHenry et al., 1990; Salgado, 1998). While the magnitude of predictive validities is typically modest, which some have interpreted as a major weakness (Morgeson et al., 2007a, 2007b), there are nevertheless certain trends.

Usually, it is natural to include personality measures in selection programs due to the content of job analyses and an understanding of job performance wider than that of task performance. Project A in the U.S. Army contributed to these matters by using quite broad criteria (Campbell & Knapp, 2001), and military job analyses typically conclude that it is essential to also select for non-cognitive competencies (Mitchell & Driskill, 1996; Sümer et al., 2001). Indeed, the mainstream personnel psychology categorization of performance into organizational citizenship and counterproductive work behavior, in addition to tasks, can help actualize personality testing (Gonzalez-Mulé et al., 2014; Sackett & Lievens, 2008).

Several *a priori* arguments thus support the use of NMPI testing in military selections, as long as it measures the FFM in a valid and reliable way. However, an important

consideration is the bandwidth debate, questioning whether broad traits or subcomponents are the best predictors for performance (Judge et al., 2013). As the NMPI does not include facets, it is reasonable to assume this could somewhat limit its predictive validity, although this is an empirical question. Study 1's supplementary file shows that the NEO facets did not obtain more evident associations toward the ratings than the domains – although the ratings did not reflect occupational performance per se.

Test use also relates to the "clinical" versus "mechanical" combination of predictors. The NAF has used its FFM measure in a typical clinical fashion since the 1990s, whereby a military psychologist uses the profiles obtained as background information in interviews and other rater-based selection methods. Such usage contrasts with the GMA selection method in the NAF, where cut-off scores serve as initial selection hurdles. For example, in order to attend selections for leadership educations in the NAF, such as the joint officer selection program described in Study 1, one must obtain a minimum score of five on a stanine scale for the GMA tests used at the conscript assessment procedure. Furthermore, in some selections, GMA scores (together with grade point averages from former education) are also used as a variable in a mathematically derived "educational prognosis."

Existing research advocates for a statistical combination of predictors to achieve optimal predictive validity (Kuncel et al., 2013; Meijer et al., 2020). Rules for making decisions about cut-off scores would then be needed if the NMPI is incorporated into a predictor algorithm. However, the lack of predictive validity studies for the NMPI makes it challenging to advocate for such cut-off scores at the time of writing. However, an initial effort to obtain predictive validity estimates might be feasible in the realm of conscription and mandatory military service. The magnitude of the Norwegian conscript assessment is significant: about 20,000 individuals are chosen from an age cohort of approximately 60,000 for further suitability evaluation for general soldier service (Køber et al., 2017). The NMPI may be a predictor at this large-scale selection program, ideally based on empirical investigations of linkages between scale scores and relevant soldier performance criteria. A weighting principle might then be applied, in which scale scores can provide different loadings toward a final suitability evaluation consisting of different cognitive and non-cognitive predictors.

In contrast to the select-in procedure outlined above, one could also apply a select-out model for advocating NMPI use. A select-out procedure is usually concerned with psychopathological conditions (King, 2014). Of particular relevance for a select-out procedure is the trait of N (or ES). Although N is a dimension of general (normal) personality

(McCrae & Costa, 2010), it is clearly associated with health concerns (Tackett & Lahey, 2016). If used in a select-out procedure, NMPI profiles may aid health evaluations in which medical doctors or clinical psychologists (knowledgeable about the FFM) assess a candidate. A feasible procedure for such usage in the conscript assessment arena might be that those who score particularly low on ES are subjected to a more thorough medical examination.

Further regarding select-in procedures, a particular challenge of using the NMPI mechanically is the findings of socially desirable responses. Thus, an individual can obtain a favorable suitability evaluation when, for example, reporting medium to high scores on C and ES while having true low scores on those scales, given that others report more truthfully. Another challenge pertains to acknowledging that people are different in their degree of self-understanding. The *APA Dictionary of Psychology* defines self-understanding as "the attainment of knowledge about and insight into one's characteristics, including attitudes, motives, behavioral tendencies, strengths, and weaknesses" (https://dictionary.apa.org/self-understanding). One's inner "third eye" is, of course, relevant when determining graded agreement with items operationalizing the FFM. As noted earlier, a possible remedy for unwanted response sets might be future research investigating the effect of a social desirability scale and forced choice formats. The latter has been shown to reduce faking, although it cannot eliminate it altogether (Wetzel et al., 2021). However, the problem of differentiated self-understanding will prevail in measures relying on self-perceived personality. One solution can be to measure others' perceptions of a person, although this is a much more effortful enterprise compared to self-report tests.

Considering the challenges of mathematically derived cut-off scores on the NMPI, using test results as background material for identifying the strengths and weaknesses of applicants and candidates can be advocated. Specially trained military psychologists or HR personnel might use a FFM test soundly in a semi-structured or structured interview by problematizing or supporting the selection criteria evaluated (e.g., competencies). By actively anchoring test result interpretation in relevant selection criteria, one may avoid a "whole-person" evaluation, which is tempting when interpreting a FFM profile. After all, a profile provides much information, not least when making hypotheses about a person based on the combination of two or more traits. Naturally, such test use is time-consuming. Therefore, such a practice is undoubtedly most feasible in selection programs that contain fewer candidates than the conscript assessment procedure.

Studies 1 and 2 highlight aspects relevant to using a FFM test non-mechanically in officer and special forces selection. For the joint selection program for officers, a broad job

analysis effort concerning military leadership was initiated in 2014. This effort resulted in formulating the competencies role model, task focus, mental robustness, cooperation, and development, presumably related to successfully employing mission command leadership principles. It was reasonable to assume a differentiated overlap with the personality scales based on the competency contents. For example, ES might be related more to mental robustness than to development, and C more to task focus than to cooperation. However, findings showed that neither the NMPI nor the NEO scales had a correlational pattern pointing to differences in single competencies' personality overlap. In fact, only minor associations toward the OARs themselves were seen (E+ and O-). These results point to a low FFM saturation in the rating practices, arguing for including the NMPI as a predictor at this selection program (i.e., the potential for incremental validity). Future studies may inspect the degree of incremental validity for investigating whether a mechanical combination of predictors is warranted.

For non-mechanical use, the NMPI can be valuable in problematizing or supporting competency ratings. Military psychologists may aid assessors in their judgments of competency levels by discussing personality test results. Nuanced competency ratings may thus be achieved and perhaps counteract biased rating practices, such as those colored by the halo effect and other heuristics (Kahneman, 2011). Based on the study findings, although characterized by low effects, an awareness of whether high scorers on E or O are rated objectively might be important. High E is likely advantageous in military settings (Darr et al., 2018). However, when high scorers of E manifest behaviors hiding low C scores, as problematized in civilian settings (Wiersma & Kappe, 2016), such a trait advantage becomes suboptimal. After all, C has demonstrated predictive validity in military and civilian studies alike (Barrick et al., 2001; Darr, 2011; Hough et al., 1990). With regard to O, while high levels of this trait might result in distractions for learning routines in initial training phases (Darr, 2011), and more so than in civilian ACs (Collins et al., 2003; Hoffman et al., 2015), this trait can be relevant in the military in the long run. One possibility is that O may be positively related to successfully developing and employing mission command leadership principles. A concrete mechanism for such a positive relatedness might be that open-mindedness is advantageous for softening micro-management tendencies and encouraging initiative among subordinates.

Considering the SOF, operators in such units can be categorized as high-risk operational personnel, a broad category often used in the operational psychology literature (e.g., Picano & Roland, 2012). According to another publication by Picano et al. (2017), U.S.

selection programs for such personnel often separate competency-related evaluations and psychological evaluations. They write that, in contrast to competency-related assessments, "psychological evaluations, including interviews and psychological testing (both cognitive and personality), often focus more heavily on broader clinical constructs than on the specific competencies identified, and yield more general or global assessments of candidates' suitability for high-risk operational work" (p. 280).

This approach is recognizable in a Norwegian context, although medically/clinically oriented assessments are perhaps more separated from the psychological. While the FSK and MJK have separate selection systems, a military psychologist supports instructors in these selections with tests of intelligence and the FFM. They may also help observe candidate behavior in AC exercises. Cut-off scores have been used for intelligence tests. In contrast, the 5PFmil 2.0. has not been used in this way. As Johnsen (2017) documented for the Norwegian police special force selection, personality test results in military selections are also treated as one of several information sources for a candidate. As I experienced some years ago, the practice was to provide a whole-person interpretation of candidates near the end of the selection phase to a board deciding to select them in or out. Here, a blend of information was synthesized to make a final decision.

While such a practice places high demands on the selection personnel and increases the risk of biased judgments, a whole-person approach may actually be warranted when considering the complexity of SOFs. Perhaps most relevantly, it is difficult – if not impossible – to formulate an adequate competency system "by the book" for such occupations. SOFs, more so than conventional forces, are expected to be innovators who develop new tactics and techniques, constantly exploring the future of warfare (Berg-Knutsen & Roberts, 2015). Furthermore, while prediction of future performance is essential, it is also particularly relevant to focus on the probability of counterproductive work behaviors. Such behaviors, which are often understood as less grave than those relevant for a security clearance process (Robinson & Bennett, 1995), can still have severe consequences when considering the critical and politically sensitive nature of special forces missions. In effect, when selecting SOF operators, a future competency profile is more uncertain relative to other armed professions (e.g., officers). Johnsen (2017) highlighted that the aiding psychologists should have extensive knowledge about the unit and the content of the total selection procedure. With such an understanding, the whole-person approach may then be targeted at relevant aspects, even though explicit competencies – as those for the officer selection – are lacking.

Study 2's findings may guide personality test usage, as the study participants were successfully selected and employed operators. The average tip-of-the-spear operator, whether from the FSK or MJK, manifested himself as lower on E and A and somewhat higher on ES than non-SOF military personnel. Thus, the specialness of operators was, to a certain degree, manifested in their personalities. The findings of E and A point to a possibility of operators generally being somewhat reserved and skeptical toward strangers. Those with combat-deployment experience scored even lower on A compared to those without such experience. Such a profile might be adaptive when considering such units' security and secrecy requirements (e.g., not revealing classified information). Furthermore, lower than average scores on these traits may fuel a combat mindset, which is believed to be a psychological state providing aggression regulation and power to act (Boe et al., 2020). A combat mindset may be relevant if, for example, direct actions are warranted (e.g., sabotage or eliminating adversaries). That said, scores of E and A were not low per se, which paints a picture of the operator as being able to be socially adept if a diplomatic approach is warranted. Some caution is indicated in concluding that this is an adaptive FFM profile due to different findings from related professions in other countries (Bech et al., 2021; Braun et al., 1994). However, such a profile might guide the Norwegian selection system when used as a rule-of-thumb.

To summarize, using the NMPI-50 measure in military selections is indeed worthwhile. The NMPI can be used to nuance competency ratings at the joint officer selection program, as the FFM and those competencies somewhat align in content. The profile that emerged among SOF operators may serve as a rule-of-thumb for identifying personality strengths for special forces candidates. However, the NMPI, lacking in facets, is somewhat short for the whole-person approach practiced in this selection. Ideally, as recommended in the literature, the NMPI should be used as a variable in a selection system that mechanically combines predictors. Such a combination may not be relevant in the SOF selections due to uncertain competency profiles. However, for the large-scale conscript assessment procedure, this should be feasible in the future, ideally with the obtainment of predictive validity data for determining rules about making decisions for cut-offs and/or predictor weights. For the officer selection program, there is a potential for incremental validity due to scarcity in personality overlap. If relevant predictive validity data can be obtained here as well, a mechanical selection system incorporating the NMPI might be possible.

4.3 Strengths and limitations

An overall strength of this thesis is in providing findings and discussions that are generally relevant for evaluating the measurement quality and practical usage of personality tests in personnel selection. However, a significant limitation concerns the lack of predictive validity data.

The main strength of all three studies concerns the provision of original Norwegian military psychology research. There is little existing FFM research relevant to Studies 1 and 2. Therefore, these studies expand the literature through empirical findings about, respectively, personality overlap with rater-based military selection methods and the characteristics of SOF operators. Using a complete age cohort in Study 3 is also a strength in Norwegian personality research, where the findings replicated the FFM of personality in a large sample. The implementation of item response analysis and measurement invariance analysis, in addition to the factor analyses, provided findings of both theoretical and practical relevance. Furthermore, by obtaining data on motivation for military service, aspects of socially desirable responses on self-report measures could be investigated.

A limitation with respect to Study 1 was that all participants had been preselected in the conscript assessment procedure, thus producing a likely restriction of range in study variables. While the study's purpose was not to generalize to the general population, more prominent associations might have emerged if correcting for range restriction based on population data. However, we investigated whether the standard deviation of variable scores changed in line with progress at the selection program. The unrestricted group was then constituted from those attending the first two days of the program, while the presumed restricted groups were those with the finalized interview and field exercise ratings. Non-existent to minor group differences in standard deviations were observed, thus legitimizing the usage of uncorrected scores for our purposes. Unfortunately, the reliability and validity of the competency ratings were unknown. Of specific concern was the lack of inter-rater reliability studies. A possible positive influence of inter-rater reliability was actualized, however, as HR personnel and military psychologists trained assessors in performing interview and field exercise ratings, though we did not collect data on the extent to which assessors used the interview scoring system and the BARS related to the field exercise.

Study 2's limitations concern first and foremost a possible sampling bias threatening the generalizability of the results. From a scientific perspective, there was a challenge with confidentiality issues about the number of employees in the SOF. From a military standpoint, this was necessary for conducting the research. Therefore, we could not calculate the response

rate and otherwise perform representativeness analyses for the study participants. The sampling bias was consequently unknown. Considering the scarce literature on personality characteristics of SOF personnel, the study findings were nevertheless valuable. Two other limitations were related to the sole usage of the NMPI-80 for FFM measuring and the lack of a comparative population group. Firstly, while the NMPI-80 was preliminarily validated in Study 1, its factor structure and other relevant psychometric investigations have not been published. Some aspects of NMPI-80's reliability and validity therefore remain undocumented. Secondly, a general population sample might have provided a more accurate estimation of SOF uniqueness in personality. However, due to feasibility issues, we could not use other established personality tests with norms, such as the NEO-PI-3. As the Study 1 sample to which the SOF operators were compared was younger than the Study 2 sample, some cautiousness was warranted in interpreting the scores.

While the NMPI-50 was thoroughly inspected for psychometric properties in Study 3, this was based on a distinct age group and in a military test administration situation. Therefore, caution is recommended in generalizing the results to other age groups and other administration contexts. As noted earlier, renewed comprehensive analyses of scale properties are needed to optimize reliable and valid score interpretations for other age groups. Measurement invariance analyses can also be expanded further.

5 Conclusion

Human factors, in addition to equipment and technological aids, are significant for performance in military activities. The complexity and use-of-force potential in conflicts, crises, and wars require that military personnel are capable of mastering service life in an occasionally extraordinary context. Therefore, psychology provides the military with tests of intelligence and personality, the two main domains concerning stable individual differences, in order to increase accuracy in the assessment of human factors. This provision has been especially relevant for personnel assessment and selection, as witnessed in the longstanding history of military psychology.

This thesis presented and discussed military studies concerning the FFM, a validated descriptive taxonomy for operationalizing personality. The FFM has advantages in research and applied settings as it is an economical and easy-to-communicate framework. It has been extensively studied in terms of the linkages between the FFM and military performance and is widely used in military selection programs. In Norway, the newly developed NMPI is envisioned as a feasible tool for supporting military selection decisions. Its psychometric qualities and potential use in mechanical predictor combinations or as background information supporting rater-based selection methods have been discussed. The findings of this short-form measure are overall promising, although it does not provide rich individual personality profiles.

The implications of this thesis for the military realm fall into two categories. Firstly, the very content of this thesis may continue to uphold the relevance of scientific personality psychology in military matters. The FFM, a remarkably agreed-upon framework in the academic community, represents a periodic table of personality that captures essential human factors. The Big Five traits secure a targeted approach toward assessing tendencies relevant for several organizational outcomes. Without this framework, I suggest that the specification and measurement of non-cognitive factors will become more diverse and subject to ideologies that are less dependent on empirical evidence. At worst, the usage of measures producing scores of unknown reliability and validity can harm selection decisions (or other activities in which such measures are used) and negatively impact the very reputation and standing of military psychology. As such, the FFM is a professional approach.

This reasoning does not mean that innovation and renewed models of "will do" factors are not welcome. For example, exciting Norwegian military psychology studies have recently emerged, such as those focusing on character strengths in selection (Bang et al., 2021) and

leadership development (Boe, 2019). Yet, the FFM stands as a particularly well-validated framework related to self-perceived personality, which enables cost-effective measurement. This first implication thus further cements the use of a thoroughly research-based personality framework in the Norwegian Armed Forces.

The second implication relates to the practical use of personality tests. Although some measurement challenges were observed for the NMPI in the conscript assessment procedure, this arena holds the most feasible opportunity for mechanical usage of the measure. By extracting a general factor in addition to the five scales, it is possible to control somewhat for social desirability effects in the responses. The scales of ES and C may be used in a predictor algorithm due to the predictive validity of the scales shown in the general military psychology literature. Where sex differences in scores on C were non-existent, a difference was observed for ES. To secure an adequate number of females in a final selection decision for mandatory soldier service, a cut-off score of ES need not be too high, since more females than males report low ES scores. Ideally, though, local validity studies including relevant criteria can inform service-related cut-off values (i.e., in line with a job analysis approach).

The implication of the findings from the officer selection arena and the SOF relate to advice when using a personality test as background material for rating and evaluating candidates. By informing assessors at the joint officer selection program of candidate personality, more nuanced competency ratings than those witnessed in Study 1 may be seen. This is particularly relevant in evaluating whether those scoring high on E and O are respectively rewarded and punished unfairly in their ratings. More so than in other military selections, the SOF operator selection programs rely on a whole-person approach regarding psychological suitability. Such an approach is demanding and increases the risk of biased judgment. Study 2's findings may guide this process, as trait patterns among successful employees were reported. However, the NMPI is not considered to be ideal for SOF selection, as this short-form measure cannot obtain rich individual profiles.

In closing, I note that we, in the role of selection practitioners, owe it to those who are accepted for a service that the assessment of human factors is as accurate as possible and used soundly. We also owe reasonable and fair explanations to those who are evaluated as psychologically unsuitable. In the field of non-cognitive human factors, the FFM succeeds in providing a validated and pragmatic operationalization that most people understand. Indeed, this personality framework can be used for other purposes as well, such as developmental programs aiming to increase the well-being and performance of employees and leaders. The NMPI is ideal for large-scale selection arenas, contingent on recognizing challenges with self-

reporting, but it needs to be used with caution if nuanced individual profiles are necessary. Lastly, I hope this thesis demonstrates how psychology can continue aiding the military in a scientific and applied way, expanding on a century-old history focusing on human factors for securing peace and the safety of our society.

References

Allport, G. (1931). What is a trait of personality? *The Journal of Abnormal and Social Psychology*, *25*, 368-372. https://doi.org/10.1037/h0075406

Allport, G., & Odbert, H. (1936). Trait-names: A psycho-lexical study. *Psychological monographs*, *47*, i-171. https://doi.org/10.1037/h0093360

Ashton, M. C., & Lee, K. (2020). Objections to the HEXACO model of personality structure - and why those objections fail. *European journal of personality*, *34*, 492-510. https://doi.org/10.1002/per.2242

Atherton, O. E., Donnellan, M. B., & Robins, R. W. (2020). Development of personality across the life span. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (2 ed., pp. 169-182). Cambridge University Press. https://doi.org/10.1017/9781108264822.017

Bang, H., Nilsen, F., Boe, O., Eilertsen, D. E., & Lang-Ree, O. C. (2021). Predicting army cadets´ performance: The role of character strengths, GPA and GMA. *Journal of Military Studies, 0*, 1-15. https://doi.org/10.2478/jms-2021-0016

Barrick, M. R., & Mount, M. K. (2012). Nature and use of personality in selection. In N. Schmitt (Ed.), *The Oxford handbook of personnel assessment and selection* (pp. 225-251). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199732579.013.0011

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. https://doi.org/10.1111/1468-2389.00160

Bech, S. C., Dammeyer, J., & Liu, J. (2021). Changes in personality traits among candidates for special operations forces. *Military Psychology*, *33*, 197-204. https://doi.org/10.1080/08995605.2021.1902178

Berg-Knutsen, E., & Roberts, N. C. (2015). *Strategic design for NORSOF 2025*. http://hdl.handle.net/10945/47444

Birkeland, S. A., Manson, T. M., Kisamore, J. L., Brannick, M. T., & Smith, M. A. (2006). A meta-analytic investigation of job applicant faking on personality measures. *International Journal of Selection and Assessment*, *14*, 317-335. https://doi.org/10.1111/j.1468-2389.2006.00354.x

Block, J. (1995). A contrarian view of the five-factor approach to personality description. *Psychological Bulletin*, *117*, 187-215. https://doi.org/10.1037/0033-2909.117.2.187

Boe, O. (2019). Selecting officers based upon character strenghts: Applications for leadership development. In L. I. Mangussen (Ed.), Disaster, diversity and emergency preparation. (pp. 109-125). IOS Press Ebooks. https://doi.org/10.3233/NHSDP190031

Boe, O., Torgersen, G.-E., & Skoglund, T. H. (2020). Does the Norwegian police need a well-functioning combat mindset? *Frontiers in Psychology*, *11:1075*. https://doi.org/10.3389/fpsyg.2020.01075

Bowles, S. V., & Bartone, P. T. (Eds.). (2017). *Handbook of military psychology: Clinical and organizational practice*. Springer International Publishing. https://doi.org/10.1007/978-3-319-66192-6.

