Running Head: SELECTION OF CONSCRIPTS

Predicting Basic Military Performance for Conscripts in the Norwegian Armed Forces

Petter Kristian Køber,[a] Ole Christian Lang-Ree,[b] Kari V. Stubberud,[c] & Monica Martinussen[d]

a Norwegian Defence Research Establishment (FFI)

b The Norwegian Military Defence University College

c The Norwegian Armed Forces' Personnel and Conscription Centre (FPVS)

[d] The Norwegian Military Defence University College and UiT – The Arctic University of Norway

Author note

Correspondence concerning this article should be addressed to Petter Kristian Køber, Norwegian Defence Research Establishment (FFI), P.O. Box 25, N-2027 Kjeller, NORWAY; e-mail: petter-kristian.kober@ffi.no

Some preliminary findings from this study were presented at the 57th International Military Testing Association conference in Stockholm, 21–25 September 2015.

Abstract

The Norwegian Armed Forces have made major changes to the enlistment and selection system for conscripts in recent years. In this paper, the predictive validity of various selection criteria for Military Performance is examined. The sample consisted of 3,276 conscripts, of whom 18 percent were female. The predictors in the analysis were General Mental Ability, Self-Perceived Physical Fitness, Social and Life Skills, Self-Perceived Military Fit, Tested Physical Fitness and Officer-Rated Suitability. Military Performance was assessed by an officer towards the end of the basic one-year military service. Bivariate correlations between the predictors and Military Performance were studied for men and women separately. The best predictors were Tested Physical Fitness for men and Officer-Rated Suitability for women. A step-wise hierarchical moderated multiple regression analysis was conducted. A small, but significant part of the variance was explained by the model. Both self-reported variables and other predictors made a small, but significant contribution to improving the model. The results indicated that the two-step selection process was valid for predicting military performance for both men and women.

Keywords: conscripts, selection, predictive validity, military performance

Public Significance Statement

The Norwegian Armed Forces have introduced a new enlistment and selection system for conscripts. This study examines how well selection criteria such as general mental ability, self-reported measures, physical fitness and officer-rated suitability predict performance after one year of service. The selection system works for both women and men, which is important to ensure that the best candidates are selected.

Predicting Basic Military Performance for Conscripts in the Norwegian Armed Forces

The selection and placement of military personnel by means of psychological tests have traditions that go back to World War I in many countries (Rumsey, 2012; Rumsey & Arabian, 2014). Various tests have been developed and validated over the years for measuring both General Mental Ability (GMA) and more specific cognitive abilities (Rumsey, 2012). Cognitive ability tests have also been widely used for personnel selection to civilian occupations. Several large-scale meta-analyses have documented the predictive validity of such measures over different countries, occupations and types of criteria (Salgado & Anderson, 2003; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Schmidt & Hunter, 1998). The mean operational validity coefficients (corrected for range restriction and criterion reliability) have been around .50 (Bertua, Anderson, & Salgado, 2005; Salgado et al., 2003; Schmidt & Hunter, 1998).

There have been similar findings for the cognitive ability tests used for military training, e.g., pilots (Carretta, Teachout, Ree, Barto, King, & Michaels, 2014; Martinussen & Torjussen, 2004) where tests measuring both GMA and more specific cognitive abilities have predicted flying performance. Corresponding results have been found for selection and assignment to the US military (Drasgow, Embretson, Kyllonen, & Schmitt, 2006; Held, Carretta, & Rumsey, 2014), and to the Canadian forces (Campbell & Catano, 2004). In general, test batteries such as the Armed Services Vocational Aptitude Battery (ASVAB) or the Canadian Forces Aptitude Test (CFAT) are considered to be good measures of GMA, where some incremental validity from adding more specific abilities to GMA has been documented (Campbell & Catano, 2004; Carretta, et al., 2014; Earles & Ree, 1992; Rumsey & Arabian, 2014).

Predictors measuring personality traits, motivation, stress tolerance, or other types of non-cognitive skills have resulted in more mixed findings (Barrick & Mount, 1991; Guion & Gottier, 1965; Rumsey & Arabian, 2014). There is evidence that personality traits (Carretta et al., 2014), hardiness (Bartone, Roland, Picano, & Williams, 2008; Thomassen, Hystad, Johnsen, Johnsen, Laberg, & Eid, 2015), and motivation (Gubata, Oetting, Weber, Feng, Cowan, & Niebuhr, 2012) all have predictive validity for military performance. Other validation studies have made an argument for using types of criteria other than training and work performance, such as psychological fitness and service completion, and that predictors other than GMA may be needed for predicting these (Bäccman, Berggren, & Norlander, 2012). The AIM (Assessment of Individual Motivation) predicted attrition and mental health disorders during the first year of service (Gubata et al., 2012). The TAPAS (Tailored Adaptive Personality Assessment System) was developed to improve the selection of new military recruits by predicting job effort, physical fitness, and drive to perform (Niebuhr, Gubata, Oetting, Weber, Feng, & Cowan, 2013). The TAPAS battery was examined in a large-scale study of US Army recruits. The results indicated that one of the dimensions – Physical Conditioning – predicted a mental disorder diagnosis and early discharge. In a study of Finnish conscripts, the best predictors of military adjustment were personal factors such as Acceptance of Authority, Sociability, Affective Commitment and Physical Health, whereas background factors such as physical and mental health problems and lack of schooling and motivation were more predictive of early attrition (Salo, 2008). A study of Swedish peace-keeping soldiers indicated that physical health and some scales measuring mental health predicted Military Capacity as rated by officers, whereas Sense of Coherence and some of the personality traits predicted Civil Adjustment after operations (Bäccman, Berggren, & Norlander, 2012).

Another non-cognitive measure often used in military selection is physical fitness. Several European and US studies have indicated that self-reported and measured physical fitness may be of importance for basic military training and service completion (Niebuhr, Scott, Powers, Li, Han, Millikan, & Krauss, 2008; Taanila, Hemminiki, Suni, Pihlajamäki, & Parkkari, 2011; Trone, Cipriani, Raman, Wingard, Schaffer, & Macera, 2013). In Finland, based on a study of military conscripts, both self-assessed health and a running test predicted medical discharge from service (Taanila, et al., 2011). Similar findings were detected for US Army enlistees examining an objective test of physical fitness, which predicted attrition during initial entry training (Niebuhr et al., 2008). Low levels of self-reported physical activity was associated with poor training outcomes among US Navy recruits (Trone et al., 2013).

The majority of validation studies for military selection have been conducted on male-majority samples. The possibility of test bias in terms of differential prediction of sub-group performance has rarely been examined (Cleary, 1968; EFPA, 2013). Some studies have indicated a bias (Berry, Clark, & McClure, 2011; Roberts & Skinner, 1996), whereas other studies have not been able to detect this (Carretta, 1997).

Military selection in Norway

In Norway, tests were developed and introduced for military selection in the years following World War II (Torjussen & Hansen, 1999). These covered the selection of conscripts, officers, pilots, and personnel for the Special Forces (Martinussen & Torjussen, 2004; Torjussen & Hansen, 1999). The selection has relied heavily on the use of cognitive ability tests, but also on physical and medical requirements. For the selection of conscripts, three basic cognitive ability tests were developed and used from 1954, measuring general reasoning ability, numerical skills, and verbal ability (Hansen, 2006). These paper-and-pencil

tests were later converted to a computer-based test battery, which was implemented from 2011 (Skoglund, Martinussen, & Lang-Ree, 2014). Norway has had conscription for all men over 18 years old since 1953, and from 2015 this has applied to women as well. However, the needs and tasks of the Armed Forces have changed over the years, resulting in a general downsizing, and also a reduced need for conscripts.

A new enlistment and selection system was introduced in 2010. It is a two-step process with an initial screening, Selection Part 1, followed by a more comprehensive testing, Selection Part 2. Selection Part 1 is an online questionnaire, which all Norwegian men and women are obliged to answer. Selection Part 2 is a one-day muster, with a full medical examination, physical capacity tests, cognitive ability tests, and an interview with a selection officer. The selection process is outlined in Figure 1.

Less than 15 percent of each annual age cohort is required to serve in the Norwegian Armed Forces. This means that the selection of conscripts is essential for the Norwegian Armed Forces, not only to fill the short-term need for conscripts, but also to ensure the long-term recruitment of military officers. The inclusion of women is expected to increase the percentage of women serving in the Norwegian Armed Forces. Hence, it is of particular interest whether there are any gender differences in the predictive validity of the selection criteria.

Insert Figure 1 here

The current study

Because of the new military demands facing the Armed Forces and the changes to the Norwegian enlistment and selection system, it was important to assess the predictive validity of the various selection criteria used for basic military service. The purpose of this study was thus, firstly, to examine the predictive validity of GMA and other selection criteria; secondly, to estimate a model for the military performance of conscripts in the Norwegian Armed Forces based on all available predictors; and, thirdly, to examine gender differences in the validity of individual predictors and of the combined model. The predictive validity of the selection criteria for the completion of basic military service had been examined in a previous study, with motivation to serve emerging as the most important criterion (Køber, 2015).