Braun, D. E., Prusaczyk, W. K., Goforth, H. W. J., & Pratt, N. C. (1994). *Personality profiles of U.S. Navy Sea-Air-Land (SEAL) personnel*. Naval Health Research Center.

Campbell, J. P. (1990). An overview of the Army selection and classification project (Project A). *Personnel Psychology*, *43*, 231-239. https://doi.org/10.1111/j.1744-6570.1990.tb01556.x

Campbell, J. P., & Knapp, D. J. (Eds.). (2001). *Exploring the limits in personnel selection and classification*. Lawrence Erlbaum Associates Publishers.

Cattell, H. E. P., & Mead, A. D. (2010). The Sixteen Personality Factor Questionnaire (16PF). In G. J. Boyle, G. Matthews, & D. H. Saklofske (Eds.), *The SAGE handbook of personality theory and assessment: Volume 2 —personality measurement and testing*. SAGE Publications Ltd. https://doi.org/10.4135/9781849200479.n7

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurment invariance. *Structural Equation Modeling*, *9*, 233-255. https://doi.org/10.1207/S15328007SEM0902_5

Cho, E., & Seonghoon, K. (2015). Cronbach´s coefficient Alpha: Well known but poorly understood. *Organizational Research Methods*, *18*, 207-230. https://doi.org/10.1177/1094428114555994

Collins, J. M., Schmidt, F. L., Sanchez–Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, *11*, 17-29. https://doi.org/10.1111/1468-2389.00223

Crawford, M. P. (1970). Military psychology and general psychology. *American Psychologist*, *25*, 328-336. https://doi.org/10.1037/h0029450

Cronbach, L. J., & Meehl, P. E. (1955). Construct validity in psychological tests. *Psychological Bulletin*, *52*, 281-302. https://doi.org/10.1037/h0040957

Darr, W. (2011). Military personality research: A meta-analysis of the Self Description Inventory. *Military Psychology*, *23*, 272-296. https://doi.org/10.1080/08995605.2011.570583

Darr, W., & Catano, V. M. (2016). Determining predictor weights in military selection: An application of dominance analysis. *Military Psychology*, *28*, 193-208. https://doi.org/10.1037/mil0000107

Darr, W. A., Ebel-Lam, A., & Doucet, R. G. (2018). Investigating the extravert advantage in training: Exploring reward sensitivity, training motivation, and self-efficacy as intermediary factors. *Canadian Journal of Behavioural Science*, *50*, 172-184. https://doi.org/10.1037/cbs0000102

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*, 417-440. https://doi.org/10.1146/annurev.ps.41.020190.002221

Driskell, J. E., & Olmstead, B. (1989). Psychology and the military: Research applications and trends. *American Psychologist*, *44*, 43-54. https://doi.org/10.1037/0003-066X.44.1.43

EFPA. (2013). *EFPA review model for the description and evaluation of psychological and educational tests, version 4.2.6* http://assessment.efpa.eu/documents-/

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. https://doi.org/10.1093/oxfordhb/9780195399325.013.0042

Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and Individual Differences*, *13*, 667-673. https://doi.org/10.1016/j.addbeh.2014.01.002

Furnham, A. (2008). *Intelligence and personality at work. Exploring and explaining individual differences at work.* Routledge.

Goldberg, L. R. (1992). The development of markers for the Big-Five factor structure. *Psychological Assessment*, *4*, 26-42. https://doi.org/10.1037/1040-3590.4.1.26

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

Goldberg, L. R., Johnsen, J. A., Eber, H. W., Hogan, R., Asthon, 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. https://doi.org/10.1016/j.jrp.2005.08.007

Gonzalez-Mulé, E., Mount, M. K., & Oh, I.-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology*, *99*, 1222-1243. https://doi.org/10.1037/a0037547

Graziano, W. G., & Tobin, R. M. (2016). Agreeableness and the five factor model. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.17

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

Hansen, I. (2006). *Bidrag til psyologitjenestens historie i Forsvaret fra 1946-2006*. Norwegian Staff College.

Hartmann, E., & Grønnerød, C. (2009). Rorschach variables and Big Five scales as predictors of military training completion: A replication study of the selection of candidates to the naval special forces in Norway. *Journal of Personality Assessment*, *91*, 254-264. https://doi.org/10.1080/00223890902794309

Hartmann, E., Sunde, T., Kristensen, W., & Martinussen, M. (2003). Psychological measures as predictors of military training performance. *Journal of Personality Assessment, 80,* 87-98. https://doi.org/10.1207/S15327752JPA8001_17

Highhouse, S. (2008). Stubborn reliance on intuition and subjectivity in employee selection. *Industrial and Organizational Psychology, 1*, 333-342. https://doi.org/ 10.1111/j.1754-9434.2008.00058.x

Hoffman, B. J., Kennedy, C. L., Lopilato, A. C., Monahan, E. L., & Lance, C. E. (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, *100*, 1143-1168. https://doi.org/10.1037/a0038707

Hogan, R. (2005). In defense of personality measurement: New wine for old whiners. *Human Performance*, *18*, 331-341. https://doi.org/10.1207/s15327043hup1804_1

Hooper, D., Coughlan, J., & Mullen, M. R. (2007). Structural equation modeling: Guidelines for determining model fit. *Electronic Journal on Business Research Methods*, *6*, 53-60.

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

Hough, L. M., Dunnette, M. D., Eaton, H. K., Kamp, J. D., & McCloy, R. A. (1990). Criterion-related validities of personality constructs and the effect of response distortion on those validities. *Journal of Applied Psychology*, *75*, 581-595. https://doi.org/10.1037/0021-9010.75.5.58

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

Hughes, D. J., & Batey, M. (2017). Using personality questionnaires for selection. In H. W. Goldstein, E. D. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell handbook of the psychology of recruitment, selection and employee retention* (pp. 151-181). John Wiley & Sons Ltd. https://doi.org/10.1002/9781118972472.ch8

Jackson, J. J., & Roberts, B. W. (2015). Conscientiousness. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.18

Johansen, R. B., Fosse, T. H., & Boe, O. (Eds.). (2019). *Militær ledelse*. Fagbokforlaget.

John, O. P., & Robins, R. W. (Eds.). (2021). *Handbook of personality: Theory and research* (4 ed.). Guilford Press.

Johnsen, B. H. (2017). Selection of police special operations officers: The role of the psychologist. In S. V. Bowles & P. T. Bartone (Eds.), *Handbook of military psychology: Clinical and organizational practice.* Springer. https://doi.org//10.1007/978-3-319-66192-6_18

Judge, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & Crawford, E. R. (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*, 875-925. https://doi.org/10.1037/a0033901

Judge, T. A., & Zapata, C. P. (2015). The person-situation debate revisited: Effect of situation strenght and trait activation on the validity of the big five personality traits in predicting job performance. *Academy of Management Journal*, *58*, 1149-1170. https://doi.org/10.5465/amj.2010.0837

Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus and Giroux.

Kaplan, R. M., & Saccuzo, D. P. (2009). *Psychological assessment and theory: Creating and using psychological tests* (8th ed.). Wadsworth Cengage Learning.

King, R. E. (2014). Personality (and psychopathology) assessment in the selection of pilots. *The International Journal of Aviation Psychology, 24:1*, 61-73. https://doi.org/10.1080/10508414.2014.860844

Koonce, J. M. (1984). A brief history of aviation psychology. *Human Factors*, *26*, 499-508. https://doi.org/10.1177/001872088402600502

Kragh, U. (1960). The Defence Mechanism Test: A new method for diagnosis and personnel selection. *Journal of Applied Psychology*, *44*, 303-309. https://doi.org/10.1037/h0046004

Kuncel, N. R., Klieger, D. M., Connelly, B. S., & Ones, D. S. (2013). Mechanical versus clinical data combination in selection and admissions decisions: A meta-analysis. *Journal of Applied Psychology*, *98*, 1060-1072. https://doi.org/10.1037/a0034156

Källmen, H., Wennberg, P., & Bergman, H. (2010). Psychometric properties and norm data of the Swedish version of the NEO-PI-R. *Nordic Journal of Psychiatry*, *65*, 311-314. https://doi.org/10.3109/08039488.2010.545433

Køber, 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*, 560-569. https://doi.org/10.1037/mil0000192

Laurence, J. H., & Matthews, M. D. (Eds.). (2012). *The Oxford handbook of military psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399325.001.0001.

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

Magnusson. D. (1961). Test teori. Almqvist & Wiksell Förlag AB.

Maheshwari, N., Kumar, V. V., & Singh, N. P. (2016). Military psychology: The centurion phenomenon. In N. Maheshwari & V. V. Kumar (Eds.), *Military psychology: Concepts, trends and interventions*. SAGE Publications Pvt. Ltd. https://doi.org/10.4135/9789353885854.n2

Martinsen, Ø. L. (2007). *Norwegian translation of the NEO PI*. Hogrefe.

Martinussen, M. (2005). Seleksjon av flygere og flygeledere. *Tidsskrift for Norsk Psykologforening*, *42*, 291-300.

Martinussen, M., Richardsen, A. M., & Vårum, H. W. (2003). Validation of an ipsative personality measure (DISCUS). *Scandinavian Journal of Psychology*, 411-416. https://doi.org/10.1111/1467-9450.00253

Martinussen, M., & Torjussen, T. M. (1998). Pilot selection in the Norweigan Armed Force: A validation and meta-analysis of the test battery. *The International Journal of Aviation Psychology*, *8*, 33-45. https://doi.org/10.1207/s15327108ijap0801_2

McCrae, R. R. (2020). The five-factor model of personality: Consensus and controversy. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (2 ed., pp. 129-141). Cambridge University Press. https://doi.org/10.1017/9781108264822.013

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

McCrae, R. R., Costa, P. T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, *84*(3), 261-270. https://doi.org/10.1207/s15327752jpa8403_05

McCrae, R. R., Terracciano, A., & 79_members_of_the_Personality_Profiles_of_Cultures_Project. (2005). Personality profiles of cultures: Aggregate personality traits. *Journal of Personality and Social Psychology*, *89*, 407-425. https://doi.org/10.1037/0022-3514.89.3.407

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. A., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology*, *43*, 335-354. https://doi.org/10.1111/j.1744-6570.1990.tb01562.x

Meijer, R. R., Neumann, M., Hemker, B. T., & Niessen, A. S. M. (2020). A tutorial on mechanical decision-making for personnel and educational selection. *Frontiers in Psychology*, *10:3002*. https://doi.org/10.3389/fpsyg.2019.03002

Melton, A. W. (1957). Military psychology in the United States of America. *American Psychologist*, *12*, 740-746. https://doi.org/10.1037/h0047927

Mischel, W. (1968). *Personality and assessment*. John Wiley & Sons Inc.

Mischel, W. (2004). Toward an integrative science of the person. *Annual Review of Psychology*, *55*, 1-22. https://doi.org/10.1146/annurev.psych.55.042902.130709

Mitchell, J. L., & Driskill, W. E. (1996). Military job analysis: A historical perspective. *Military Psychology*, *8*, 119-142. https://doi.org/10.1207/s15327876mp0803_1

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*, 1029-1049. https://doi.org/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*, 683-729. https://doi.org/10.1111/j.1744-6570.2007.00089.x

Murphy, K. R., & Davidshofer, C. O. (2001). *Psychological testing: Principles and applications* (5th ed.). Prentice-Hall, Inc.

NAF. (2021). *Lederkandidatstudien*. https://www.forsvaret.no/forskning/prosjekter/lederkandidatstudien

NATO. (2012). *Psychological and physiological selection of military special operations forces Personnel*. https://www.sto.nato.int/publications/Pages/default.aspx

Nye, C. D., White, L. A., Drasgow, F., Prasad, J., Chernyshenko, O. S., & Stark, S. (2020). Examining personality for the selection and classification of soldiers: Validity and differential validity across jobs. *Military Psychology*, *32*, 60-70. https://doi.org/10.1080/08995605.2019.1652482

Olsen, T., & Thormodsen, M. (2014). *Forging Norwegian special operation forces*. https://calhoun.nps.edu/handle/10945/42699

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

Page, G. D. (1996). Clinical psychology in the military: Developments and issues. *Clinical Psychology Review*, *16*, 383-396. https://doi.org/10.1016/0272-7358(96)00020-7

Picano, J. J., & Roland, R. R. (2012). Assessing psychological suitability for high-risk military jobs. In J. H. Laurence & M. D. Matthews (Eds.), *The Oxford handbook of military psychology*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195399325.013.0056

Picano, J. J., Roland, R. R., Williams, T, & Bartone, P. T. (2017). Assessment of elite operational personnel. In S. V. Bowles & P. T. Bartone (Eds.), *Handbook of military psychology: Clinical and organizational practice*. Springer. https://doi-org.mime.uit.no/10.1007/978-3-319-66192-6

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. https://doi.org/10.1093/oxfordhb/9780199928309.013.0008

Psykologen. (1964). *Forsvarets psykologitjeneste*. Bilag til Psykologen Oktober 1964.

Reynolds, C. R., Altmann, R. A., & Allen, D. N. (2021). *Mastering modern psychological testing: Theory and methods* (2nd ed.). Springer Nature Switzerland AG. https://doi.org/10.1007/978-3-030-59455-8

Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, *2*, 313-345. https://doi.org/10.1111/j.1745-6916.2007.00047.x

Roberts, B. W., & Mroczek, D. K. (2008). Personality trait change in adulthood. *Current Directions in Psychological Science*, *17*, 31-35. https://doi.org/10.1111/j.1467-8721.2008.00543.x

Robinson, S. L., & Bennett, R. J. (1995). A typology of deviant workplace behaviors: A multidimensional scaling study. *Academy of Management Journal, 38*, 555–572. https://doi.org/10.5465/256693

Roth, P. L., Van Iddekinge, C. H., Huffcutt, A. I., Eidson Jr, C. E., & Schmit, M. J. (2005). Personality saturation in structured interviews. *International Journal of Selection and Assessment*, *13*, 261-273. https://doi.org/10.1111/j.1468-2389.2005.00323.x

Sackett, P. R., & Lievens, F. (2008). Personnel selection. *Annual Review of Psychology*, *59*, 419-450. https://doi.org/10.1146/annurev.psych.59.103006.093716

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

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, *11*, 299-324. https://doi.org/10.1080/13594320244000184

Sanchez-Roige, S, Gray, J. C., MacKillop, J. Chen, C-H., & Palmer, A.A. (2017). The genetics of human personality. *Genes, Brain, and Behavior, 17*, 1-13. https://doi.org/10.1111/gbb.12439

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*, 262-274. https://doi.org/10.1037/0033-2909.124.2.262

Schultz, D. P., & Schultz, S. E. (1996). *A History of modern psychology* (6 ed.). Harcourt Brace College Publishers.

Sells, S. B. (1955). Development of a personaity test battery for psychiatric screening of flying personnel. *Journal of Aviation Medicine, 26*, 35-45.

Simms, L., Williams, T. F., & Simms, E. N. (2016). Assessment of the five factor model. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.28

Skoglund, T. H., Martinussen, M., & Lang-Ree, O. C. (2014). Papir vs PC. *Tidsskrift for Norsk Psykologforening, 51*, 450-452.

Spulak, R. G. J. (2009). *A theory of special operations: The origin, qualities, and use of SOF*. http://www.ifpa.org/confrncNworkshp/confrncNworkshpPages/nscc_march2010/PDF3.pdf

Standards. (2014). *The standards for educational and psychological testing*. American Educational Research Association; American Psychological Association; National Council on Measurement in Education.

Stark, S., Chernyshenko, O. S., Drasgow, F., Nye, C. D., White, L. A., Heffner, T., & Farmer, W. L. (2014). From ABLE to TAPAS: A new generation of personality tests to support military selection and classification decisions. *Military Psychology, 26*, 153-164. https://doi.org/10.1037/mil0000044

Statacorp. (2017). *StataCorp. 2019. Stata statistical software: Release 16*. Statacorp LLC.

Streiner, D. L. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*. https://doi.org/10.1207/S15327752JPA8001_18

Strickhouser, J. E., Zell, E., & Krizan, Z. (2017). Does personality predict health and well-being? A metasyntesis. *Health Psychology, 36*, 797-810. https://doi.org/10.1037/hea0000475

Staal, M. A., & Stephenson, J. A. (2013). Operational psychology post-9/11: A decade of evolution. *Military Psychology, 25*, 93-104. https://doi.org/10.1037/h0094951

Sutin, A. R. (2015). Openness. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.16

Sümer, H. C., Sümer, N., Demirutku, K., & Cifci, O. S. (2001). Using a personality-oriented job analysis to indentify attributes to be assessed in officer selection. *Military Psychology, 13*, 129-146. https://doi.org/10.1207/S15327876MP1303_1

Sørlie, H. O., Hetland, J., Dysvik, A., Fosse, T. H., & Martinsen, Ø. L. (2020). Person-organization fit in a military selection context. *Military Psychology, 32*, 237-246. https://doi.org/10.1080/08995605.2020.1724752

Tackett, J. L., & Lahey, B. B. (2016). Neuroticism. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.14

Tett, R. P., & Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *Journal of Applied Psychology, 88*, 500-517. https://doi.org/10.1037/0021-9010.88.3.50

Tett, R. P., & Christiansen, N. D. (2007). Personality tests at the crossroads: A response to Morgeson, Campion, Dipboye, Hollenbeck, Murphy, and Schmitt (2007). *Personnel Psychology, 60*, 967-993. https://doi.org/10.1111/j.1744-6570.2007.00098.x

Thomas, J. L., Dickson, M. W., & Bliese, P. D. (2001). Values predicting leader performance in the U.S. Army reserve officer training corps assessment center: Evidence for a personality-mediated model. *The Leadership Quarterly, 12*, 181-196. https://doi.org/10.1016/S1048-9843(01)00071-6

Thoresen, C. J., Bradley, J. C., Bliese, P. D., & Thoresen, J. D. (2004). The Big Five personality traits and individual job performance growth trajectories in maintenance and transitional job stages. *Journal of Applied Psychology, 89*, 835-853. https://doi.org/10.1037/0021-9010.89.5.83

Torjussen, T. M., & Hansen, I. (1999). Forsvaret, best i test? *Tidsskrift for Norsk Psykologforening, 36*, 772-779.

Tupes, E., & Christal, R. E. (1992). Recurrent personality factors based on trait ratings. *Journal of Personality*, *60*, 225-251. https://doi.org/10.1111/j.1467-6494.1992.tb00973.x

Ubostad, B., & Ellertsen, B. (1999). MMPI, MMPI-2 og MMPI-A: Norske versjoner av et utbredt klinisk kartleggingsinstrument. *Tidsskrift for Norsk Psykologforening*, *36*, 730-735.

Van der Linden, D., Dunkel, C. S., & Wu, P. (2021) Is there a meaningful general factor of personality? *The Spanish Journal of Psychology, 24.* https://doi.org/10.1017/SJP.2021.2

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substative and error influence. *Journal of Applied Psychology*, *90*, 108-131. https://doi.org/10.1037/0021-9010.90.1.108

Weiss, A., & Deary, I. J. (2020). The trait approach. In P. J. Corr & G. Matthews (Eds.), *The Cambridge handbook of personality psychology* (2 ed., pp. 31-43). Cambridge University Press. https://doi.org/10.1017/9781108264822.005

Wetzel, E., Frick, S., & Brown, A. (2021). Does mulitdimensional forced-choice prevent faking? Comparing the susceptibility of the multidimensional forced-choice format and the rating scale format to faking. *Psychological Assessment*, *33*, 156-170. https://doi.org/10.1037/pas0000971

White, L. A., Young, M. C., & Rumsey, M. G. (2001). ABLE implementation issues and related research. In J. P. Campbell & D. J. Knapp (Eds.), *Exploring the limits in personnel selection and classification*. Lawrence Erlbaum Associates Publishers.