Predictors included not only GMA tests, but also other predictors such as physical fitness, social and life skills, self-perceived military fit and the selection officer's rating of the candidates. Based on previous findings (Bertua et al., 2005; Carretta, et al., 2014; Salgado et al., 2003; Schmidt & Hunter, 1998) it was expected that GMA tests would predict military performance during the first year of service. Findings from Finland and Norway had indicated that social skills might be important for completing service (Salo, 2008; Køber, 2015), but might not necessarily be as important for predicting performance in the way it is examined in this study. Finally, it was expected that physical fitness would predict performance, as in previous studies from Finland and the US (Niebuhr et al., 2008; Taanila, et al., 2011; Trone, et al, 2013). The selection officer's rating of the candidates was based on a semi-structured interview. Based on previous meta-analyses assessing the validity of such interviews in general (Schmidt & Hunter, 1998), and findings from the selection of pilots for the Norwegian Air Force (Martinussen & Torjussen, 2004), it was expected that interview ratings would have some predictive validity for military performance. Previous studies show that

rating measures often are better predicted by non-cognitive measures (Campbell & Knapp, 2001).

Method

Participants and procedure

The study looked at 3,276 conscripts, born in 1994, who had completed the basic one-year of military service. A total of 592 of the participants (18 percent) were female.

The participants in the study had completed Selection Part 1 in 2011. Most of the participants took part in Selection Part 2 and were selected for military service between August 2012 and June 2013. All of the participants had initial enrollment for military service between July and October 2013. Military performance was assessed towards the end of their service in 2014.

Measures

General Mental Ability (GMA). GMA was measured with a computer-based test battery at Selection Part 2 (Skoglund et al., 2014; Torjussen & Hansen, 1999). The test battery consisted of three different sub-tests measuring Numerical Ability (30 items in 25 minutes), General Reasoning Ability (36 items in 20 minutes) and Word Similarities (54 items in 6 minutes). For all tests, the number of correct answers was recorded and transformed into a stanine score. The total score was calculated as the mean stanine score of the three tests. Only the overall score was available for this study. The test–retest reliability r_{XX} for each of the three sub-tests was .84, .72, and .90, respectively (Sundet, Tambs, Magnus, & Berg, 1988).

Self-Perceived Physical Fitness. The degree of Self-Perceived Physical Fitness was measured through three items from Selection Part 1: frequency of training, physical endurance, and physical strength. Frequency of training was measured on a four-point scale: "less than once a week," "once or twice a week," "3 or 4 times a week," and "5 times a week or more." The answers were transformed into a numeric scale ranging from 1 to 4. Physical endurance and physical strength were assessed compared to peers of the same gender and age. Answers were on a five-point Likert scale, ranging from "much worse" to "much better". The answers were then transformed into a numeric scale ranging from 1 to 4, with the answers "worse" and "much worse" both assigned the lowest value. The total score for Self-Perceived Physical Fitness was the sum of the three items, ranging from 3 to 12. Cronbach's alpha for this scale has, based on the current sample, been estimated at .67 and .70 for men and women, respectively.

Social and Life Skills. The level of Social and Life Skills was measured with 13 items from Selection Part 1. The questions measuring these skills had been in use since 2009, and were based on a template used by the Swedish Armed Forces (The National Service Administration, 2007). Examples of items were: "I like to take responsibility," "I am generally at ease with my peers at school and in my spare time," and "I cope well with stress and deadlines." The answers were on a four-point Likert scale ranging from "strongly disagree" to "strongly agree." The answers were then transformed into a numeric scale 1–4. The total score for Social and Life Skills was the sum of the scores for the 13 items, ranging from 13 to 52. Cronbach's alpha for the scale has, based on the current sample, been estimated at .74 for both men and women.

Self-Perceived Military Fit. The degree of Self-Perceived Military Fit was measured through two items from Selection Part 1: motivation for military service and self-reported

suitability for military service. Both questions were answered on a four-point Likert scale ranging from "strongly disagree" to "strongly agree." The answers were transformed into a numeric scale 1–4. The total score for Self-Perceived Military Fit was the sum of the scores for the two items, ranging from 2 to 8. Cronbach's alpha for the scale has, based on the current sample, been estimated at .73 and .75 for men and women, respectively.

Tested Physical Fitness. The degree of Tested Physical Fitness was measured through two tests at Selection Part 2: physical endurance and physical strength. Physical endurance was measured with a treadmill test, and the result was transformed into a stanine score. Physical strength was measured with bench press and leg press tests, and the total result was transformed into a stanine score. The results were not gender adjusted, i.e., the norm used was the same for men and women. The total score for Tested Physical Fitness was the sum of the two tests, ranging from 2 to 18.

Officer-Rated Suitability. The degree of Officer-Rated Suitability for basic military service was assessed at Selection Part 2. A military recruitment officer rated the candidates on a four-point scale: "unsuitable," "less suitable," "suitable," and "very suitable." The results were transformed into a numeric scale 1–4. The assessment was made on the basis of an interview, and relevant factors were physical and mental robustness, attitudes, motivation, ambition, and social skills. Results from medical tests, GMA tests, and physical tests and the candidate's conduct on selection day were taken into account as well.

Military Performance. The Military Performance of the conscripts was assessed by an officer towards the end of the basic military service on a four-point scale: "below expectations," "according to expectations," "above expectations," and "excellent." All conscripts were assessed with regard to responsibility, cooperation and communication, initiative, accomplishment, independence, and conduct in the service, as well as an

independent overall impression of Military Performance. Only the overall impression assessment was recorded and made available for this study. The results were transformed into a numeric scale 1–4, with 4 corresponding to "excellent."

Statistical analyses

The statistical analyses were performed with the Statistical Package for Social Sciences (SPSS 21). In order to examine the predictive validity of the various predictors used in the selection process, descriptive statistics and bivariate correlations were calculated for men and women.

In addition, Cronbach's alpha was calculated for all multi-item scales. Values from .70 are considered adequate, from .80 good, and from .90 excellent (EFPA, 2013). The test–retest reliability of the paper-and-pencil versions of the GMA tests was assessed in a previous study (Sundet et al., 1988). Values from .60 are considered adequate, from .70 good, and from .80 excellent (EFPA, 2013). The equivalence of paper-and-pencil and computer-based versions of the GMA tests was explored in a previous study (Skoglund et al., 2014).

Hierarchical moderated multiple regression analysis was used to estimate a prediction model for Military Performance. A gender variable, with value 1 for women and 0 for men, was introduced to evaluate gender differences in the predictive validity in addition to interaction terms involving gender. First, all six continuous predictors, were centered according to the procedures outlined in Aiken and West (1991). GMA was entered at step 1 of the analysis, because it is a well-established predictor of performance (Drasgow et al., 2006; Held et al., 2014; Schmidt & Hunter, 1998). Three different self-reported variables were entered at step 2 of the analysis in order to assess whether these added incremental validity to the prediction of military performance. Similarly, the remaining two predictors

from Selection Part 2 were entered at step 3 of the analysis. Gender was introduced at step 4 of the analysis. Finally, six interaction variables between gender and the other six predictors were introduced at step 5 of the analysis.

Results

Bivariate correlations

Descriptive statistics and bivariate correlations for men and women were calculated separately (Table 1 and Table 2). All predictors were significantly correlated with Military Performance for men (*r* ranging from .05 to .19). The best predictor was Tested Physical Fitness (*r* = .19), followed by Officer-Rated Suitability (*r* = .13) and GMA (*r* = .13). Of the self-reported variables, Self-Perceived Physical Fitness (*r* = .11) and Social and Life Skills (*r* = .10) had slightly higher correlations with the criterion than Self-Perceived Military Fit (*r* = .05) (Table 1).

For women, the best predictor was Officer-Rated Suitability (*r* = .26), followed by Tested Physical Fitness (*r* = .19) and GMA (*r* = .17). Social and Life Skills (*r* = .12) was significantly correlated with the criterion , whereas the other two self-reported predictors, Self-Perceived Physical Fitness and Self-Perceived Military Fit, were not (Table 2).

_____

Insert Table 1 and Table 2 here

_____

Regression analysis

Hierarchical multiple regression analysis was used to estimate a prediction model for Military Performance (Table 3). Overall, the model explained 7 percent of the variance in Military Performance.