Widiger, T. A. (Ed.). (2015). *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.001.0001.

Wiersma, U., & Kappe, R. (2016). Selecting for extroversion but rewarding for conscientiousness. *European Journal of Work and Organizational Psychology*, *26*, 314-323. https://doi.org/10.1080/1359432X.2016.1266340

Wilt, J., & Revelle, W. (2016). Extraversion. In T. A. Widiger (Ed.), *The Oxford handbook of the five factor model*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199352487.013.15

Ziegler, M., & Buehner, M. (2008). Modeling social desirable responding and its effects. *Educational and Psychological Measurement*, *69*, 548-565. https://doi.org/10.1177/0013164408324469

Appendix A

NMPI-80 items (*reversed; items in bold were continued in the NMPI-50)

Openness to experience	Emotional Stability	Extroversion
Am a down to earth person*	**Worry a lot***	**Often keep in the background***
Am not so concerned of my feelings*	**Am easily stressed***	**Do not like attention***
Do prefer stable routines*	**Become nervous easily***	**Am introverted***
Am open and tolerant of others´ way of life	**Am afraid of many of things***	Thrive best when alone*
Respect others´ views of right and wrong	**Often feel tense***	**Like to be the center of social attention**
Am fascinated by patterns and colors	Often doubt myself*	Like to be the center of attention
Think art is important	**Often feel guilty***	**Often start conversations**
Have a vivid imagination	**Am often sad***	Make friends easily
Am good at creative thinking	Am scared easily*	Talk to new people
Know many words and concepts	**Am often afraid of making a fool of myself***	Like to be where people gather
Notice beautiful objects	**Often feel that others are better than me***	**Like to lead others**
Generates many good ideas	My mood changes fast*	Am an energetic person
Am curious about other cultures	**Often think about what others think of me***	**Likes to decide**
Ask questions that no one else asks	Get angry easily	**Am dominant and self-confident**
Am curious and like to learn new things	Am for the most part quite relaxed	Like to compete
Often seek new experiences	-	Am action oriented
Understand things quickly	-	Like to have influence

Conscientiousness	Agreeableness	Anchors
Have a lot of clutter around me*	Keep my distance to other people*	1. Strongly disagree
Often waste my time*	Rarely trust others*	2.
Do things half-way*	Am though to get to know*	3.
Am hardworking and like to get things done	Am concerned of others well-being	4.
Have clear goals and work systematically to achieve them	**Often feel compassion for others**	5.
Am good at putting things in order	**Am percieved as warm and friendly**	6.
Am a punctual person	Think the best of people	7. Strongly agree
Completes my duties right away	**Am helpful towards others**	-
Always meet prepared	**Like to helping others**	-
Am detail oriented	**Take time for others**	-
Always keep appointments	**Am interested in other people**	-
Have high ambitions	**Easily grow fond of other people**	-
Reach my goals	**Have something nice to say about everyone**	-
Always do something	Like to cooperate with others	-
-	**Am good at understanding the needs of others**	-
-	Am usually polite	-

Appendix B

NMPI-50 items (*reversed; items in bold are not part of the NMPI-80)

Openness to experience		Emotional Stability		Extroversion	
o1	Notice beautiful objects	n1	Am easily stressed*	e1	**View myself as an outgoing person**
o2	**Wish to engage in self-development**	n2	Worry a lot*	e2	**Frequently becomes the leader of groups**
o3	**Spend time reflecting on life**	n3	Am often afraid of making a fool of myself*	e3	Like to be the center of social attention
o4	Am fascinated by patterns and colors	n4	Often think about what others think of me*	e4	Do not like attention*
o5	Have a vivid imagination	n5	Often feel guilty*	e5	Likes to decide
o6	Generates many good ideas	n6	Am often sad*	e6	Often start conversations
o7	Know many words and concepts	n7	Become nervous easily*	e7	Often keep in the background*
o8	Think art is important	n8	Often feel that others are better than me*	e8	Am dominant and self-confident
o9	Am good at creative thinking	n9	Often feel tense*	e9	Am introverted*
o10	Am curious about other cultures	n10	Am afraid of many things*	e10	Like to lead others

Conscientiousness		Agreeableness		Anchors
c1	Always keep appointments	a1	Am good at understanding the needs of others	1. Strongly disagree
c2	Reach my goals	a2	Am perceived as warm and friendly	2.
c3	Am detail oriented	a3	Have something nice to say about everyone	3.
c4	Am a punctual person	a4	Take time for others	4.
c5	Am good at putting things in order	a5	Am interested in other people	5.
c6	Completes my duties right away	a6	Easily grow fond of other people	6.
c7	Always meet prepared	a7	Am helpful to others	7. Strongly agree
c8	Am hardworking and like to get things done	a8	Often feel compassion for others	_
c9	Have a lot of clutter around me*	a9	Like helping others	_
c10	Have clear goals and work systematically to achieve them	a10	**Trust others**	_

Study 1

SCANDINAVIAN JOURNAL OF WORK AND ORGANIZATIONAL PSYCHOLOGY

ORIGINAL ARTICLE

Candidate Personality Traits Associated with Ratings in a Military Officer Selection Setting

Tom H. Skoglund*, Thomas H. Fosse†, Ole Christian Lang-Ree‡, Øyvind L. Martinsen§ and Monica Martinussen*

While studies of the relationship between applicant personality and selection assessment ratings are widespread within personnel psychology, studies on military samples are scarcer. The current study aimed to investigate the associations between the Five-Factor Model of personality and ratings in (1) a military selection interview and (2) a field selection exercise simulating a war-like scenario. Research participants were candidates attending a selection program for military officer schools in the Norwegian Armed Forces (N = 901). The study used the NEO-PI-3 for measuring personality, and a shorter military personality measure (NMPI) was added for purposes of convergent test validation. Military selection officers rated candidates based on mission command leadership competencies. After controlling for sex, age, and general mental ability, results demonstrated small predictive effects of extraversion (+) and openness (–) toward mean competency ratings in both the interview and the field selection exercise. Furthermore, the selection officers used somewhat global evaluations of candidates in their ratings due to high correlations between competencies. Based on the findings, we suggest that personality testing may be useful in the selection process and contribute to a more nuanced rating of competencies.

Keywords: military; personality; Five-Factor Model; personnel selection; assessment center; interview

A primary emphasis in personnel selection research identifies determinants of future work performance, where tests of general mental ability and personality traits, employment interviews, and work sample tests are examples of established and valid predictors (Hunter & Schmidt, 1998; Schmidt et al., 2016). However, some selection methods, such as work sample tests (e.g., assessment center exercises) and interviews, are more complex relative to psychometric tests in terms of measurement methodology. While results from standardized psychometric tests generally are dependent on item responses from the test-takers themselves, assessment ratings obtained from interviews and assessment center exercises rely on how a candidate is perceived and interpreted by others. Accordingly, psychometric tests often provide an operationalization of well-defined constructs. In contrast, human judgment produces assessment ratings where the constructs measured are less well defined (Roth et al., 2005). Due to this essential difference, research investigating the overlap between test results and ratings from assessors

is vital for an increased psychological understanding of complex predictors.

While several meta-analyses have examined if candidate ratings obtained in employment interviews and assessment center exercises are associated with cognitive abilities and personality traits (Berry et al., 2007; Collins et al., 2003; Salgado & Moscoso, 2002; Hoffman et al., 2015; Huffcutt et al., 2001), military studies on this matter are scant and often limited to the trait of extraversion (e.g., Darr et al., 2018; Thomas et al., 2001). Due to likely differences in civilian and military selection programs, such as job competencies rated and assessment center exercises employed, there is a need for further military studies. Detailed knowledge of military predictors is relevant for optimizing selection programs, which is important considering that military organizations often recruit candidates to life-long careers where employees might participate in high-risk activities and operations (Campbell et al., 2010; Rumsey & Arabian, 2014). The present study aimed at providing such knowledge by applying the well-established Five-Factor Model (FFM) for studying if candidate personality traits were associated with ratings in a military interview and a field selection exercise at an officer selection program.

Personality traits and interviews

Although studies investigating the predictive validity of military selection interviews exist (e.g., Darr & Catano,

* UiT The Arctic University of Norway, NO

† Norwegian Defence University College, NO

‡ Norwegian Armed Forces Joint Medical Services, NO

§ BI-Norwegian Business School, NO

Corresponding author: Tom H. Skoglund (tomhskoglund@gmail.com)

2016; Køber et al., 2017), there is a lack of published military studies relevant to the present investigation. However, the civilian personnel selection literature reports that the FFM[1] personality traits, to some degree, are embedded in interview ratings. The meta-analysis by Salgado and Moscoso (2002) documented that observed mean correlations ranged between $r = .12–.17$ between traits and ratings obtained in conventional interviews. Regarding behavioral interviews, where structure and questions are pre-planned in more detail relative to conventional interviews, smaller correlations were found. Here, the strongest correlations were mean $r = .10$ for extraversion and mean $r = .08$ for conscientiousness. Those findings were in line with results from another meta-analysis that focused exclusively on structured interviews (Roth et al., 2005).

Other authors have tried to explain the associations between extraversion, conscientiousness, and interview ratings. For instance, some have hypothesized that these traits fuel advantageous self-efficacy mechanisms, meaning that the likelihood of interview success is related to interview self-efficacy (Tay et al., 2006). A more recent study by Wiersma and Kappe (2016) offered a somewhat different explanation by noting that extraversion and conscientiousness may be considered as incentive-enhancing preferences, where the former trait leads to such characteristics as assertiveness, decisiveness, and ambitiousness, and the latter to intrinsic motivation and a striving to perform above average. These authors claimed that extraversion is typically visible in interviews and, therefore, more easily assessed relative to conscientiousness, which is perhaps intuitively compelling. This explanation is supported by the findings of Salgado and Moscoso (2002) for conventional interviews, where extraversion had a higher association with ratings compared with conscientiousness. However, in structured interviews, extraversion seems to have a somewhat lower association relative to conscientiousness (Roth et al., 2005; Salgado & Moscoso, 2002).

Personality traits and assessment centers

The field selection exercise investigated in the present study was quite different from a standard civilian assessment center (AC), both in duration and content. It lasted seven days and nights in the outdoors (i.e., the field), where candidates participated in a war-like scenario. Nevertheless, in terms of measurement methodology, the field selection exercise mirrors an AC as the candidates were rated by assessors when engaging in behavioral simulation exercises (International Taskforce on Assessment Center Guidelines, 2015). The civilian research concerning AC construct embeddedness is voluminous (see e.g., Hoffman et al. (2015) and Thornton and Gibbons (2009) for reviews). For the present study, research focusing on personality and AC overall assessment ratings (OARs) and studies using military samples are of primary interest. One meta-analysis based on civilian samples reported quite high operational validities corrected for unreliability in criteria for the FFM personality traits (.16–.47), where extraversion had the largest validity (Collins et al., 2003).

While conscientiousness was not investigated in this study, the authors also found a high operational validity for cognitive ability (.65). On the other hand, more modest results were reported in the review and meta-analysis by Hoffman et al. (2015), also focusing on civilian sample—where several operational validities were less than .10 for the FFM traits.

Paralleling the results of large-scale meta-analyses using civilian samples (e.g., Barrick et al., 2001), meta-analyses using military samples have also detected that the FFM relates to job performance, where especially conscientiousness has demonstrated predictive validity (Darr, 2011; Salgado, 1998). In the few personality studies regarding military ACs and short-duration training performance, however, extraversion has received the most attention in research. In a five-week training and evaluation course for the U.S. Army Reserve Officer Corps (Thomas et al., 2001), it was found that extraversion was positively related to leadership ratings among approximately 800 cadets ($r = .14$). Darr et al. (2018) argued that extraversion might be advantageous in basic military training, as such contexts are typically collective, allowing participants to interact with and lead others. In their study of 251 candidates undergoing basic officer training in a 15-week course, Darr et al. (2018) found that the dominance aspect of extraversion was positively related to performance ($r = .16$). An investigation of 60 junior officers completing a five-week course measured all FFM traits (Calleja et al., 2019) and found that conscientiousness was related to "planning performance" ($r = .27$).

Three Norwegian studies relevant to the present study aim have used samples from the same selection program as the current investigation (Hystad et al., 2011; Martinsen et al., 2020; Sørlie et al., 2020). Of those studies, Sørlie et al. (2020) included the same sample used in the present study. While Sørlie et al. (2020) investigated the predictive validity of a person-organization fit measure, the authors reported that extraversion and openness had minor isolated predictive impacts ($\beta = .17$ and $-.10$) on the field selection exercise OAR when other variables were included in the regression model. We will expand on those results in the present study by including more nuanced field selection exercise ratings and also including the interview ratings. Hystad et al. (2011) and Martinsen et al. (2020) used data obtained some years earlier at the same selection program for their studies. The first study documented some predictive validity of dispositional hardiness toward final admission decisions and is thus more indirectly relevant as neuroticism is frequently found to be negatively associated with resilience variables (Oshio et al., 2018). The second study (Martinsen et al., 2020) also focused on final admission decisions and found that those offered officer training had lower scores on neuroticism and higher scores on extraversion and conscientiousness relative to those not selected.

Study aim and hypotheses

The purpose of the present study was to explore if candidate personality traits were associated with ratings in a selection program for military officers. Selection officers

rated candidates in a competency-based interview and a field selection exercise simulating a war-like scenario. We measured the FFM personality traits with the established NEO-PI-3 (McCrae et al., 2005). A shorter military FFM test was added (the Norwegian Military Personality Inventory; NMPI) for obtaining construct validity estimates for the less comprehensive NMPI. Additionally, it was of interest to investigate if this test would produce the same results as the NEO, considering that the NMPI was developed for the military. As general research has shown advantages for contextualized personality measures with respect to predictive validity (e.g., Shaffer & Postlethwaite, 2012), the NMPI may be a promising tool for military selection. Due to the impact of cognitive ability on performance (Schmidt & Hunter, 1998; Schmidt & Hunter, 2004), including in studies of ratings in employment interviews (Berry et al., 2007) and AC methods (Collins et al., 2003), scores on general mental ability (GMA) were used for purposes of statistical control.

At the selection program, the criteria upon which candidates were rated are formulated as military leadership competencies. The competencies rated are believed to be in line with general research on effective leadership (Yukl, 2012) and individual prerequisites for successful development into a mission command leader—the espoused leadership philosophy of the Norwegian Armed Forces (Defence Staff Norway, 2012). The competencies, and the gist of their content, are *role model*, acts in line with NAF´s core values, is open to feedback, shows integrity; *task focus*, takes the initiative, works systematically toward goals, prioritizes adequately; *mental robustness*, can cope with high demands and stressful life events, is emotionally stable, and adapts to uncertain circumstances; *cooperation*, gains trust from others, communicates efficiently, delegates, and supports others; and *development*, stimulates autonomy in others and encourages reflection, original thinking, and self-development in others.

Hypotheses concerning personality and interview ratings

Due to the lack of military studies, civilian findings formed the basis for hypothesis development for the interview. Small positive associations between extraversion and conscientiousness on the one hand and the interview OAR on the other, in line with the findings of Salgado and Moscoso (2002) and Roth et al. (2005), were expected. The general arguments put forth by Tay et al. (2006) and Wiersma and Kappe (2016) regarding the relevancy of these traits for interviews also supported this expectation. The following hypothesis was thus formulated:

> H_1: Extraversion and conscientiousness will show statistically significant positive associations with the interview OAR—after controlling for age, sex, and GMA.

In the interview, the three competencies of role model, mental robustness, and development were rated. The following hypothesis was formulated based on content similarities between personality traits and the competencies:

> H_2: Conscientiousness will be positively associated with role model ratings, neuroticism negatively with mental robustness, and openness positively with development.

Hypotheses concerning personality and field selection exercise ratings

Expected findings would necessarily parallel the results from Sørlie et al. (2020), considering that those authors used parts of the same data set as the present study (the field selection exercise OAR and the NEO). However, as the current study also included the NMPI, we expanded the hypothesis development pertaining to the field selection exercise. Although acknowledging that civilian ACs and military field selection exercises are equivalent in terms of measurement methodology, hypothesis development was clouded by the different results reported in meta-analyses of the former (e.g., Collins et al., 2013; Hoffman et al., 2015). Furthermore, likely differences in duration and content between civilian ACs and military field selection exercises limited hypothesis development. Based on the military studies of Thomas et al. (2001), Darr et al. (2018), and Sørlie et al. (2020), however, we expected a positive association between extraversion and the field selection exercise OAR—and also a negative association with respect to openness based on the findings from the latter study. It was also reasonable to expect a positive association between conscientiousness and the OAR considering the predictive validity of this trait in military studies (Calleja et al., 2019; Darr, 2011; Fosse et al., 2015; Martinsen et al., 2020). Finally, based on the harsh elements of the field selection exercise, together with the findings of Martinsen et al. (2020) and Hystad et al. (2011), it was reasonable to expect that neuroticism would be negatively associated with the OAR.

> H_3: Extraversion and conscientiousness will show statistically significant positive associations with the field selection exercise OAR; whereas, neuroticism and openness will show statistically significant negative associations—after controlling for age, sex, and GMA.

In the field selection exercise, all five competencies were rated. Expectations in terms of competency-level associations had parallels with those for the interview. Additonally, we hypothesized that conscientiousness might be positively associated with task focus, and extraversion and agreebleness positively with cooperation, due to content similarities.

> H_4: Neuroticism will be negatively associated with mental robustness ratings, extraversion and agreeableness positively with cooperation, openness positively with development, and conscientiousness positively with role model and task focus.

Our study can contribute to the selection literature in several ways. First, it may increase the understanding

of complex military predictors and provide results of similarities or differences relative to civilian findings. Second, the study may be valuable for evaluating the potential usefulness of incorporating a personality test in selection programs that use military interviews and field selection exercises. Third, as the present study used both OARs and ratings of specific military leadership competencies, study findings can uncover if these different competencies are unequal in personality overlap. Those findings may be of interest as job competencies are often less precisely defined and operationalized compared to psychological constructs (Furnham, 2008).

Method
Participants and procedure
All participants in the present study were candidates attending a selection program for basic officer schools in the summer of 2016. The selection program lasted two weeks and selected officer education applicants for either the Army, Navy, or Air Force. Initially, there were 1287 candidates attending. The personality data were collected during the first two days, where candidates were introduced to the research project in a classroom setting. The final sample consenting to participate was $N = 901$, which included all candidates that attended the classroom brief. Accordingly, 386 candidates did not participate in the classrooms due to early termination at the selection program. Some may have chosen not to attend the classroom brief while continuing the selection program, but we did not register this number. Thus, the response rate was 70%, calculated from the total number of registered candidates. There were 207 women (23%) and 694 men (77%) in the final sample, and the age range was 18–34 years ($M = 19.6$, $SD = 1.86$).

This study was part of a larger research program at the Norwegian Defence University College: "The Leadership Candidate Study" (NAF, 2020), which is approved by the Norwegian Centre for Research Data. Acknowledging ethical challenges related to informed consent when obtaining data at military selection areas (e.g., obedience to authority and conformity pressure), we thoroughly informed candidates of the study purpose. Candidates were also informed that the decision to participate would not affect selection decisions and that the personality data would not be registered in their military records. Furthermore, we evaluated that the measures used were unobtrusive by measuring normal personality traits in which foreseeable adverse psychological consequences of responding were unlikely.

During Week 1, in which the interviews took place, 91 candidates from the initial research-participant pool left the selection program for various reasons (e.g., self-choice, medical conditions, failed physical tests), resulting in 810 participants with complete interview ratings. The field selection exercise was carried through during the second week. A further reduction in participants was witnessed, primarily due to self-choice, where candidates left the program during the first two days of the field selection exercise; in addition, some participants had already been selected out due to unsuccessful interview ratings. The final number of participants with complete field selection

exercise ratings was 551. There was no systematic dropout with respect to sex or age.

Measures
General mental ability
All participants had, at age 17, passed the conscript assessment procedure in Norway, undergoing GMA testing. Since the 1950s, the Norwegian Armed Forces has used a GMA test that includes three subtests measuring reasoning, numerical, and verbal abilities (Køber et al., 2017). The administration time is one hour. A previous study (Skoglund et al., 2014) documented adequate parallel-form reliability by correlating the total score from a paper and a computerized version of the GMA test ($r = .85$). The present study used the mean GMA stanine score.