GMA ($B = 0.05$, $p < .001$) explained 2 percent of the variance in Military Performance. Three self-reported variables entered at step 2 explained an additional 2 percent of the variance in Military Performance. Of the individual predictors in step 2, only Social and Life Skills was significant ($B = 0.02$, $p < .01$). In the third step, Tested Physical Fitness and Officer-Rated Suitability added another 2 percent to the explained variance. Both added predictors were significant, with positive coefficients for both Tested Physical Fitness ($B = .05$, $p < .001$) and Officer-Rated Suitability ($B = 0.10$, $p < .001$). In the fourth step, gender ($B = 0.29$, $p < .01$) added another 1 percent to the explained variance. In the fifth and final step, where the interaction terms with gender were added, only one of the interaction terms was significant. Moreover, the step did not make a significant contribution to the explained variance, and thus no individual interaction effects were further explored.

Insert Table 3 here

Discussion

The purpose of this study was to examine the predictive validity of GMA and other predictors used in the selection process, to estimate a model for predicting military performance, and to examine gender differences in the predictive validity.

Individual predictors of Military Performance

GMA was a significant predictor of Military Performance for both men and women. This was in line with previous studies based on both civilian and military samples (Salgado & Anderson, 2003; Salgado et al., 2003; Schmidt & Hunter, 1998).

Social and Life Skills was also a significant predictor for both men and women. One possible explanation was that good social and life skills represent a protective factor that facilitates adjustment and social support from others.

The study included two predictors related to physical fitness, one self-reported and the other tested. Strong correlations were observed, in particular for Tested Physical Fitness. This supported previous results that indicated that physical fitness is important for military performance and service completion (Niebuhr et al., 2008; Taanila et al., 2011; Trone et al., 2013).

Two predictors involved aspects of motivation and job–person fit, which have in some studies proven to predict future performance (Gubuta et al., 2012). Self-Perceived Military Fit had a very low correlation with performance for men, and was non-significant for women. One possible explanation was the relatively long time span (approximately two-and-a-half years) between Selection Part 1 and the completion of service. In addition, the participants were relatively young, and motivation and plans for the future can change quickly. Officer-Rated Suitability is significantly correlated with Military Performance, and more so for women compared to men.

The observed correlations could in general be described as small according to Cohen's criteria (1988). However, they represented a conservative estimate of the predictive validity as the correlations were attenuated by range restriction and predictor reliability. In addition,

the criterion variable Military Performance had a relatively small variance in the current sample. Fewer than 10 percent of the participants in the study received any of the extreme values (1 or 4). Moreover, the participants in the study had served in different military units and the assessments may have varied between units and between individual officers.

A model for predicting Military Performance

The regression model predicted military performance, explaining 7 percent of the variance in military performance. The first four steps of the regression model were significant and added incremental validity to the prediction. GMA was entered in step 1, and step 2 included three self-reported variables where the only significant individual predictor was Social and Life Skills. Step 3 included Tested Physical Fitness in addition to Officer-Rated Suitability. When performing the interview, the officer had access to all of the information about the candidate, in addition to their own assessments about motivation and how well the candidate would fit military service. However, the ratings by the officer had relatively low inter-correlations with the other predictors, indicating that this assessment was not heavily influenced by any of the other predictors and, thus, added incremental validity to the model. Step 4 included Gender, which explained a small but significant part of the variance after controlling for the other predictors.

Overall, only a small part of the variance in Military Performance was explained by the regression model. However, as already explained, the criterion variable had relatively low variance. The criterion reliability may have been reduced by different expectations in the various types of service the conscripts were assigned to. In addition, meta-analyses have indicated that performance ratings by instructors in general have low reliability (mean r_{xx} = .52) as indicated by a meta-analysis of instructor ratings (Viswesvaran, Ones, & Schmidt, 1996). Furthermore, previous studies suggest that the correlations between cognitive

measures and performance ratings, like GMA and Military Performance in this study, are lower than between cognitive measures and more task-related performance tests (Campbell & Knapp, 2001).

One possibility for improving the model would be to include other relevant predictors, e.g., medical assessments, personality traits (Big Five), or hardiness. However, even if the overall explained variance was relatively modest, even predictors with small predictive validities may be useful in selection and increase the utility of the selection system (Brogden, 1946). Another possibility for improving the model would be to use a criterion which is more directly related to task or service performance. This could increase the predictive validity of GMA and possibly of other potential predictors, like tests assessing more specialized cognitive abilities or measures of work interests.

Gender differences

The moderated regression analysis did not indicate any significant interaction between gender and the predictors, as this step was non-significant. This suggests that the same regression model may be used for men and women. The individual variable Gender was significant, but explained only a small proportion of the variance in Military Performance (1 percent). These findings need to be further explored and replicated in a larger sample, especially since conscription in Norway now includes women as well. This change may result in more variation in motivation compared to the current sample, in which only motivated women were selected.