NEO-PI-3
The NEO-PI-3 is a 240-item self-report 5-point Likert scale test, aiming to measure the FFM of personality, including six facets for each of the five domains (neuroticism, extraversion, openness to experience, agreeableness, and conscientiousness) (McCrae & Costa, 2010). The present study used raw scores based on the Norwegian version of the NEO-PI-3 (Martinsen, 2007). Test completion time was approximately 45 minutes. Factor analyses supporting the five-factor structure of this inventory are reported in Martinsen et al. (2020). Based on the present sample, Cronbach's alpha values were .91 (neuroticism), .88 (extraversion), .87 (openness), .88 (agreeableness), and .91 (conscientiousness).

NMPI
The NMPI is a self-report personality test consisting of 79 items developed by the Norwegian Armed Forces (Skoglund et al., 2020). The items are rated on a 7-point Likert scale, aiming to measure the five factors of (number of items in parenthesis) emotional stability (15), extraversion (18), openness to experience (17), agreeableness (16), and conscientiousness (13). In the development of the NMPI, most items were extracted from the International Personality Item Pool database (Goldberg et al., 2006) and then translated to Norwegian. Experienced military psychologists provided the remaining items in the NMPI. There are no facet scores for this test. The present study used raw scores, and the test completion time was approximately 15 minutes. Based on the present sample, Cronbach's alpha values were .89 (emotional stability), .89 (extraversion), .78 (openness), .90 (agreeableness), and .87 (conscientiousness).

Interview
Candidates were interviewed and rated by an experienced officer if they had passed a medical examination and physical tests. The interview lasted one hour, and the personality test results were not available. The three competencies of role model, mental robustness, and development were rated by a 9-point scale (1 indicated the weakest score and 9 the strongest). The interviewers were expected to provide their ratings based on a detailed scoring system that operationalized each competency into example answers along the 9-point scales. An OAR

was calculated as the mean competency score and used for scoring interview leadership potential, a variable used in the final selection decision. There were five or six questions for each competency, and a mix of behavioral and situational questions was applied. Examples of behavioral questions were "Please tell me of one episode where you received challenging feedback about yourself" (role model), "Can you remember a situation of unpredictability in your studies/work?" (mental robustness), and "Can you give an example of finding a new and original solution to a problem" (development). No estimate of the interview interrater reliability was available.

Field selection exercise

In the seven-day field selection exercise, candidates took turns solving ongoing work sample cases as leaders within teams of five to seven. Typical for these cases were threats from hostile forces while maneuvering in difficult terrain, establishing camps, or providing first aid to wounded soldiers. The field selection exercise was physically demanding, and the candidates experienced frequent discomfort, including some lack of food and sleep. An experienced military selection officer followed the team and rated the candidates. With a few exceptions, officers did not rate the same candidates they had interviewed the week before. All five competencies were rated on a 9-point scale, operationalized in behaviorally anchored rating scales (BARS) adapted to the different work sample cases. The personality test results were not available to the selection officers. While isolated competencies were rated as per work sample, an across-exercise (i.e., work samples and other observations) system was applied in the end. This meant that the final competency rating was the product of multiple observations in different settings. Some selection officers used a mathematical approach (i.e., mean score) and others did not. A final OAR based on mean competency scores was used for scoring the field exercise leadership potential, used in the final selection decision. Interrater reliability estimates of the field exercise ratings were not available.

Statistical analyses

IBM SPSS 26.0 was used for all statistical analyses. Only two to four subjects were missing the mean domain scores on the NEO. However, on the NMPI, there was a larger portion of missing data, with 45–58 subjects missing the mean scores on the factors. This difference in missing data was most likely due to respondent fatigue, as the NMPI came last in the questionnaire used. For those with complete ratings on the field selection exercise, the missing data for the NMPI had dropped to 21–34. One to two participants were missing GMA scores. The analyses used pairwise deletion of cases to handle missing data. Initial inspections of normality, linearity, multicollinearity, and homoscedasticity did not reveal any serious violations of the statistical assumptions. As the field selection exercise ratings were obtained when the team leader was part of a candidate group, a possibility of dependency of observations was actualized. However, Sørli et al. (2020) reported no need for multilevel modeling based on their dataset (which included the sample used in the present

study) by investigating the group-level variation of ratings based on a fixed model and a random model of the data.

For investigating if correction for range restriction was appropriate, the standard deviations of GMA and personality trait scores in three groups based on the selection hurdles were inspected: (1) candidates attending the first two days of the selection program; (2) candidates obtaining interview ratings; and (3) candidates obtaining field selection exercise ratings. These standard deviations showed only minor differences. Furthermore, it would be imprecise to use GMA and NEO population norms as an unrestricted group for our purpose of investigating military officer candidates. Thus, because data from military studies relevant for correcting interrater unreliability were also lacking, observed associations were used for testing our hypotheses.

We chose to omit the intercorrelations between ratings in **Table 1** for increasing readability and report these results here. There were large correlations between the competency ratings in both the interview ($r = .72$–$.79$) and the field selection exercise ($r = .75$–$.84$). The interview OAR and the field selection exercise OAR did not correlate strongly ($r = .26$), indicating a large amount of nonshared variance between these variables. Hierarchical regression analyses were used to test H_1 (interview) and H_3 (field selection exercise), with the candidate OARs as dependent variables. In the first analytic step, the control variables sex, age, and GMA were entered, followed by the FFM traits in the second. For testing H_2 (interview) and H_4 (field selection exercise), the correlations between the specific competency ratings and the FFM traits were used. Although the study aim did not include an investigation of the associations between NEO facets and candidate ratings, we have provided these correlations in a supplementary file (appendix).

Results

Table 1 provides the correlations between the NEO domains and the NMPI factors on the one side and the interview and field selection exercise ratings on the other. Small statistically significant correlations were observed between neuroticism/emotional stability, extraversion, agreeableness, and conscientiousness and the interview OAR ($r = -.09/.10 - .23$). Due to the opposed direction of neuroticism and emotional stability, a minus sign indicates the correlation for the former. For the field selection exercise OAR, only openness demonstrated a statistically significant correlation ($r = -.10$).

Table 2 summarizes the hierarchical multiple regression analyses using age, sex, and GMA as control variables. The inclusion of the NMPI factors provided a significant contribution to explaining the rating variance in both the interview (7%) and field selection exercise (3%); whereas, the NEO domains only showed a significant contribution for the interview (5%). Thus, personality variables contributed to an overall marginal increment in explained variance above that provided by the control variables, somewhat higher for the interview ratings relative to the field selection exercise ratings.

For testing hypotheses 1 and 3, the standardized regression coefficients were used. The NMPI demonstrated

Table 1: Descriptive Statistics and Correlations for Study Variables (N = 551–901).

	M	SD	Range	Sex	Age	GMA	N	E	O	A	C
Control											
Sex	–	–	–	–							
Age	19.62	1.86	18–34	.00	–						
GMA	6.62	1.23	3–9	–.20**	–.15**	–					
Personality							**NEO-NMPI correlations**				
NEO N (NMPI ES)	1.31 (5.24)	0.42 (0.86)	0.25–3.02 (1.73–6.93)	.15** (.19**)	–.03 (.06)	–.15** (.15**)	–.82**	.39**	–.23**	–.32**	–.41**
NEO E (NMPI E)	2.71 (5.19)	0.37 (0.74)	1.40–3.67 (2.61–7.00)	.15** (.06)	.10** (.07*)	–.11** (–.02)	–.31**	.80**	.33**	.61**	.28**
NEO O (NMPI O)	2.43 (4.83)	0.38 (0.61)	1.33–3.58 (2.76–6.82)	.09* (.01)	–.06 (–.05)	.18** (.15**)	–.12**	.32**	.80**	.31**	.10**
NEO A (NMPI A)	2.60 (5.72)	0.36 (0.67)	1.08–3.60 (2.13–7.00)	.20** (.18**)	–.06 (–.08*)	–.05 (–.07)	–.08*	.02*	.08*	.62**	.27**
NEO C (NMPI C)	2.93 (5.70)	0.37 (0.68)	1.52–3.88 (3.62–7.00)	.12** (.15**)	.01 (.02)	–.00 (–.05)	–.33**	.29**	.19**	.39**	.82**
Ratings							**NEO (NMPI)**				
							N(ES)	E	O	A	C
Interview											
Role model	6.26	1.67	1–9	.09**	.03	.00	–.06 (.07)	.19** (.19**)	.03 (.00)	.06 (.15**)	.09* (.10**)
Mental robust	6.05	1.69	1–9	.09**	.01	.02	–.11** (.13**)	.20** (.23**)	.03 (.01)	.04 (.13**)	.12** (.12**)
Development	5.96	1.68	1–9	.09**	.02	.03	–.07* (.08*)	.19** (.20**)	.05 (.02)	.06 (.15**)	.11** (.11**)
Field exercise											
Role model	5.68	2.07	1–9	.02	.17**	–.08	–.00 (–.01)	.05 (.04)	–.09* (–.11*)	.04 (.02)	.03 (.04)
Task focus	5.44	2.07	1–9	.05	.15**	–.03	–.03 (.02)	.05 (.09*)	–.09* (–.10*)	–.02 (–.00)	.02 (.03)
Mental robust	5.66	2.08	1–9	–.02	.17**	–.04	–.07 (.03)	.07 (.07)	–.09* (–.10*)	–.00 (–.01)	.03 (.05)
Cooperation	5.56	1.85	1–9	.05	.17**	–.04	–.01 (–.00)	.08 (.06)	–.04 (–.09)	.06 (.02)	.06 (.07)
Development	5.56	1.83	1–9	.07	.14**	–.03	–.01 (.02)	.06 (.07)	–.03 (–.04)	.03 (.02)	.07 (.08)
OAR											
Interview	6.09	1.53	1–9	–.10**	.02	.02	–.09* (.10**)	.21** (.23**)	.04 (.01)	.06 (.16**)	.11** (.12**)
Field exercise	5.58	1.82	1–9	.03	.18**	–.05	–.03 (.01)	.07 (.07)	–.08 (–.10*)	.02 (.01)	.05 (.06)

Note: Sex coded 1 = male and 2 = female; GMA = General Mental Ability; N = Neuroticism; ES = Emotional Stability; E = Extraversion; O = Openness to Experience; A = Agreeableness; C = Conscientiousness; OAR = Overall Assessment Rating; In the NEO-NMPI correlations matrix, the diagonal refers to correlations between NEO domains (row) and corresponding NMPI factors (column).

* p < .05; ** p < .01 (two-tailed).

Table 2: Summary of Hierarchical Regression Analyses for NEO domains and NMPI Factors Predicting Interview OAR and Field Selection Exercise OAR, controlling for Sex, Age, and GMA. Values within Brackets refer to Models using the NMPI.

	Selection method			
	Interview OAR (N = 810)		Field Exercise OAR (N = 551)	
Predictor	Step 1 β	Step 2 β	Step 1 β	Step 2 β
Sex	.11**	.09*(.09*)	.03	.02 (.02)
Age	.03	.05 (.05)	.17***	.18***(.18***)
GMA	.04	.08*(.07)	−.02	.02 (−.01)
Neuroticism[a]		−.00 (.01)		.02 (−.04)
Extraversion		.22*** (.25***)		.11*(.17**)
Openness		−.05 (−.14**)		−.11*(−.16**)
Agreeableness		−.01 (.05)		.02 (−.02)
Conscientiousness		.06 (.03)		.03 (.05)
R^2	.01*	.06*** (.08***)	.03***	.05** (.06***)
R^2 change	−	.05***(.07***)	−	.02 (.03*)
F	3.43*(3.23*)	6.43***(7.94***)	6.05***(5.70**)	3.44**(4.08***)

Note: OAR = Overall Assessment Rating; β = Standardized regression coefficient; Sex coded 1 = male and 2 = female; GMA = General Mental Ability; [a]Reversed in the NMPI (emotional stability).
* p < .05; ** p < .01; *** p < .001 (two-tailed).

a unique statistical contribution for extraversion and openness in both selection methods (β = .25/.17 and −.14/−.16). The NEO showed a unique statistical contribution for extraversion concerning ratings in both selection methods (β = .22/.11). However, for openness, a statistically significant beta value was reached for the field selection exercise only (β = −.11). Therefore, while extraversion was to some extent positively related to candidate ratings, openness was somewhat negatively related when controlling for age, sex, and GMA. The individual predictors of neuroticism, agreeableness, and conscientiousness were not significantly related to the candidate ratings.

Regarding hypotheses 2 and 4, concerning associations between the personality traits and the specific competency ratings, the correlations provided in **Table 1** were used for investigation. Inspection of the correlations between the personality measures and the interview competency ratings (role model, mental robustness, and development) indicated no clear competency-dependent associative patterns with the five personality traits. Regarding the field selection exercise, a marginal tendency of associations was found between openness and the competencies of role model, task focus, and mental robustness, but not for cooperation and development.

The NEO and the NMPI coincide mostly in terms of associations toward the ratings, likely due to the correlation between these measures (shown in the upper right of **Table 1**). For agreeableness, however, there was a more moderate correlation (r = .62) pointing to somewhat different operationalizations between the measures regarding this trait. This may explain why NEO and NMPI agreeableness had different correlations toward the interview ratings.

Discussion

Studies of construct embeddedness in interviews and assessment centers are important for an increased understanding of predictors (Berry et al., 2007; Collins et al., 2003, Roth et al., 2005). We sat forth to investigate this in a military setting by scrutinizing associations between FFM personality traits and ratings at an officer selection program when controlling for the well-known performance predictor of cognitive ability. Four hypotheses targeted expected associations between candidate personality and ratings of both OARs and five leadership competencies. Overall, our findings indicated that (1) personality traits were to some degree related with the OARs, where extraversion and openness demonstrated small isolated positive and negative associations, respectively, and (2) there was a lack of expected associations between FFM traits and single competencies. Thus, findings partially supported hypotheses 1 and 3, while no support was found for hypotheses 2 and 4.

The association between extraversion and candidate ratings was expected and is in agreement with previous civilian and military studies (Collins et al., 2003; Salgado & Moscoso, 2002; Sørlie et al., 2020; Thomas et al., 2001). Trait activation theory may shed light on extraversion embeddedness. Tett and Burnett (2003) theorized that situational cues increase the relevancy of a given personality trait—and arguably, a military selection arena holds several such cues for extraversion activation (Darr et al., 2018). However, our results for openness contrast civilian findings (Collins et al., 2003; Salgado & Moscoso, 2002), which may point to differences in civilian and military selection processes concerning the attractiveness of this trait. Interestingly, a large-scale study documented that individuals low on openness were

more likely to enter service in the German military than those scoring higher on this trait (Jackson et al., 2012). We speculate that, on average, action-oriented concrete thinkers may thrive more in military organizations relative to individuals with tendencies toward abstract thinking and aesthetic interests. As such, it is perhaps an advantage to be somewhat conventional when answering military interview questions and when persevering in a demanding exercise in the outdoors. It could also be that the interview questions and the work samples in the field selection exercise were suboptimal for triggering individual differences in openness.

It was somewhat surprising to observe the marginal personality overlap with the two selection methods, and especially so for the field selection exercise considering the duration of seven days. Acknowledgment of situational strength may contribute to understanding the scarcity in personality embeddedness. The gist of situational strength theory is that a strong situation provides guidelines or cues for expected behavior, whereas a weak situation does not (Judge & Zapata, 2015). In the field selection exercise, candidates were provided with uniforms and basic military equipment, and they certainly understood that they were under observation. Such contextual factors most likely constituted a strong situation, possibly constraining the manifestation of personality differences between candidates. While Darr (2011) reported an overall generalizability of published meta-analytic FFM estimates concerning the prediction of military job performance, it could be that situational forces are more salient in a selection setting. We also note that considering the bandwidth debate, questioning whether broad traits or its subcomponents are the best predictors for performance (e.g., Judge et al., 2013), NEO facets could perhaps show higher associations with the ratings relative to the NEO domains. However, as shown in the supplementary file, correlations at the NEO facet level did not show clearly more evident associations toward the ratings than the NEO domains. Some nuances can be seen, though, first and foremost regarding the field selection exercise where no NEO domains correlated significantly with the OAR. However, the facet of activity (extraversion) did show a significant positive association ($r = .18$), and the facets of ideas (openness), compliance (agreeableness), and depression (neuroticism) did show significant negative associations ($r = -.13, -.11,$ and $-.09$).

By testing hypotheses of differential personality-competency associations, we could investigate possible personality overlaps in a more nuanced way relative to the usage of the OARs. However, our hypotheses were not supported, and there was otherwise no clear pattern in the correlations. This lack of a clear pattern, due to the high intercorrelations between the competency ratings, most likely point to a practice where interviewers and assessors rated candidates based on global evaluations. The high intercorrelations may, of course, be due to a "g" factor, where candidates who excel on one leadership competency excel on others as well—such tendencies have, for example, been demonstrated in ratings of job performance (Viswesvaran et al., 2005). However, it is

also relevant to note findings from decision-making psychology, where several cognitive biases fuel so-called "system 1" thinking characterized by fast and intuitive information processing (Kahneman, 2011), which can potentially threaten the use of the interview scoring system and the BARS. One such likely bias is the halo effect, whereby a global evaluation of a person influences judgments of specific attributes (Nisbett & Wilson, 1977; Viswesvaran et al., 2005). We did not, however, obtain data on the decision-making processes of selection officers. Based on the high intercorrelations of competency ratings, it is difficult to argue for aspects of the construct validity of the competencies themselves. The present study shows that when using the leadership competencies in a practical selection context, ratings of the isolated competencies intercorrelate highly.

The results revealed thus a suboptimal rating practice at the officer selection program. Still, we do not intend to criticize the selection methods of the interview and the field exercise per se. With stronger associations between constructs measured by cost-friendly psychometric tests and judgment-based ratings, one could, from a predictive perspective, argue for the unnecessity of employing the more costly rater-based selection methods (Collins et al., 2003). As seen in our study, with low associations, the argument can be turned around, thus pointing to a potential for incremental validity when using psychometric tests as predictors. Although acknowledging the weak embeddedness of the established predictors of cognitive ability and personality traits in the ratings, we do not know the predictive validity of the interview and field selection exercise ratings toward military job performance. Nevertheless, such costly selection methods are valuable for other purposes than purely predictive. Among those are realistic job previews and beginning socialization into a military identity, which may foster positive applicant reactions and acceptance rates for chosen candidates.

Study limitations
The present study has some limitations. First, as all research participants were preselected through a conscript assessment procedure and also had actively applied for attending the selection program, there was some risk of range restriction in study variables. However, our purpose was not to generalize to the general population but to preselected candidates for officer selection. Thus, our relevant unrestricted group would be those attending the first two days at the selection program (i.e., before the interview and the field selection exercise). Considering, for example, the GMA scores, there were only minor differences in the standard deviations between unrestricted and assumed restricted groups: (1) candidates attending the first two days of the selection program, $M = 6.62$, $SD = 1.23$; (2) candidates obtaining interview ratings, $M = 6.61$, $SD = 1.23$; and (3) candidates obtaining field selection exercise ratings, $M = 6.65$, $SD = 1.24$. For NEO neuroticism: (1) $M = 1.31$, $SD = 0.42$; (2) $M = 1.30$, $SD = 0.41$; (3) $M = 1.27$, $SD = 0.40$. We also note that some social desirability in the self-report personality

measures may have occurred as the data were collected in a selection setting and thus contributed to skewed distributions.

Second, the reliability and validity of competency ratings were unknown, first and foremost due to a lack of interrater reliability studies and information of selection officers' actual decision-making processes. As our criteria were a military interview and a field exercise in a simulated war context, not directly comparable to interrater reliability estimates from civilian employment interviews and assessment centers, we did not have relevant data for correcting the criteria for attenuation. Considering the field selection exercise, the authors did not have access to work sample specific ratings. Perhaps more nuanced personality competency-rating relationships would be observed with such data.

Third, unfortunately, it was not possible in our dataset to identify whether candidates applied for officer school in the Army, Navy, or Air Force. Because incumbents in these three branches within the Norwegian Armed Forces seem to have different military identities (Johansen et al., 2013), and also because they have had experiences of somewhat different traditions for officer selection prior to the implementation of the joint selection program (Hansen, 2006), personality saturation might have been somewhat branch dependent.

Future studies may investigate the reliability and predictive validity of the competency ratings. Findings would be important for evaluating if the competencies are adequately measurable and are valid indicators of individual prerequisites for a mission command leader. Considering the oftentimes impreciseness of job competencies relative to psychological constructs as emphasized by Furnham (2008), such a study may be valuable. Another research line might be investigations of selection officers' decision-making processes when expected to use interview scoring systems and BARS. Such a study would be theoretically interesting and valuable for evaluating the practical adequacy of structured selection systems in military settings.

Conclusion

While low associations between candidate personality and ratings point to the usefulness of personality testing in the selection program (e.g., the potential for incremental validity), we also suggest that personality test scores may help assessors achieve more nuanced competency assessments of candidates. There are content similarities between the FFM and the mission command competencies that military psychologists and assessors may discuss for counteracting the tendency toward global evaluations of candidates. At the time of writing, personality testing is not systematically used at the selection program (i.e., as a predictor). The NMPI developed in-house may be a promising tool for future test usage, where possible advantages for military organizations are a short administration time and a lack of propriety restriction. However, further reliability and predictive validity analyses of the NMPI are warranted before operational use.

In closing, we suggest that an awareness of whether high scorers on extraversion or openness are rated objectively can be important in military selections. While extraversion might be advantageous to some degree in military settings (Darr et al., 2018), a possible extraversion favorability in a selection program is suboptimal. Such favorability may especially be counterproductive when camouflaging low conscientiousness scores (Wiersma & Kappe, 2016) considering the predictive validity of the latter trait in the context of military job performance (Darr, 2011; Fosse et al., 2015; Salgado, 1998). Although high openness scorers presumably are few in military organizations (Jackson et al., 2012), this trait may very well be relevant for success in both educational programs and ultimately in the execution of leadership in unpredictable and potentially dangerous contexts in which military officers might operate (Campbell et al., 2010). For example, open-mindedness and creativity are possibly more adaptive than rigidity and conventionality when engaging hostile forces under changing circumstances. Furthermore, we speculate that openness can also be advantageous when developing and employing mission command leadership principles (i.e., encouraging decentralized and disciplined initiatives), such as being generally self-reflective and forthcoming when subordinates present original solutions to challenges and problems.

Note

[1] While the five-factor model and the Big Five taxonomy belong to the questionnaire and lexical traditions, respectively, the contents of the five main personality factors are essentially equivalent (Simms et al., 2017).

Additional File
The additional file for this article can be found as follows:

· **Appendix.** Correlations between the NEO-PI-3 Facets and the Performance Ratings. DOI: https://doi.org/10.16993/sjwop.135.s1

Competing Interests
The authors have no competing interests to declare.

Author Contributions
Tom H. Skoglund designed the study and wrote the manuscript. Tom H. Skoglund, Thomas Fosse and Ole Christian Lang-Ree collected the data. All authors analyzed and interpreted the data and gave feedback on the manuscript.

References
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*(1–2), 9–30. DOI: https://doi.org/10.1111/1468-2389.00160

Berry, C. M., Sackett, P. R., & **Landers, R. N.** (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology*, *60*(4), 837–874. DOI: https://doi.org/10.1111/j.1744-6570.2007.00093.x

Calleja, J. A., Hoggan, B. L., & **Temby, P.** (2019). Individual predictors of tactical planning performance in junior military officers. *Military Psychology*, *32*(2), 149–163. DOI: https://doi.org/10.1080/08995605.2019.1691405

Campbell, D. J., Hannah, S. T., & **Matthews, M. D.** (2010). Leadership in military and other dangerous contexts: Introduction to the special topic issue. *Military Psychology*, *22*, S1–S14. DOI: https://doi.org/10.1080/08995601003644163

Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & **Le, H.** (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment*, *11*(1), 17–29. DOI: https://doi.org/10.1111/1468-2389.00223

Darr, W. A. (2011). Military personality research: A meta-analysis of the Self Description Inventory. *Military Psychology*, *23*(3), 272–296. DOI: https://doi.org/10.1080/08995605.2011.570583

Darr, W. A., & **Catano, V. M.** (2016). Determining predictor weights in military selection: An application of dominance analysis. *Military Psychology*, *28*(4), 193–208. DOI: https://doi.org/10.1037/mil0000107

Darr, W. A., Ebel-Lam, A., & **Doucet, R. G.** (2018). Investigating the extravert advantage in training: Exploring reward sensitivity, training motivation, and self-efficacy as intermediary factors. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 50*(3), 172–184. DOI: https://doi.org/10.1037/cbs0000102

Defence Staff Norway. (2012). *Chief of Defence's (CHOD) Policy on Leadership in the Norwegian Armed Forces*. Norway: Norwegian Armed Forces.

Fosse, T. H., Buch, R., Säfvenbom, R., & **Martinussen, M.** (2015). The impact of personality and self-efficacy on academic and military performance: The mediating role of self-efficacy. *Journal of Military Studies*, *6*(1), 47–65. DOI: https://doi.org/10.1515/jms-2016-0197

Furnham, A. (2008). *Intelligence and personality at work. Exploring and explaining individual differences at work.* Routledge. DOI: https://doi.org/10.4324/9780203938911

Hansen, I. (2006). *Bidrag til psykologitjenestens historie i Forsvaret fra 1946–2006* [Contributions to the History of the Norwegian Military Psychology Service 1946–2006]. Norwegian Defence University College.

Hoffman, B. J., Kennedy, C. L., Lopilato, A. C., Monahan, E. L., & **Lance, C. E.** (2015). A review of the content, criterion-related, and construct-related validity of assessment center exercises. *Journal of Applied Psychology*, *100*(4), 1143–1168. DOI: https://doi.org/10.1037/a0038707

Huffcutt, A. I., Conway, J. M., Roth, P. L., & **Stone, N. J.** (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology*, *86*(5), 897–913. DOI: https://doi.org/10.1037/0021-9010.86.5.897

Hystad, S. W., Eid, J., Laberg, J. C., & **Bartone, P. T.** (2011). Psychological hardiness predicts admission into Norwegian military officer schools. *Military Psychology*, *23*(4), 381–389. DOI: https://doi.org/10.1080/08995605.2011.589333

International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, *41*, 1244–1273. DOI: https://doi.org/10.1177/0149206314567780

Jackson, J. J., Thoemmes, F., Jonkmann, K., Lüdke, O., & **Trautwein, U.** (2012). Military training and personality trait development: Does the military make the man, or does the man make the military? *Psychologcial Science*, *23*(3), 270–277. DOI: https://doi.org/10.1177/0956797611423545

Johansen, R. B., Laberg, J. C., & **Martinussen, M.** (2013). Military identity as predictor of perceived military competence and skills. *Armed Forces & Society*, *40*(3), 521–543. DOI: https://doi.org/10.1177/0095327X13478405

Jugde, T. A., Rodell, J. B., Klinger, R. L., Simon, L. S., & **Crawford, E. R.** (2013). Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives. *Journal of Applied Psychology*, *98*(6), 875–925. DOI: https://doi.org/10.1037/a0033901

Judge, T. A., & **Zapata, C. P.** (2015). The person-situation debate revisited: Effect of situation strength and trait activation on the validity of the Big Five personality traits in predicting job performance. *Academy of Management Journal, 58*(4), 1149–1170. DOI: https://doi.org/10.5465/amj.2010.0837

Kahneman, D. (2011). *Thinking, Fast and Slow.* US: Farrar, Straus and Giroux.

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

Martinsen, Ø. L. (2007). *Norwegian translation of the NEO PI.* Hogrefe.

Martinsen, Ø. L., Furnham, A., McCrae, I., & **Fosse, T. H.** (2020). Personality traits as predictors of leadership progression in the armed forces. *Submitted manuscript.*

McCrae, R. R., & Costa, P. T. (2010). NEO Inventories Professional Manual. *Psychological Assessment Resources*.

McCrae, R. R., Costa, P. T., & Martin, T. A. (2005). The NEO-PI-3: A more readable revised NEO personality inventory. *Journal of Personality Assessment*, *84*(3), 261–270. DOI: https://doi.org/10.1207/s15327752jpa8403_05

NAF (Norwegian Armed Forces). (2020). Lederkandidatstudien [Leadership Candidate Study]. Retrieved from https://forsvaret.no/hogskolene/Forskning/lederkandidatstudien

Nisbett, R., & Wilson, T. D. (1977). Telling more than we can know: Verbal reports on mental processes. *Psychological Review*, *84*(3), 231–259. DOI: https://doi.org/10.1037/0033-295X.84.3.231

Oshio, A., Taku, K., Hirano, M., & Saeed, G. (2018). Resilience and Big Five personality traits: A meta-analysis. *Personality and Individual Differences*, *127*, 54–60. DOI: https://doi.org/10.1016/j.paid.2018.01.048

Roth, P. L., Van Iddekinge, C. H., Huffcutt, A. I., Eidson, C. E., Jr., & Schmit, M. J. (2005). Personality saturation in structured interviews. *International Journal of Selection and Assessment*, *13*(4), 261–273. DOI: https://doi.org/10.1111/j.1468-2389.2005.00323.x

Rumsey, M. G., & Arabian, J. M. (2014). Military Enlistment Selection and Classification: Moving Forward. *Military Psychology*, *26*(3), 221–251. DOI: https://doi.org/10.1037/mil0000040

Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civil occupations: A European perspective. *Human Performance*, *11*(2), 271–288. DOI: https://doi.org/10.1207/s15327043hup1102&3_8

Salgado, J. F., & Moscoso, S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology*, *11*(3), 299–324. DOI: https://doi.org/10.1080/13594320244000184

Schmidt, F. L., & Hunter, J. (1998). The validity and utility of selection methods in personnel psychology. *Psychological Bulletin*, *124*(2), 262–274. DOI: https://doi.org/10.1037/0033-2909.124.2.262

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. DOI: https://doi.org/10.1037/0022-3514.86.1.162

Schmidt, F. L., Oh, I.-S., & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings. *Working paper.*

Shaffer, J. A., & Postlethwaite, B. E. (2012). A matter of context: A meta-analytic investigation of the relative validity of contextualized and noncontextualized personality measures. *Personnel Psychology*, *65(3)*, 445–493. DOI: https://doi.org/10.1111/j.1744-6570.2012.01250.x

Simms, L., Williams, T. F., & Simms, E. N. (2017). Assessment of the Five Factor Model. In T. A. Widiger (Ed.), *The Oxford Handbook of the Five Factor Model.* Oxford University Press. DOI: https://doi.org/10.1093/oxfordhb/9780199352487.013.28

Skoglund, T. H., Brekke, T.-H., Steder, F. B., & Boe, O. (2020). Big Five personality profiles in the Norwegian Special Operations Forces. *Frontiers in Psychology*, *11*, 747. DOI: https://doi.org/10.3389/fpsyg.2020.00747

Skoglund, T. H., Martinussen, M., & Lang-Ree, O. C. (2014). Papir versus PC [Paper versus PC]. *Journal of the Norwegian Psychological Association, 51*(6), 450–452.

Sørlie, H. O., Hetland, J., Dysvik, A., Fosse, T. H., & Martinsen, Ø. L. (2020). Person-Organization Fit in a military selection context. *Military Psychology*, *32*(3), 237–246. DOI: https://doi.org/10.1080/08995605.2020.1724752

Tay, C., Ang, S., & Van Dyne, L. (2006). Personality, biographical characteristics, and job interview success: A longitudinal study of the mediating effects of interviewing self-efficacy and the moderating effects of internal locus of causality. *Journal of Applied Psychology*, *91*(2), 446–454. DOI: https://doi.org/10.1037/0021-9010.91.2.446

Thomas, J. L., Dickson, M. W., & Bliese, P. D. (2001). Values predicting leader performance in the U.S. Army Reserve Officer Training Corps Assessment Center: Evidence for a personality-mediated model. *The Leadership Quarterly*, *12*(2), 181–196. DOI: https://doi.org/10.1016/S1048-9843(01)00071-6

Thornton, G. C., & Gibbons, A. M. (2009). Validity of assessment centers for personnel selection. *Human Resource Management Review*, *19*(3), 169–187. DOI: https://doi.org/10.1016/j.hrmr.2009.02.002

Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, *90*(1), 108–131. DOI: https://doi.org/10.1037/0021-9010.90.1.108

Wiersma, U., & Kappe, R. (2016). Selecting for Extroversion but rewarding for Conscientiousness. *European Journal of Work and Organizational Psychology, 26*(2). DOI: https://doi.org/10.1080/1359432X.2016.1266340

Yukl, G. (2012). Effective leadership behavior: What we know and what questions need more attention. *Academy of Management Perspectives*, *24*(4), 66–85. DOI: https://doi.org/10.5465/amp.2012.0088

Study 2

Big Five Personality Profiles in the Norwegian Special Operations Forces

Tom Hilding Skoglund[1,2]*, Thor-Håvard Brekke[3], Frank Brundtland Steder[4] and Ole Boe[2,5]

[1] Faculty of Health Sciences, UiT The Arctic University of Norway, Tromsø, Norway, [2] Psychology Department, Bjørknes College, Oslo, Norway, [3] Norwegian Armed Forces, Oslo, Norway, [4] Norwegian Defence Research Establishment, Oslo, Norway, [5] University of South-Eastern Norway, School of Business, Drammen, Norway

This study is the first to report on Big Five personality traits of employees in the Norwegian military Special Operations Forces (NORSOF). Three research questions were formulated for this study, aiming to investigate (1) whether age, number of combat-deployments and rank (OR/OF) had an impact on the personalities of NORSOF employees, (2) possible personality differences between personnel organized in the underlying departments of the NORSOF, and (3) if there were personality differences between SOF-operators and conventional forces applicants. SOF-operators from the Norwegian Special Operations Commando (FSK) and the Norwegian Naval Special Operations Commando (MJK) constituted 40% of the total NORSOF sample (N = 190), whilst the term SOF-support categorized the larger proportion of non-operators. Results indicated that younger employees tended to be lower on emotional stability than older colleagues, and that those without any combat-deployments were somewhat higher on agreeableness and a bit lower on emotional stability relative to employees with such experience. Additionally, personnel with officer ranks (OF) were higher on extraversion compared to specialists (OR). Results did not show any significant intradepartmental differences in mean personality trait scores. Compared to male applicants for basic officer training in conventional forces (N = 662), SOF-operators (all males) were less extroverted, less agreeable, and slightly more emotionally stable. The authors conclude that the NORSOF attracts and recruits personnel with similarities in their Big Five personalities. Furthermore, we suggest that the personality profile that emerged for the "average" Norwegian SOF-operator is a functional one, especially when considering the desired future image of the Special Forces operative as a Warrior-Diplomat.

Keywords: military, SOF, special forces, Big Five, personality

INTRODUCTION

Special Operations Forces (SOFs) are typically characterized by stringent personnel selection systems, advanced military training regimes and versatile mission forms, including asymmetrical warfare, making such units one the most demanding of all military specialties. Such units have a wider range of capabilities and often operate more independently, usually in smaller teams, relative to regular military forces. Having a flexible structure, SOFs can function as liaisons between military

forces, civilian organizations and different law enforcement agencies in international operations – in addition to being a tactical level asset for special missions. Personnel requirements such as versatility, agility, effectiveness, speed and surprise, working both independently and in direct support to others, are often highlighted in published materials concerning SOFs (NATO, 2012). To distinguish a SOF-operator from soldiers in other military units, Spulak (2007) used the term "Elite warrior" in order to pinpoint that operators typically demonstrate superior task performance relative to the performance of the larger military population, thus underlining the "specialness" of SOFs. The term "Warrior-Diplomat" has been used to describe the ideal future image of SOF-personnel (Berg-Knutsen and Roberts, 2015), whereby operators are capable of combining tactical level competency with insights in societal perspectives, such as political and cultural processes, thus enabling optimal job performance across sectors and institutions in an international context.

More detailed job analyses from SOFs and specific psychological requirements for its personnel are, however, confidential information in most countries. This clouds the possibility of gaining more precise knowledge of the psychological characteristics of employees, especially for its "tip of the spear" operators. Published empirical studies with samples from SOFs are scant. Thus, studies with such samples, acknowledging security concerns such as personnel anonymity, are needed in the military psychology literature. Empirical studies can contribute to improvements in selection systems and other human resource strategies within these important military assets, and ultimately support the SOF truth that "people are more important than hardware" (USSOCOM, 2019, p. 57).

The Norwegian Special Operations Forces (NORSOF) are organized with two Special Forces and a joint staff. The Norwegian Special Operations Commando (NORSOC or FSK in Norwegian) is a formerly Army affiliated department, whereas the Norwegian Naval Special Operations Commando (NORNAVSOC or MJK in Norwegian) is a formerly Navy affiliated department. The former has its main historical roots in paratrooper operations, the latter mainly in frogman operations (for a historical perspective, see Olsen and Thormodsen, 2014). Both the FSK and the MJK are on national and international standby for special operations and counter-terrorism actions, and both units have frequent deployments to conflict areas abroad, where the capacity building of security forces in other states has been one of the main official missions (Hedenstrom and Kristiansen, 2016).

The operator selection to the FSK and the MJK are separate, and both selection systems are considered extremely demanding regarding physical and psychological requirements. Thus, only a few candidates complete the selection phase and become trained as combat-ready operators. In addition to the operators, the NORSOF has different personnel categories, collectively named "SOF-support" for the purpose of the present study. This category has personnel with different military specialties and backgrounds, and with the common characteristic that they have not been through an operator selection process, either in the FSK or the MJK, and therefore are not fully trained nor approved as SOF-operators. SOF-support personnel contribute to the operational capacity of the FSK and the MJK in many ways, for example by participating in training for operations, by providing contributions in diverse missions and combat-deployments, or by executing leadership from different hierarchies in the organization.

By measuring the psychological characteristics of personnel in the NORSOF, this study aims to investigate possible intradepartmental differences in individual attributes, and examine if the carefully selected and highly trained operators have psychological dissimilarities relative to those in other military personnel categories. The present study aims to examine these questions by means of Big Five personality testing. The Big Five framework is essentially equivalent to the Five Factor Model (FFM) in the concepts used for personality descriptions (John et al., 2008). The Big Five (or FFM) is a system for describing the phenomenon of personality differences from a lexical standpoint, a method for conceptualizing personality dating back to at least Allport and Odbert's (1936) classical work of extracting 18,000 person-describing words from the English dictionary. During the 1980s, the Big Five framework, and measurement methods thereof, emerged as a cemented discipline within personality psychology (Digman, 1990; Scroggins et al., 2008), spawning voluminous research publications (see for example Widiger, 2017). While the NORSOF has used Big Five testing for internal educational and organizational purposes, published Big Five personality studies of personnel employed by the NORSOF are to the authors' knowledge non-existent.

The present study used a Big Five test recently developed by the Norwegian Armed Forces, designed to measure factors with the same meaning as the definitions seen in the widespread NEO-PI-R/3 test (Costa and McCrae, 1992; McCrae et al., 2005; McCrae and Costa, 2010): Neuroticism, emotional instability rather than adjustment; Extraversion, represented by a need to be outgoing and active as opposed to being an introvert; Openness to experience, being imaginative and having broad interests as opposed to being traditional and down to earth; Agreeableness, a compassionate style toward others rather than being antagonistic; Conscientiousness, a tendency to be well organized and goal-oriented as opposed to an easy-going- and non-directional style. Big Five testing (and diverse personality tests with resemblance to the Big Five) have been widespread in military contexts, especially in the domain of personnel selection (Salgado, 1998; Campbell and Knapp(eds), 2001; Stark et al., 2014). Personality traits are often considered relevant for selection decisions in both military and civilian contexts, as they (especially conscientiousness) tend to add incremental validity after controlling for general mental ability (Schmidt and Hunter, 1998; Darr, 2009).

The Big Five traits are viewed as stable characteristics, disposing for individual patterns of cognitions, emotions and behaviors (McCrae and Costa, 2010). Through adult years, trait level changes have, however, been demonstrated. Changes seem to predominate in the age span 20–40, for example by increasement of emotional stability (Roberts and Mroczec, 2008). Further, the maturity principle in personality psychology suggests that an age related increasement in the traits of agreeableness, conscientiousness, in addition to emotional stability, could

take place (Caspi et al., 2005). Regarding military contexts, a longitudinal German study found that training in conventional forces was associated with a lasting reduction in agreeableness, as measured 5 years after the end of service (Jackson et al., 2012). This finding demonstrates a possible impact of military experience on agreeableness, opposite the change direction forecasted by the general maturity principle.

Published empirical personality studies of employees in SOFs are scant, most likely due to security issues and the secretive culture typically surrounding such units. Two Norwegian studies, using applicants, investigated the predictive validity of a Big Five test used in the operator selection to the MJK, but findings did not reveal clear associations between test scores and pass/fail results in an upcoming strenuous practical selection period (Hartmann et al., 2003; Hartmann and Grønnerød, 2009). The authors of the 2003 study did claim, however, that high emotional stability and low extraversion increased the probability of success. Bartone et al. (2008) noted that prediction of success in selection courses for elite military units have been met with limited success. These authors did, however, find a significant group difference in mean scores between those who passed and those who failed a 4-week selection and assessment course among United States Army Special Forces candidates, when investigating a measure of psychological hardiness. The authors described hardiness as a personal stress-resiliency resource and argued for the relevance of this psychological construct in SOFs.

A study of 139 NEO-PI-R profiles from operators in the United States Navy Sea-Air-Land (SEAL) Commando, found that SEALs had lower scores on neuroticism and agreeableness, the same to lower on openness, and higher scores on conscientiousness and extraversion, relative to the norms for adult American males (Braun et al., 1994). This study also reported that more-experienced SEALs scored higher on conscientiousness and lower on extraversion than less-experienced SEALs, and that commissioned officers had higher scores on both of these factors when compared to enlisted operators. Regarding the differences in personality trait levels based on amount of experience, the authors concluded that age was responsible for the effect, not warfare experience *per se*. A more recent study of a police Special Force reported that its personnel had significantly higher scores on all five factors (emotional stability being reversed neuroticism) when compared to the population mean scores of males, and also for career soldiers, in Italy (Garbarino et al., 2012). Police Special Forces can be somewhat different from military SOFs in their primary mission forms, although obvious similarities exist regarding stringent personnel selection, advanced tactical training, and special missions such as counter-terrorism (Johnsen, 2017).

Although not investigating Big Five traits *per se*, a recent Norwegian study by Boe et al. (2017) investigating character strengths in the FSK sheds light on desired personality functioning in this unit. The authors reported that 27 officers from the FSK evaluated the character strength named open-mindedness as the most important for successful duty as a Special Forces officer, and that this sample evaluated humility/modesty, curiosity, and forgiveness and mercy as more important compared to character strength evaluations done by Norwegian Army officers.

At a more general level for "high-risk operational personnel," such as SOF-operators, clandestine intelligence operatives and astronauts, Picano and Roland (2012) wrote that six attribute dimensions are commonly required for successful performance: emotional stability, adaptability, teamwork abilities, physical stamina and fitness, sound judgment and decision-making, and intrinsic motivation. The first three of these attribute dimensions point to a low degree of neuroticism, and shed light on the relevance of the interpersonal traits of extraversion and agreeableness. Intrinsic motivation can perhaps be related to the conscientiousness factor, a trait that has demonstrated predictive validity across different jobs in both civilian and military settings (Barrick and Mount, 1991; Salgado, 1998; Darr, 2009).

Summarized, the personality picture in SOFs, whether it is for applicants or employees, or for the broader category of high-risk operational personnel, is not necessarily a clear one, except for repeated findings of low neuroticism (high emotional stability). The lack of success in predicting performance in SOF selections based on personality variables, and the scant empirical personality findings of employees, clouds the picture. By reporting on additional personality data, the present study contributes further to the personality psychology knowledge base of SOFs.

Aims of the Study

The purpose of this study was to investigate personality characteristics of personnel in the NORSOF by using a Big Five test called the Norwegian Military Personality Inventory (NMPI). NMPI-data were obtained from three personnel categories, including: (1) FSK-operator, (2) MJK-operator, and (3) SOF-support. We also had information on age, number of combat-deployments and whether the participants were educated as officers (OF) or not (OR – other ranks). By investigating the impact of personnel category, age, number of combat-deployments and rank on levels of personality traits, a discussion of the organizational psychology within the NORSOF can be done – more specifically, we hypothesized that if clear personality dissimilarities were found, this could be a challenge for the climate among colleagues. Further, a database consisting of NMPI scores of applicants to basic officer training (1-year education to become a non-commissioned officer – NCO) in the Norwegian Armed Forces was used for comparative analyses. Such analyses are relevant for examining whether the SOF-operators have a different personality profile than other categories of military personnel, considering the differences in selection systems and subsequent military service for SOFs and conventional forces, respectively. We set forth to investigate if the "specialness" of the operators was reflected in their personalities. Three research questions (RQ) were formulated for this study:

(1) For the total NORSOF sample, are there significant group-differences in personality traits based on age groups, number of combat-deployments groups, and rank (OF/OR)?

(2) Are there significant group-differences in personality traits between the three NORSOF personnel categories (FSK-operator, MJK-operator, and SOF-support)?

(3) Are there significant group-differences in personality traits between SOF-operators (who were all males) and male NCO-applicants?

Based on the maturity principle in personality psychology (Caspi et al., 2005), we expected that older NORSOF personnel would score higher on agreeableness, conscientiousness and emotional stability than younger colleagues. We were unsure of potential differences based on combat-deployments and rank, however, especially since these variables were not strongly associated with age in our sample. Considering the SOF status of both the FSK and the MJK, we had no educated reason to expect clear differences in levels of personality traits between operators in the two units, even though these Special Forces divisions are separate and have their historical roots in the Army and Navy, respectively. Based on previous studies (Braun et al., 1994; Bartone et al., 2008; Garbarino et al., 2012), we expected that SOF-operators would score higher on emotional stability and conscientiousness than SOF-supports and NCO-applicants. Serving as one of "the quiet professionals," a term often used to describe Special Forces operatives, a lower degree of extraversion and agreeableness were also expected – although higher extraversion relative to the population was found in the studies of Braun et al. (1994) and Garbarino et al. (2012). The study of Jackson et al. (2012), demonstrating that military training was associated with agreeableness reduction, further supported this expectation, although we did not find personality change studies from SOF environments. The study of Boe et al. (2017) gave expectations of higher openness to experience among operators relative to other personnel categories.

MATERIALS AND METHODS

Participants

Personnel employed by the NORSOF were asked to participate in this study. Conscripts in the "Fallskjermjegertroppen" (Paratroopers) and the all-female "Jegertroppen" (Hunter Troop) were not included, as these soldiers are not employees. The final sample was N = 190. The response rate is unknown, as the actual size of this military department is classified information. Personnel selected and trained as SOF-operators either in the FSK or the MJK constituted 76 individuals (40%) in the final sample, whereas the remaining 113 (60%) were support personnel (one person did not report background).

Measures

Demographic Information

The questionnaire used included the following demographic variables: Background (FSK-operator; MJK-operator; other); Age (under 30; 30–40; above 40); Number of deployments (0; 1–5; 6–10; above 10); Rank (Other Ranks; Officer Rank; Civilian). Age and number of deployments were coded in categories for

minimizing anonymity concerns of respondents. A deployment in the NORSOF means participation in international combat operations, in which the duration can vary. In the last decade, personnel in the NORSOF have usually been deployed 4–6 months at a time in international conflict areas. To gain an officer rank in Norway, a 3-year-long education at a military academy either in the Army, Navy or Air Force, resulting in a Bachelor's degree, is required.

Norwegian Military Personality Inventory (NMPI)

The Big Five personality dimensions were measured using the NMPI. This is a self-report seven-point Likert scale factor-level test consisting of 79 items, developed by the Norwegian Armed Forces (Antonsen, 2016; Skoglund, 2017; Skoglund et al., unpublished). The five factors (number of items in parenthesis) of the NMPI are called Extraversion (18), Agreeableness (16), Conscientiousness (13), Emotional Stability (15), and Openness to new experiences (17). Skoglund et al., (unpublished) reported Cronbach's alphas of α = 0.78–0.90 for the NMPI factors, and the following correlations between the NMPI factors and the NEO-PI-3 factors based on a sample of 850 applicants for basic officer training (NCO): E, r = 0.80; A, r = 0.62; C, r = 0.82; ES (reversed Neuroticism), r = −0.82; O, r = 0.80. For the sample in the present study, Cronbach's alphas for the NMPI factors were: α = 0.88 for E and A; α = 0.85 for C; α = 0.87 for ES; α = 0.75 for O.

After promising initial validation findings (Antonsen, 2016; Skoglund, 2017), the NMPI are undergoing a research phase in the NAF, where norms and practical usage aspects are yet to be developed (at the time of writing). The authors of the present study, therefore, had limited choices regarding comparative analyses between the SOF-sample and other samples, and thus formulated RQ3 in line with accessible datasets. NMPI's short form, in-house copyright permission and complete data control without third parties were necessary requirements set forth by the NORSOF for this study. Thus, we could not use established tests, such as the NEO-PI-3 (McCrae and Costa, 2010).

Procedure

The Norwegian Social Science Data Service, the research commission at the Norwegian Defence University College, and the NORSOF approved this study. The Chief Psychologist in the Norwegian Armed Forces has the copyright permission for the NMPI, and gave permission to use the test for research purposes. The questionnaire with demographic variables and the NMPI were distributed through a military mail-system during May 2018.

Statistical Analyses

IBM SPSS 25.0 was used for all statistical analyses. Descriptive analyses investigated characteristics of the sample, whereas one-way between-groups analyses of variances (ANOVAs) and independent-samples t-tests were conducted for comparing groups. Missing values were excluded analysis by analysis. Effect sizes are reported as eta squared for ANOVAs and as Cohen's d for t-tests. Eta squared and Cohen's d were interpreted in line with Cohen (1988), classifying 0.01 and 0.2 as a small effect,

0.06 and 0.5 as a medium effect and 0.14 and 0.8 as a large effect, respectively.

RESULTS

Demographic Information and Evaluation of Parametric Assumptions

Table 1 presents a summary of demographic information for the sample. There were more participants reporting a background as an FSK-operator compared to an MJK-operator. All participants reporting operator backgrounds were males. Having asked for background in the questionnaire, the authors did not have information of the proportion of operators no longer on active duty as operatives in saber squadrons (operational units). There were a few females in the SOF-support group, but not enough for analyses of potential gender differences in personality scores. Age distribution had a somewhat similar shape for the FSK-operators and the MJK-operators, whereas there were a higher proportion of participants above 40 years of age in the larger SOF-support group. The majority of participants in all three groups had one to five combat-deployments. The majority of both operator groups had a specialist rank (OR), whereas there were a greater proportion of participants with officer ranks (OF) in the support group. As only a handful reported civilian positions, they were removed from the sample. There was one missing value on the variables of rank, background and number of combat-deployments.

Investigating the total NORSOF sample, the score distributions on the NMPI factors of agreeableness, conscientiousness and emotional stability were somewhat negatively skewed (−0.95, −0.47, and −0.56, respectively), whereas the distribution of scores on extraversion and openness were quite symmetrical (0.04 and −0.23, respectively). A visual inspection of histograms and Q-Q (quantile-quantile) plots

revealed no serious violations of normality, legitimizing usage of parametric testing. When extracting the SOF-operators as a group, the same pattern in score distributions emerged. Further, Levene's test for homogeneity of variance are reported on for the specific statistical tests used.

Research Question 1

One-way between-groups ANOVAs and independent samples t-tests were conducted to explore the impact of age, number of combat-deployments and rank on NMPI factor scores. The separate analyses of age and number of combat-deployments were supported by a correlation of $r = 0.39$ between the two variables, indicating a relationship of a low-medium magnitude (Cohen, 1988). To gain statistical power (that is, avoiding too few participants in the demographics categories), the total NORSOF sample was used for analyses investigating RQ1. Levene's tests for homogeneity of variance were not significant for the analyses related to RQ1, with the exception of agreeableness scores based on the grouping variable of combat-deployments ($p < 0.05$). As this indicated a violation of homogeneity of variance, Welch's F test was used for this specific analysis. Precisions of statistically significant findings are indicated by the 95% confidence intervals of the differences between the means.

Table 2 summarizes the ANOVA analyses. There was a statistically significant difference in emotional stability scores for the three age groups (under 30, 30–40, above 40): $F(2,187) = 5.17$, $p = 0.007$. The eta squared effect size was 0.05. *Post hoc* comparisons using the Tukey HSD test demonstrated that the mean score for those less than 30 years of age ($M = 5.16$, $SD = 0.73$) was significantly different from those above 40 years of age ($M = 5.63$, $SD = 0.72$), 95% CI [0.09, 0.58].

There were only three participants who had more than 10 combat-deployments, thus the categories for number of combat-deployments were re-coded into three groups (None; 1–5 deployments; above five deployments). There was a statistically significant difference in scores on agreeableness for the three groups: *Welch's* $F(2,41.61) = 5.33$, $p = 0.009$. The eta squared effect size was 0.04. *Post hoc* testing using the Games-Howell procedure showed that the mean score of those without any deployments ($M = 5.81$, $SD = 0.61$) was significantly different from those with both 1–5 deployments ($M = 5.32$, $SD = 0.63$), 95% CI [0.06, 0.91], and those with more than five deployments ($M = 5.24$, $SD = 0.94$), 95% CI [0.10, 1.04].

In addition, there was a statistically significant difference in scores on emotional stability for the three groups $F(2,186) = 3.22$, $p = 0.042$. The eta squared effect size was 0.03. *Post hoc* comparisons using the Tukey HSD test demonstrated that the mean score for those without any deployments ($M = 5.15$, $SD = 0.70$) was significantly different from those with more than five deployments ($M = 5.64$, $SD = 0.75$), 95% CI [0.01, 0.98]. These statistically significant differences in agreeableness and emotional stability based on combat-deployments are, however, somewhat imprecise (large confidence intervals) – most likely due to the small sample size in the non-deployment group ($N = 18$).

Finally, for RQ1, independent-samples t-tests were conducted to explore differences in mean factor scores based on rank, OR

TABLE 1 | Demographic information of the sample (in percentages).

Demographic variable	FSK-operators (*N* = 48)	MJK-operators (*N* = 28)	SOF-supports (*N* = 113)
Gender			
Male	100	100	92.9
Female	–	–	7.1
Age			
<30	20.8	14.3	23.9
30–40	54.2	57.1	36.3
>40	25.0	28.6	39.8
Number of deployments			
None	4.2	3.6	13.4
1–5	60.4	78.6	66.1
6–10	35.4	17.9	17.9
10	–	–	2.7
Rank			
Officer (OF)	33.3	39.3	56.6
Other (OR)	66.7	60.7	43.4

TABLE 2 | NMPI factor means, standard deviations and one-way between-groups analyses of variance using age, number of combat-deployments and personnel category as predictors.

	Age				
NMPI factor	**Group 1 (N = 41): under 30**	**Group 2 (N = 83): 30–40**	**Group 3 (N = 66): above 40**	**F (η²)**	**Significant effects**
Extraversion	4.53 (0.88)	4.74 (0.79)	4.85 (0.82)	1.95 (0.02)	ns
Agreeableness	5.43 (0.70)	5.28 (0.69)	5.37 (0.79)	0.70 (0.01)	ns
Conscientiousness	5.74 (0.74)	5.65 (0.71)	5.52 (0.68)	1.31 (0.01)	ns
Emotional stability	5.16 (0.73)	5.40 (0.77)	5.63 (0.72)	5.17* (0.05)	1 < 3
Openness	4.74 (0.65)	4.84 (0.53)	4.93 (0.59)	1.38 (0.01)	ns
	Number of combat-deployments				
	Group 1 (N = 18): none	**Group 2 (N = 126): 1–5**	**Group 3 (N = 45): above 5**		
Extraversion	4.92 (0.72)	4.66 (0.81)	4.87 (0.85)	1.61 (0.02)	ns
Agreeableness[1]	5.81 (0.61)	5.32 (0.63)	5.24 (0.94)	4.34* (0.04)	1 > 2, 1 > 3
Conscientiousness	5.92 (0.66)	5.62 (0.73)	5.54 (0.67)	1.88 (0.02)	ns
Emotional stability	5.15 (0.70)	5.40 (0.75)	5.64 (0.75)	3.22* (0.03)	1 > 3
Openness	4.96 (0.61)	4.82 (0.57)	4.93 (0.56)	0.89 (0.01)	ns
	Personnel category				
	Group 1 (N = 48): FSK-operator	**Group 2 (N = 28): MJK-operator**	**Group 3 (N = 113): SOF-support**		
Extraversion	4.61 (0.74)	4.75 (0.97)	4.78 (0.82)	0.66 (0.01)	ns
Agreeableness	5.31 (0.63)	5.17 (0.77)	5.40 (0.75)	1.19 (0.01)	ns
Conscientiousness	5.68 (0.75)	5.61 (0.68)	5.61 (0.70)	0.17 (0.00)	ns
Emotional stability	5.55 (0.78)	5.51 (0.80)	5.35 (0.74)	1.38 (0.01)	ns
Openness	4.86 (0.51)	4.88 (0.50)	4.84 (0.63)	0.05 (0.00)	ns

[1]Welch's F and Games-Howell post hoc comparison used; *p < 0.05; η², eta squared; < or >, significant group differences after post hoc comparisons using the Tukey HSD test; ns, non-significant.

(N = 98) and OF (N = 91). There was a statistically significant difference in extraversion scores for the OR group (M = 4.54, SD = 0.79) and the OF group [M = 4.93, SD = 0.81; t(187) = −3.41, p = 0.00, two-tailed], 95% CI [0.17–0.63]. The effect size for this difference, using Cohen's d, was 0.50.

Research Question 2

The lower part of **Table 2** presents the NMPI mean factor scores of FSK-operators, MJK-operators, and the SOF-support group, and summarizes results from ANOVAs relevant for RQ2. The Levene's tests for homogeneity of variance were not significant for the analyses related to RQ2, thus indicating equal variances. A visual inspection revealed quite similar mean factor scores for the three groups. One-way between-groups ANOVAs did not show any statistically significant differences in NMPI factor scores between the three groups, although it should be noted that the sample sizes of FSK- and MJK-operators are quite small with 48 and 28 individuals, respectively. A power analysis using the G*power software (Faul et al., 2009) showed that the projected sample needed for a detection of a mean score difference between two groups with an effect size of Cohen's d = 0.50, and with alpha = 0.05 and power = 0.80, is N = 64 in each group. Thus, FSK- and MJK-operators were categorized as one group containing 76 individuals. Independent-samples t-tests were conducted to investigate differences in NMPI-factor

mean scores between the SOF-operator group and SOF-supports, resulting in no significant differences for any of the five factors.

Research Question 3

This study had access to a database used for validation purposes of the NMPI, based on scores from applicants attending a common assessment center for basic officer training (NCO) in the Army, Navy and Air Force in the summer of 2016 (Antonsen, 2016; Skoglund, 2017; Skoglund et al., unpublished). All participants at the assessment center were asked to participate in the validation study. They were informed that the NMPI was attached to a research project, and that scores on this test would not affect upcoming selection decisions. The final sample of NCO-applicants with NMPI factor scores were N = 843–856, giving a response rate above 90%. As all NORSOF-operators were males, male NCO-applicants were used for answering RQ3, controlling for possible gender-effects. Because of missing data on some items, there were 648–662 males with NMPI factor scores in the sample – this was approximately 77% of the total NCO-sample. Among the males, age varied from 18 to 34 (M = 19.62, SD = 1.89). The applicants had limited military experience (about half did not have any) before participating in the assessment center, in which 53% of the males either canceled the selection phase themselves or were evaluated too low on the selection criteria.

Independent-samples *t*-tests were conducted to investigate RQ3. Equal variances for the two groups were confirmed by non-significant results on Levene's tests. As **Table 3** shows, there were three findings reaching statistical significance.

A statistically significant difference in mean scores on extraversion was found for the SOF-operator group ($M = 4.67$, $SD = 0.83$) and NCO-applicant group [$M = 5.16$, $SD = 0.76$; $t(722) = -5.37$, $p = 0.00$, two tailed], 95% CI [0.32, 0.68]. For mean scores on agreeableness, there was also a statistically significant difference between SOF-operators ($M = 5.26$, $SD = 0.68$) and NCO-applicants [$M = 5.65$, $SD = 0.68$; $t(736) = -4.72$, $p = 0.00$, two tailed], 95% CI [0.23, 0.55]. Finally, there was a significant difference in mean scores on emotional stability for SOF-operators ($M = 5.54$, $SD = 0.78$) and NCO-applicants [$M = 5.33$, $SD = 0.82$, $t(729) = 2.07$, $p = 0.04$, two tailed], 95% CI [0.01, 0.40]. Cohen's *d* effect sizes were 0.62 for extraversion, 0.57 for agreeableness, and 0.26 for emotional stability.

DISCUSSION

The present study is the first to report on Big Five personality traits of Norwegian SOF personnel, by using the NMPI. The NMPI is a factor-level test demonstrating sound psychometric properties, and strong convergent validity toward the NEO-PI-3 (Antonsen, 2016; Skoglund et al., unpublished). Three research questions investigating the impact of demographic information and different personnel categories on levels of personality traits were formulated. Of special interest were potential personality differences between SOF-operators and other personnel categories, considering the uniqueness in the selection, training and mission forms of the former compared to the latter (Spulak, 2007; NATO, 2012; Berg-Knutsen and Roberts, 2015). Statistically significant findings were found. Care in the interpretation of results is warranted, though, primarily since the actual mean differences observed in scores on personality traits were small. Results demonstrated firstly that younger personnel in the NORSOF were somewhat lower on emotional stability than their older colleagues, and that those without any combat-deployments scored a little higher on agreeableness and slightly lower on emotional stability relative to employees with such experiences. In addition, personnel with officer ranks (OF) reported, to some degree, higher extraversion

compared to specialists (OR). Secondly, there were no significant differences in personality trait scores between SOF-operators from both the FSK and the MJK, and the SOF-support group. Operators from the two units had very similar mean personality trait scores (a mean difference of 0.08 for the five factors), and there were no significant differences in mean trait scores between SOF-operators as a group and SOF-support personnel. Thirdly, compared to applicants for basic officer training (to become an NCO) in the conventional forces, SOF-operators were less extroverted, less agreeable and to a certain extent more emotionally stable.

The results demonstrate an impact tendency of age and number of combat-deployments on emotional stability levels. This points to possible important organizational psychological characteristics within the NORSOF. As older operators and support personnel, as well as those with combat-deployment experiences, reported higher scores on emotional stability relative to younger and less-experienced ones, they can be well-suited role models in a military department where managing stressors is of great importance. Contrary to our findings, Braun et al. (1994) did not, however, report differences in neuroticism based on age or warfare experience among SEALs. The finding regarding age is in line, though, with general research on personality trait change in adulthood, where a decrease in neuroticism is sometimes seen (Caspi et al., 2005; Roberts and Mroczec, 2008).

While it may be obvious that high levels of neuroticism can be counterproductive in stressful settings, high levels of agreeableness have somewhat contradictorily been associated both with less stressors in life (Leger et al., 2016), and with an increase in subjective distress when encountering interpersonal conflicts (Suls et al., 1998). Thus, one can argue that different levels of agreeableness seem both dysfunctional and functional, perhaps depending on the levels of other traits and on contextual factors. It may be that combat-deployments, thus seeing conflict and war up close, can take its toll on the characteristics associated with agreeableness – although causality cannot be inferred from the present study. Trust, altruism and tender-mindedness, the NEO facets correlating most strongly to the NMPI factor agreeableness (Skoglund et al., unpublished), are necessary building blocks for team cohesion, but can be exploited if such tendencies are too strong toward others who have hostile intentions. Those with combat-deployment experiences may have "balanced" their agreeableness more than those who have not been deployed yet, or maybe the more experienced

TABLE 3 | NMPI mean factor scores, standard deviations and independent samples *t*-tests for SOF-operators and male NCO-applicants.

NMPI factor	SOF-operators (*N* = 76)		NCO-applicants (*N* = 648–662)			
	M	*SD*	*M*	*SD*	*t*	Cohen's *d*
Extraversion	4.67	0.83	5.16	0.76	−5.37*	0.62
Agreeableness	5.26	0.68	5.65	0.68	−4.72*	0.57
Conscientiousness	5.65	0.72	5.65	0.69	0.08	0.00
Emotional stability	5.54	0.78	5.33	0.82	2.07*	0.26
Openness	4.87	0.50	4.83	0.61	0.59	0.09

*$p < 0.05$.

personnel have this balance naturally. Braun et al. (1994) did not, however, find any differences in agreeableness based on amount of experience of SEALs, but noted that personality differences based on warfare experience were most likely explained by age. Except for the difference in scores on emotional stability based on age groups, we did not obtain results in line with expectations based on the maturity principle in personality psychology (Caspi et al., 2005). The differences in agreeableness based on number of combat-deployment are somewhat in line, though, with the findings demonstrating that conventional military training can reduce levels of this trait (Jackson et al., 2012).

The finding that officers (OF) reported themselves as slightly more extraverted relative to specialists (OR), gives rise to speculations of a possible functionality of this trait in the selection to the 3-year-long officer education in Norway and the later performance of leadership. More often than not, operators and support personnel in the NORSOF attend the officer education after some years of service as specialists, pointing to a possibility that the extraverted are more likely to become officers, or that the less extraverted prefer to be specialists. This finding is in line with results from Braun et al. (1994), as these authors reported higher degrees of extraversion among SEAL-officers relative to SEAL-enlisted personnel. The comparison to the SEALs, however, is somewhat imprecise as our sample were operators and support personnel combined.

The empirical studies of Bartone et al. (2008) and Garbarino et al. (2012) support the theory of "specialness" of SOFs by demonstrating differences in psychological hardiness and Big Five scores among successful and non-successful SOF-applicants and police Special Force officers and career soldiers, respectively. Comparing operators to population norms, these studies, together with the Braun et al. (1994) study, reported significant differences in Big Five trait levels, further underlining a kind of specialness. The operators of the FSK and the MJK were similar in their Big Five personalities, and they did not differ significantly from support personnel. Thus, there were no clear intradepartmental differences regarding personality trait levels, suggesting that the specialness of SOF-operators was not reflected in their personalities when compared with personnel within the NORSOF.

The present study did, however, support the theory of specialness when comparing SOF-operators with NCO-applicants. A personality profile of the operator emerged, characterized by lower extraversion and agreeableness, and somewhat higher emotional stability, relative to male applicants to basic officer training in conventional forces. The effect sizes in Cohen's *d* were medium for differences in extraversion and agreeableness, and small for the difference in emotional stability. Although not directly comparable to the male population in Norway, these applicants were not militarily experienced individuals, and about half of them did not manage successful NCO-selection. Both SEALs (Braun et al., 1994) and Special Forces police officers (Garbarino et al., 2012) had higher scores on extraversion relative to population norms, and also when compared to career-soldiers for the police operators, which can be said to be opposite to our findings (although we did not compare to population norms). Our finding of lower

agreeableness is in line with Braun et al. (1994), but not with Garbarino et al. (2012). Emotional stability has especially been highlighted as important for high-risk operational personnel (Braun et al., 1994; Bartone et al., 2008; Garbarino et al., 2012; Picano and Roland, 2012), and it may not come as a surprise that the present study found the same trend as earlier empirical studies, demonstrating higher levels of this trait for SOF-operators relative to other samples. The SOF personality profile that emerged is somewhat contradictory on a conceptual level to the character strengths study of Boe et al. (2017), where open-mindedness was evaluated as especially important by FSK-officers, and where humility/modesty, curiosity, and forgiveness and mercy were evaluated as more important by these officers, relative to evaluations done by officers in the conventional Army. It could be that Big Five personality traits and character strengths tap into different psychological phenomena, although this is debated (Peterson and Seligman, 2004).

Braun et al. (1994) suggested a personality description of the "average" SOF-operator (SEAL) based on their findings, resulting in a profile that is intuitively comparable to what one may think is functional for high-risk operational personnel, highlighting such attributes as hardiness and persistence, and some skepticism of others' intentions. Where these authors had access to population norms of the personality test used (the NEO), we did not. Nonetheless, interpreting the findings in the present study, it becomes clear that the Norwegian "tip of the spear" operators typically do not have very low or very high scores on any of the Big Five factors, and that they tend to be less extraverted and agreeable, and more emotionally stable, when compared to those who seek general purpose forces. This profile is not counter-intuitive, and implies an overall flexible personality functioning, drawing a picture of an emotionally stable individual with an initial reserved attitude toward strangers. Considering the future SOF-operator termed the "Warrior-Diplomat" (Berg-Knutsen and Roberts, 2015), operating in diverse contexts, we suggest that this profile is adaptive. Being emotionally stable and somewhat cautious with interpersonal interactions may be seen as functional for serving in a unit operating in high-stress environments, and in which security issues and secrecy are necessary. Raw scores on the interpersonal traits of extraversion and agreeableness were not low, making it reasonable to think of the SOF-operator as socially adept if the circumstances call for diplomatic attitudes.

As the emerged average personality profile of the operator is interpreted as functional, the selection processes in both the FSK and the MJK seem to function well regarding the evaluation of personality characteristics of applicants. The personality tests implemented in the NORSOF-operator selections serve primarily as background materials for military psychologists' advices concerning applicants' strengths and weaknesses. This practice is comparable to the procedure in the Norwegian Police Special Operations Officer selection, as documented by Johnsen (2017). For an optimal evaluation of personality test use, both the possibility of personality change based on military experience (Jackson et al., 2012) and the perspective of trait activation based on situational cues (Tett and Burnett, 2003; Judge and Zapata, 2015)

could be important aspects to consider. The first point could be researched upon with a repeated measures design. The authors suppose, though, that it would be challenging to operationalize a sound situational taxonomy matching SOF job performance, in which a persistent adaptability to unforeseen circumstances, in high stress environments, is the primary attribute to select for. Personality factors associated with this attribute in a selection context are perhaps not obvious, with the exception of broadly formulated characteristics (Picano and Roland, 2012).

Study limitations should be noted. Sampling bias is an unknown factor in this study, due to the confidentiality of the number of employees at the time of data extraction. We do not know the response rate and the exact representativeness of study-participants for the NORSOF as a whole, although we note that conscript personnel were not included in this study. Our final sample size was small, primarily because NORSOF employees are few in numbers and belong to one of the most secretive branches in our society. The results from the analyses, therefore, require cautiousness in interpretation. Our sample is nonetheless valuable for the military psychology literature, considering the scant empirical studies from such highly specialized environments. The personality measure used, the NMPI, is a newly developed Big Five measure for the Norwegian Armed Forces. Although validated with good results, norms are yet to be developed, and operational use of the test in the military has not commenced at the time of writing. Specifically, an unfortunate limitation was the missing possibility of comparative analyses between the NORSOF sample and population norms. Of special note for the present study is the difference in age for the NORSOF sample and the NCO-applicants; ideally, personality comparative data should be on the same age groups, considering possible changes in some personality traits with increased age. The reported personality trait differences would perhaps be more salient, resulting in larger effect sizes, if comparative analyses were done with civilian samples, and population norms. Some range restriction in the personality scores of NCO-applicants is reasonable to assume, as this group was preselected through the conscript assessment procedures in Norway. Demand characteristics could also be relevant for the NMPI scores of NCO-applicants, as the data-collection for this sample was done in a selection context.

CONCLUSION

In conclusion, the findings demonstrated that FSK-operators, MJK-operators and SOF-supports did not differ in a clear way in their group personality trait scores. Therefore, the SOF environment in Norway seems to attract and recruit people that have similarities in their Big Five makeup, although there are different selection systems for the operators and diverse military backgrounds among the supports. The authors did not find this counterintuitive, as being employed by the NORSOF is only for the few – although our expectations of

differences between the operators and the supports were not met. The climate among colleagues is of course dependent on many factors, one being personality variables. We did not find clear personality-based obstacles for cooperation between the departments. Questions related to potential organizational benefits with increased personality diversity in the NORSOF are unanswered based on the present study. This study replicated the most consistent findings regarding the personalities of high-risk operational personnel, namely higher emotional stability among SOF-operators compared to other samples. This finding is, however, somewhat questionable regarding the small effect size ($d = 0.26$), and the age difference between the operators and the NCO-applicants (most operators were in their 30s, whereas the applicants had a mean age of 19–20). More contradictory findings relative to earlier studies are the results of lower agreeableness and extraversion among SOF-operators compared to the NCO-applicants, although this was in line with our expectations. It was surprising, however, that the operators did not score higher than NCO-applicants on conscientiousness – especially since this trait has demonstrated predictive validity for job performance in military settings (Salgado, 1998; Darr, 2009). The expectation of higher openness was not met, perhaps due to construct differences in character strengths and Big Five personality traits. Nonetheless, this study supports that these carefully selected men are somewhat special as aspects of their personalities are a little bit different compared to those who apply for service in general purpose forces. The authors also note that the ability of the NMPI to detect differences between military personnel categories further supports the validity of this test.

Future quantitative studies concerning psychological characteristics in SOF environments should be aware of sampling bias and should acknowledge the importance of statistical power, while balancing this with the fact that samples from these secretive environments can be challenging to obtain. Where possible, studies should compare the personalities of SOF personnel with civilian samples for further investigation of the assumed uniqueness of the operators. Repeated personality measures would be valuable for investigating the possible impact of the SOF experience on personality traits. Based on findings from the present study, controlling for the effect of age, number of combat-deployments and rank can be important for investigating personality questions in SOFs. We suggest that the Big Five personality profile that emerged for the average NORSOF-operator is a functional one considering the Warrior-Diplomat role required in modern Special Forces operations, and that some cautiousness in interpersonal settings is functional when serving as one of "the quiet professionals."

DATA AVAILABILITY STATEMENT

The data analyzed in this study were obtained from the Norwegian Special Operations Command and the Norwegian Defence University College. Due to a confidentiality clause, the

raw data will not be made publicly available. Requests to access these datasets should be directed to the Norwegian Defence University College (mail: post.fhs@mil.no).

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by The Norwegian Social Science Data Service. The patients/participants provided their written informed consent to participate in this study.

AUTHOR CONTRIBUTIONS

TS wrote the manuscript. All authors analyzed and interpreted the data and gave feedback on the manuscript. T-HB designed the study and collected the data.

FUNDING

REFERENCES

Allport, G. W., and Odbert, H. S. (1936). Trait-names: a psycho-lexical study. *Psychol. Monogr.* 47, 1–171. doi: 10.1037/h0093360

Antonsen, I. (2016). *Exploring Aspects of Construct and Criterion Validity of the Norwegian Military Personality Inventory (NMPI).* Master's thesis, Norwegian University of Science and Technology, Trondheim.

Barrick, M. R., and Mount, M. K. (1991). The big five personality dimensions and job performance: a meta-analysis. *Pers. Psychol.* 44, 1–26. doi: 10.1111/j.1744-6570.1991.tb00688.x

Bartone, P., Roland, R. R., Picano, J., and Williams, T. J. (2008). Psychological hardiness predicts success in US Army Special Forces candidates. *Int. J. Select. Assess.* 16, 78–81. doi: 10.1111/j.1468-2389.2008.00412.x

Berg-Knutsen, E., and Roberts, N. (2015). *Strategic design for NORSOF 2025.* Available online at: https://www.ffi.no/no/Publikasjoner/Documents/NORSOF%202025.pdf (accessed October 10, 2019).

Boe, O., Nilsen, F. A., Kristiansen, O., Krogdahl, P., and Bang, H. (2017). "Measuring important character strengths in Norwegian Special Forces officers," in *Proceedinga of the International Conference on Education and New Learning Technologies,* Barcelona.

Braun, D. E., Prusaczyk, W. K., Goforth, H. W. Jr., and Pratt, N. C. (1994). *Personality Profiles of U.S. Navy Sea-Air-Land (SEAL) Personnel.* San Diego, CA: Naval Health Research Center.

Campbell, J. P., and Knapp, D. J. (eds) (2001). *Exploring the Limits Inpersonnel Selection and Classification.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Caspi, A., Roberts, B. W., and Shiner, R. L. (2005). Personality development: stability and change. *Annu. Rev. Psychol.* 56, 453–484. doi: 10.1146/annurev.psych.55.090902.141913

Cohen, J. W. (1988). *Statistical Power Analysis for the Behavioral Sciences,* 2nd Edn. Hillsdale, NJ: Lawrence Erlbaum Associates.

Costa, P. T., and McCrae, R. R. (1992). *NEO PI-R. Professional Manual.* Odessa, FL: Psychological Assessment Resources.

Darr, W. (2009). *Big Five Personality Research in the Military: A Meta-Analysis.* College Station, TX: Texas A&M university.

Digman, J. M. (1990). Personality structure: emergence of the five-factor model. *Annu. Rev. Psychol.* 41, 417–440. doi: 10.1146/annurev.ps.41.020190.002221

Faul, F., Erdfelder, E., Buchner, A., and Lang, A.-G. (2009). Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. *Behav. Res. Methods* 41, 1149–1160. doi: 10.3758/BRM.41.4.1149

Garbarino, S., Chiorri, C., Magnavita, N., Piattino, S., and Cuomo, G. (2012). Personality profiles of Special Force police officers. *J. Police Crim. Psychol.* 27, 99–110. doi: 10.1007/s11896-011-9099-6

Hartmann, E., and Grønnerød, C. (2009). Rorschach variables and Big Five scales as predictors of military training completion: a replication study of the selection of candidates to the naval Special Forces in Norway. *J. Pers. Assess.* 91, 254–264. doi: 10.1080/00223890902794309

Hartmann, E., Sunde, T., Kristensen, W., and Martinussen, M. (2003). Psychological measures as predictors of military training performance. *J. Pers. Assess.* 80, 87–98. doi: 10.1207/S15327752JPA8001_17

Hedenstrom, A., and Kristiansen, M. (2016). *NORSOF Military Assistance Capability Development.* Master's thesis, Naval Postgraduate School, Monterey, CA.

Jackson, J. J., Thoemmes, F., Jonkmann, K., Lüdke, O., and Trautwein, U. (2012). Military training and personality trait development: does the military make the man, or does the man make the military? *Psychol. Sci.* 23, 270–277. doi: 10.1177/0956797611423545

John, O. P., Naumann, L. P., and Soto, C. J. (2008). "Paradigm shift to the integrative Big-Five trait taxonomy: history, measurement, and conceptual issues," in *Handbook of Personality: Theory and Research,* eds O. P. John, R. W. Robins, and L. A. Pervin (New York, NY: Guilford Press), 114–158.

Johnsen, B. H. (2017). "Selection of police special operations officers: the role of the psychologist," in *Handbook of Military Psychology,* eds S. Bowles and P. T. Bartone (New York, NY: Springer International Publishing), 291–299. doi: 10.1007/978-3-319-66192-6_18

Judge, T. A., and Zapata, C. P. (2015). The person-situation debate revisited: effect of situation strength and trait activation on the validity of the big five personality traits in predicting job performance. *Acad. Manag. J.* 58, 1149–1170. doi: 10.5465/amj.2010.0837

Leger, K. A., Charles, S. T., Turiano, N. A., and Almeida, D. M. (2016). Personality and stressor-related affect. *J. Pers. Soc. Psychol.* 111, 917–928. doi: 10.1037/pspp0000083

McCrae, R. R., and Costa, P. T. (2010). *NEO Inventories Professional Manual.* Odessa, FL: Psychological Assessment Resources.

McCrae, R. R., Costa, P. T., and Martin, T. A. (2005). The NEO-PI-3: a more readable revised NEO personality inventory. *J. Pers. Assess.* 84, 261–270. doi: 10.1207/s15327752jpa8403_05

NATO (2012). *Psychological and Physiological Selection of MILITARY SPECIAL Operations Forces Personnel.* Brussels: NATO.

Olsen, T., and Thormodsen, M. (2014). *Forging Norwegian Special Operation Forces.* Master's thesis, Naval Postgraduate School, Monterey, CA.

Peterson, C., and Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification.* New York, NY: Oxford University Press.

Picano, J. J., and Roland, R. R. (2012). "Assessing psychological suitability for high-risk military jobs," in *The Oxford Handbook of Military Psychology,* eds J. H. Laurence and M. D. Matthews (New York, NY: Oxford University Press), 148–157.

Roberts, B. W., and Mroczec, D. (2008). Personality trait change in adulthood. *Curr. Direct. Psychol. Sci.* 17, 31–35. doi: 10.1111/j.1467-8721.2008.00543.x

Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civil occupations: a European perspective. *Hum. Perform.* 11, 271–288. doi: 10.1207/s15327043hup1102\&3_8

Schmidt, F., and Hunter, J. (1998). The validity and utility of selection methods in personnel psychology: practical and theoretical implications of 85 years of research findings. *Psychol. Bull.* 124, 262–274. doi: 10.1037/0033-2909.124.2.262

Scroggins, W. A., Thomas, S. L., and Morris, J. A. (2008). Psychological testing in personnel selection, part I: a century of psychological testing. *Public Pers. Manag.* 37, 99–109. doi: 10.1177/009102600803700107

Skoglund, T. H. (2017). The norwegian military personality inventory (NMPI): conscript assessment. *Paper presented at the International Military Testing Association Conference,* (Bern: CRISTIN).

Spulak, R. G. Jr. (2007). *A Theory of Special Operations (JSOU Report 07-7).* Hurlburt Field, FL: Joint Special Operation University.

Stark, S., Drasgow, F., White, L. A., Heffner, T., Chernyshenko, O. S., Nye, C. D., et al. (2014). From ABLE to TAPAS: a new generation of personality tests to support military selection and classification decisions. *Military Psychol.* 26, 153–164. doi: 10.1037/mil0000044

Suls, J., Martin, R., and David, J. P. (1998). Person-environment fit and its limits: agreeableness, neuroticism, and emotional reactivity to interpersonal conflict. *Pers. Soc. Psychol. Bull.* 24, 88–98. doi: 10.1177/0146167298241007

Tett, R. P., and Burnett, D. D. (2003). A personality trait-based interactionist model of job performance. *J. Appl. Psychol.* 88, 500–517. doi: 10.1037/0021-9010.88.3.500

USSOCOM (2019). *Fact book United States Special Operations Command.* Tampa: USSOCOM.

Widiger, T. A. (ed.) (2017). *The Oxford Handbook of the Five Factor Model.* New York, NY: Oxford University Press.

Study 3

Personality and Social Psychology

The psychometric properties and norm data of the Norwegian military personality inventory (NMPI)

MORTEN NORDMO,[1] TOM H. SKOGLUND,[2] OLE C. LANG-REE,[1] SARA K. AUSTAD[1] and MONICA MARTINUSSEN[1,3]

[1]*Norwegian Armed Forces Joint Medical Services, Sessvollmoen, Norway*
[2]*Department of Psychology, Bjørknes University College, Oslo, Norway*
[3]*RKBU Nord, UiT The Arctic University of Norway, Tromsø, Norway*

Nordmo, M., Skoglund, T. H., Lang-ree, O. C., Austad, S. K. & Martinussen, M. (2021). The psychometric properties and norm data of the Norwegian military personality inventory (NMPI). *Scandinavian Journal of Psychology*.

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 olds ($N = 54,355$), 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.

Key words: Personality, big-five, psychometrics, personnel selection, military motivation.

Morten Nordmo, Norwegian Armed Forces Joint Medical Services, Norway. E-mail: mortennordmo@gmail.com

INTRODUCTION

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, Campion, Dipboye, Hollenbeck, Murphy & Schmitt, 2007a, 2007b). 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 US 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 & Føllesdal, 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 *et al.* 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. Online Appendix S1 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 a priori five-factor solution, bi-factor solution and a three-factor solution based on the Eysenck assertion of broad traits (Eysenck & 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 < 0.08. 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 < 0.08 based on recommendations from Hu and Bentler (1999). CFI and TLI cut-off values were set at >0.90 (Hooper, Coughlan & Mullen, 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, 2019). 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, Drasgov & Parsons, 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, Schmidt, De Beuckelaer, Lek & Zondervan-Zwijnenburg, 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 (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, Guo, Parker *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.

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 online Appendix S1 (Fig. 1). 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

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)	**0.91**	0.52*	0.23*	0.31*	0.44*	–
2. Extroversion	42.66 (12.31)	0.47*	**0.90**	0.63*	0.62*	0.61*	–
3. Openness to experience	46.10 (10.74)	0.11*	0.42*	**0.85**	0.68*	0.62*	–
4. Agreeableness	53.69 (10.12)	0.18*	0.48*	0.56*	**0.91**	0.72*	–
5. Conscientiousness	49.07 (10.75)	0.35*	0.47*	0.48*	0.61*	**0.90**	–
6. Motivation for military service	2.70 (1.52)	0.46*	0.41*	0.30*	0.38*	0.44*	–

Notes: Chronbach's alfa in bold along the diagonal.
*p < 0.01(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	0.05	0.25	0.05	−0.03	**0.42**
o2	Desire for self-development	**0.31**	0.30	0.15	0.02	0.11
o3	Takes time to reflect	0.10	0.09	0.03	−0.31	**0.40**
o4	Fascinated by shapes and colors	−0.06	0.08	−0.13	−0.06	**0.65**
o5	Vivid imagination	−0.12	0.06	−0.02	0.11	**0.74**
o6	Generates ideas easily	0.14	−0.01	0.11	0.12	**0.61**
o7	Familiar with many concepts	**0.31**	0.03	0.15	−0.01	0.25
o8	Appreciates art	−0.16	0.05	−0.11	−0.03	**0.70**
o9	Thinks creatively	0.06	0.03	0.00	0.16	**0.71**
o10	Curious about other cultures	−0.01	0.20	0.04	0.00	**0.44**
e1	View myself as outgoing	−0.01	0.31	**0.63**	0.03	−0.14
e2	Frequently becomes a leader	0.29	−0.12	**0.64**	−0.06	0.07
e3	Likes being the center of attention	−0.09	0.14	**0.76**	−0.09	−0.02
e4	Do not like attention (r)	−0.23	0.09	**0.63**	0.14	−0.03
e5	Likes to make decisions	0.30	−0.13	**0.62**	−0.06	0.12
e6	Starts conversations	−0.00	0.28	**0.59**	0.01	0.04
e7	Keeps in the background (r)	−0.15	0.07	**0.63**	0.23	−0.14
e8	Dominant and self-confident	0.14	−0.12	**0.59**	0.08	0.10
e9	Is introverted (r)	−0.19	0.18	**0.56**	0.17	−0.23
e10	Enjoys leadership	0.28	−0.16	**0.66**	−0.06	0.15
n1	Easily stressed (r)	0.10	0.04	−0.05	**0.65**	0.09
n2	Worries a lot (r)	0.00	0.07	−0.05	**0.76**	0.01
n3	Fear embarrassment (r)	−0.10	−0.02	0.19	**0.61**	0.11
n4	Thinks about other's view of me (r)	−0.09	−0.10	−0.03	**0.68**	0.05
n5	Feelings of guilt (r)	0.06	−0.14	−0.04	**0.64**	−0.03
n6	Frequently sad (r)	0.13	0.11	−0.03	**0.64**	−0.11
n7	Prone to nervousness (r)	−0.00	−0.03	0.15	**0.71**	0.08
n8	Feelings of inferiority (r)	0.10	−0.15	0.15	**0.57**	0.01
n9	Feel tense (r)	0.02	0.03	0.00	**0.69**	−0.05
n10	Frequently scared (r)	0.09	0.08	0.06	**0.62**	0.02
c1	True to appointments	**0.61**	0.20	−0.12	0.05	−0.11
c2	Reaches goals	**0.57**	0.08	0.17	0.09	−0.03
c3	Attention to details	**0.58**	−0.05	−0.01	−0.12	0.18
c4	Punctual person	**0.72**	0.14	−0.07	−0.02	−0.07
c5	Systemizes frequently	**0.72**	0.01	0.00	−0.04	0.09
c6	Completes duties diligently	**0.71**	0.12	−0.09	0.08	−0.09
c7	Generally prepared	**0.73**	0.15	−0.05	0.04	−0.05
c8	Very industrious	**0.70**	0.12	−0.03	0.01	−0.07
c9	Surrounded by clutter (r)	**0.50**	0.02	−0.06	0.23	−0.15
c10	Sets goals and works to reach them	**0.70**	−0.01	0.09	0.02	0.02
a1	Understands others' needs	0.23	**0.44**	0.09	0.05	0.10
a2	Attention to the wellbeing of others	0.13	**0.63**	0.02	0.01	−0.04
a3	Something nice to say about everyone	0.11	**0.61**	−0.02	0.02	0.07
a4	Takes the time to listen to others	0.14	**0.66**	0.08	−0.01	0.02
a5	Interested in other people	0.02	**0.58**	0.16	−0.03	0.11
a6	Easily cares for others	−0.06	**0.64**	0.20	−0.15	0.03
a7	Very helpful	0.27	**0.61**	−0.04	0.02	0.07
a8	Compassionate	−0.01	**0.71**	−0.04	−0.13	0.12
a9	Likes to help others	0.11	**0.69**	−0.05	0.01	0.12
a10	Trusts others	0.09	**0.60**	−0.05	0.11	0.01

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

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.

Correlations from latent traits after CFA are shown in Table 1. 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 3 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

Table 3. *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	0.00	0.067	0.087	0.819	0.809
Three-factor solution	269951	1172	0.00	0.092	0.121	0.661	0.646
Bi-factor solution	69018	1115	0.00	0.058	0.062	0.875	0.862
Extroversion	23727	35	0.00	0.158	0.076	0.834	0.786
Openness	22444	35	0.00	0.153	0.088	0.788	0.728
Neuroticism	8400	35	0.00	0.094	0.040	0.941	0.924
Conscientiousness	7393	35	0.00	0.087	0.038	0.944	0.922
Agreeableness	4005	35	0.00	0.065	0.027	0.975	0.961

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

Table 4. *Bivariate correlations among the five personality domains and motivation for military service in a bi-factor model*

Variable	1	2	3	4	5	6
1. Emotional stability	–					
2. Extroversion	0.19*	–				
3. Openness to experience	−0.07*	−0.10*	–			
4. Agreeableness	−0.08*	0.04*	0.17*	–		
5. Conscientiousness	0.12*	−0.11*	0.13*	0.27*	–	
6. Motivation for military service	0.54*	0.03*	0.09*	0.16*	0.28*	–

*p < 0.01 (two-tailed).

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 = 0.05, SRMR = 0.05, CFI = 0.91 & TLI = 00.90). 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 = 0.44, p < .001, 95% CI = 0.43 –0.45). 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-consciously 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 4. The strength of the relationship increases for emotional stability, and decreases for the four 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.

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 3. 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.

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, Baron & Campbell, 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 ($\chi^2 = 11822(24)$, RMSEA = 0.097, SRMR = 0.031, TLI = 0.924 and CFI = 0.959). 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 = 0.55). 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 ($\chi^2 = 16150$ (24), RMSEA = 0.097, SRMR = 0.031, TLI = 0.924 and CFI = 0.959). 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 and 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 ($B = 0.49$) 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 5. 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, that is, 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 n3 (*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*). Figs. IIa–e shows individual item characteristic curves and can be viewed in online Appendix S2.

Graphs of the overall item information can be seen in Fig. 1. All five scales yield more information at lower levels of theta (θ) values. The emotional stability scale achieves SE > 0.5 from $\theta = -3.1$ to 2.3. The openness scale demonstrates the narrowest range of acceptable test information, from $\theta = -2.9$ to 1.7. In contrast, the agreeableness shows the largest range of acceptable test information from $\theta = -3.9$ to 2.0. The conscientiousness scale achieves SE > 0.5 from $\theta = -3.6$ to 2.1, and the extroversion scale from $\theta = -2.9$ 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.

Observed, metric and scalar gender differences

The gender differences in observed scores can be viewed in Table 6 and Fig. 2. We found that the mean female agreeableness

Table 5. *Discrimination (α) and difficulty (β) parameter estimates for the NMPI-50 based on five graded response models*

Item	α	$\beta1$	$\beta2$	$\beta3$	$\beta4$	$\beta5$	$\beta6$
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	−1.76	−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	−1.52	−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.

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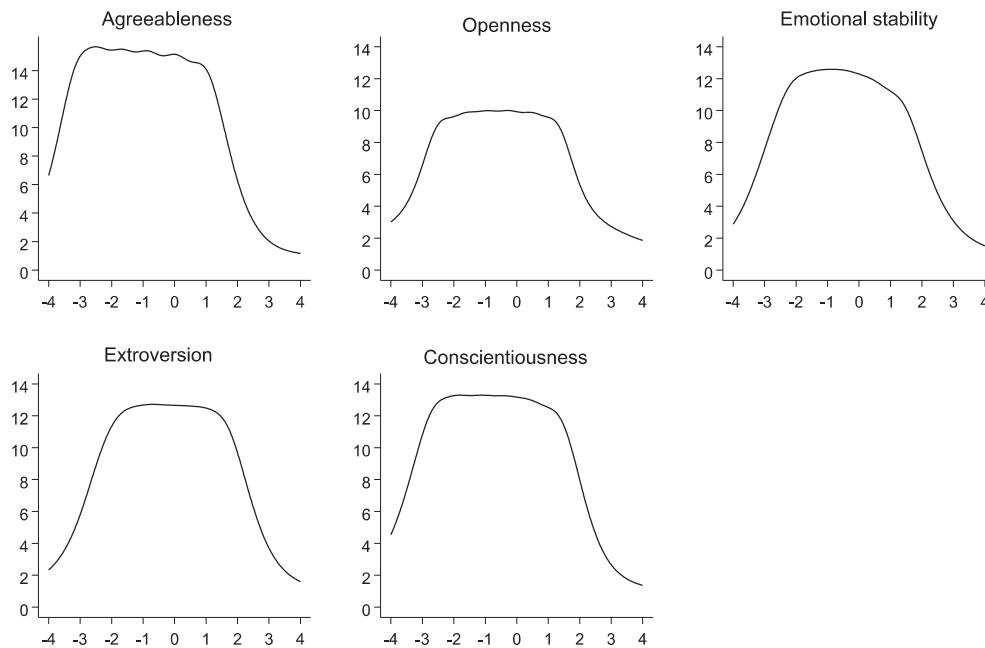
Fig. 1. Total test information curves over theta after the graded response models across five traits. Total information is the sum of information across all trait levels and response options for each item.

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

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.

The means, standard deviations, and data distributions (as seen in Fig. 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, Hřebíčková, Kuppens, Realo & Allik, 2013). Research using non-self-report measures, that is, 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).
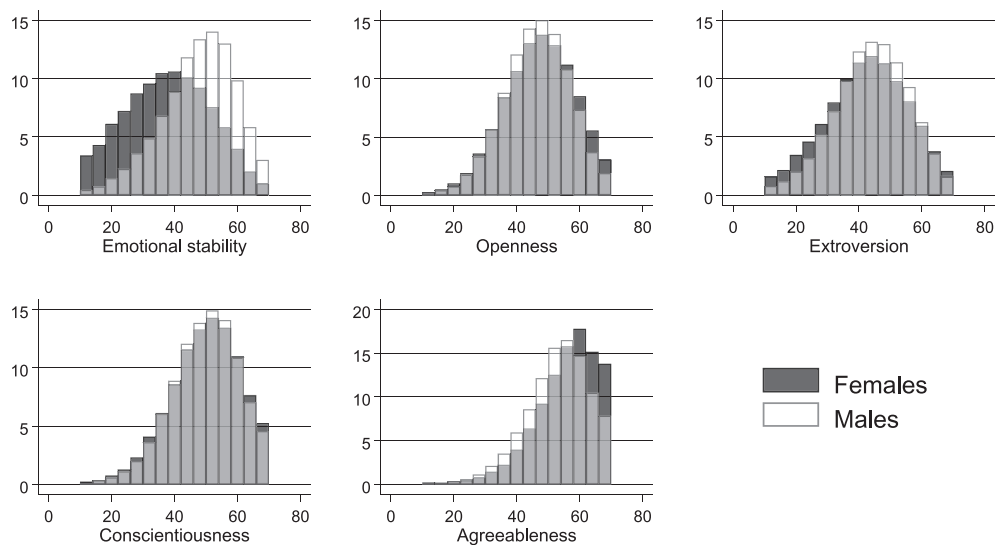

Fig. 2. Histogram of percentage of un-adjusted observed scores on personality factors over gender as measured by the NMPI-50.

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	0.936	0.918	0.095	–
	Full metric	17393	79	0.934	0.924	0.091	Yes
	Full scalar	19476	88	0.926	0.924	0.091	Yes
Conscientiousness	Baseline	14775	70	0.944	0.928	0.089	–
	Full metric	14883	79	0.943	0.935	0.084	Yes
	Full scalar	17275	88	0.934	0.933	0.086	Yes
Openness	Baseline	39516	70	0.809	0.754	0.146	–
	Full metric	40452	79	0.804	0.777	0.139	Yes
	Full scalar	52776	88	0.745	0.739	0.151	No
	Partial scalar	46183	86	0.777	0.766	0.143	No
Extroversion	Baseline	49488	70	0.830	0.782	0.164	–
	Full metric	50076	79	0.828	0.804	0.155	Yes
	Full scalar	52522	88	0.820	0.816	0.150	Yes
Agreeableness	Baseline	7969	70	0.972	0.964	0.068	–
	Full metric	8009	79	0.972	0.968	0.062	Yes
	Full scalar	11678	88	0.959	0.959	0.071	No
	Partial scalar	9799	87	0.967	0.966	0.064	Yes

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

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.

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.

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

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

[a]*Notes*: Based on Cheung and Rensvold (2002) CFI criteria.
*SE < 0.5.

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, Watson, Biderman & Ghorbani, 2015). 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.,* 2015) as well as adaptive social functioning and lack of maladjustment (Watson, Morris & Hood, 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, Morse, de Vries, MacCann & Marty, 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, Björklund & Larsson, 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, Lüdtke, Muthén *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, Shanahan & Bauer, 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, Smillie & Corr, 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, Wennberg & Bergman, 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 alpha, 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, Aluja & García, 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, Morin & St-Hilaire, 2013). The mean differences in personality factors between men and women were similar to results from comparable cultural samples (Källmen *et al.,* 2011; Martinsen, Nordvik & Østbø, 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.

DISCLAIMER

The views expressed are solely those of the authors and do not reflect the policy or position of the Norwegian Armed Forces or Norwegian government.

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*, 669–684.

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*, 335–344.

Baker, F. B. (2001). *The basics of item response theory*. College Park, MD: 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*, 74–84.

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.

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*, 49–60.

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*, 135–144.

Brannick, M. T. (1995). Critical comments on applying covariance structure modeling. *Journal of Organizational Behavior, 16*, 201–213.

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*, 127–143.

Campbell, J. P. & Knapp, D. J. (Eds.) (2001). *Exploring the limits in personnel selection and classification*. Mahwah, NY: 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*, 216–243.

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

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*, 688–691.

Costa, P. 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*. Lutz, FL: Psychological Assessment Resources.

Darr, W. (2011). Military personality research: A meta-analysis of the Self Descriptive Inventory. *Military Psychology, 23*, 272–296.

del Barrio, V., 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*, 677–681.

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

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: Oxford University Press.

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

Engvik, H. & Føllesdal, H. (2005). The Big Five Inventory på Norsk. *Tidsskrift for Norsk Psykologforening, 42*, 128–129.

Eysenck, H. J. & Eysenck, S. B. G. (1975). *Manual of the Eysenck Personality Questionnaire*. London, UK: 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*, 789–796.

FHI. (2018). *Medisinsk fødselsregister—Statistikkbank*. Retrieved from http://statistikkbank.fhi.no/mfr/. Last Accessed June 18, 2020.

Geldhof, G. J., Preacher, K. J. & Zyphur, M. J. (2014). Reliability estimation in a multilevel confirmatory factor analysis framework. *Psychological Methods, 19*, 72–91.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American Psychologist, 48*, 26–34.

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*, 430–450.

Hambleton, R. K. & Rogers, H. J. (1989). Solving criterion-referenced measurement problems with item response models. *International Journal of Educational Research, 13*, 145–160.

Hooper, D., Coughlan, J. & Mullen, M. (2007). Structural equation modeling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*, 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*, 332–346.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika, 30*, 179–185.

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*, 1–55.

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

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*, 311–314.

Kelloway, E. K. (1995). Structural equation modelling in perspective. *Journal of Organizational Behavior, 16*, 215–224.

Køber, 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*, 560–569.

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*, 739–743.

MacCallum, R. C., Browne, M. W. & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1*, 130–149.

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*, 524–545.

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*, 471–491.

Martinsen, Ø. L. N. H., Nordvik, H. & Østbø, L. (2005). Norske versjoner av NEO PI-R og NEO FFI. *Tidsskrift for norsk psykologforening, 42*, 421–423.

McCrae, R. R. (2010). The place of the FFM in personality psychology. *Psychological Inquiry, 21*, 57–64.

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

McCrae, R. R. & Costa, P. T. (2007). Brief versions of the NEO PI-3. *Journal of Individual Differences, 28*, 116–128.

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

Mehmetoglu, M. (2019). FSDET: Stata module to compute Factor score determinacy coefficient. Statistical Software Components S457982. Boston College Department of Economics.

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*, 1029–1049.

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*, 683–729.

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* (pp. 39–70). New York, NY: Guilford Press.

Ones, D. S., Dilchert, S., Viswesvaran, C. & Judge, T. A. (2007). In support of personality assessment in organizational settings. *Personnel Psychology, 60*, 995–1027.

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: 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.

Reise, S. P., Bonifay, W. E. & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*, 129–140.

Salgado, J. F. (1997). The five factor model of personality and job performance in the European community. *Journal of Applied Psychology*, *82*, 30–43.

Salgado, J. F. (1998). Big Five personality dimensions and job performance in army and civilian occupations: A European perspective. *Human Performance*, *11*, 271–288.

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*, 910–932.

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*, 262–274.

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* (Chapter 17, pp. 381–402). Oxford: 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*, https://doi.org/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*, 4–70.

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*, 337.

Widiger, T. A. (Ed.) (2017). *The Oxford handbook of the Five Factor Model*. Oxford: 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*, 478–497.

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

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article:

Appendix S1
Appendix S2