Study limitations

The main limitation of this study was restriction of range due to the fact that the predictor variables were criteria used in the selection process. Approximately 70 percent of

the men and 40 percent of the women who were tested at Selection Part 2 were selected for basic military service.

Moreover, some of the predictors were based on self-assessment. This means that impression management and social desirability may have influenced some of the answers, as well as the desire to be selected for military service.

Finally, the criterion variable Military Performance had low reliability and little variance. It is still a highly relevant criterion as the Norwegian Armed Forces use the rating for further recruitment of conscripts. A more precise and detailed criterion could have resulted in a model with a larger part of the variance explained.

Conclusion

A small, but significant, part of the variance in Military Performance was explained by the model. Predictors from both parts of the selection process made a small, but significant, contribution to improving the model. This indicated that the two-step selection process was valid for predicting military performance for both men and women. No significant gender difference in the predictive validity of the overall model was observed.

Declaration of Interest

One of the authors is affiliated to the Norwegian Armed Forces' Personnel and Conscription Centre, which is responsible for selection of conscripts to the Norwegian Armed Forces. Two of the authors are affiliated to the Norwegian Military Defence University College, which is responsible for the ability testing of conscripts. All authors are affiliated to

the Norwegian defense sector. The opinions expressed are those of the authors and not necessarily those of the Norwegian Armed Forces.

References

Aiken, L. S., and West, S. G. (1991), *Multiple regression: Testing and interpreting interactions*, Sage Publications Inc., London, UK.

Bäccman, C., Berggren, A. W., & Norlander, T. (2012). Military capacity and civil adjustment: Assessments of the 're-usable' peacekeeping soldier for development of a selection system. *International Journal of Selection and Assessment*, *20*, 171–181. doi: 10.1111/j.1468-2389.2012.00590.x

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Pscyhology, 44*, 1–26.

Bartone, P. T., Roland, R. R., Picano, J. J., & Williams, T. J. (2008). Psychological hardiness predicts success in US Army Special Forces candidates. *International Journal of Selection and Assessment*, *16*, 78–81.

Berry, C. M., Clark, M. A., & McClure, T. K. (2011). Racial/ethnic differences in the criterion-related validity of cognitive ability tests: A qualitative and quantitative Review. *Journal of Applied Psychology*, *96*, 881–906. doi: 10.1037/a0023222881

Bertua, C., Anderson, N., & Salgado, J. F. (2005). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology*, *78*, 387–409. doi: 10.1348/096317905X26994

Brogden, H. E. (1946b). On the interpretation of the correlation coefficient as a measure of predictive efficiency. *Journal of Educational Psychology*, *37*, 65-76.

Campbell J. P. & D. J. Knapp, D. J. (Eds.). (2001). Exploring the limits of personnel selection and classification. Mahwah, NJ: Erlbaum.

Campbell, S. K., & Catano, V. M. (2004). Using measures of specific abilities to predict training performance in Canadian Forces operator occupations. *Military Psychology, 16,* 183–201. doi:10.1207/s15327876mp1603_3

Carretta, T. (1997). Male–female performance on US Air Force pilot selection tests. *Aviation Space and Environmental Medicine*, *68*, 818–823.

Carretta, T. R., Teachout, M. S., Ree, M. J., Barto, E. L., King, R. E., & Michaels, C. F. (2014). Consistency of the relations of cognitive ability and personality traits to pilot training performance. *The International Journal of Aviation Psychology, 24,* 247–264. doi: 10.1080/10508414.2014.949200

Cleary, T. A. (1968). Test bias: Prediction of grades of negro and white students in integrated colleges. *Journal of Education Measurement, 5,* 115–124

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Drasgow, F., Embretson, S. E., Kyllonen, P. C., & Schmitt, N. (2006). *Technical review of the Armed Services Vocational Aptitude Battery* (Final Rep. No. 06–25). Alexandria, VA: Human Resources Research Organization.

Earles, J. A., & Ree, M. J. (1992). The predictive validity of the ASVAB for training grades. *Educational and Psychological Measurement, 52,* 721–725. doi:10.1177/0013164492052003022

European Federation of Psychologists' Association (EFPA) (2013). *EFPA Review model for the description and evaluation of psychological tests: Test review form and notes for reviewers, v 4.2.6*. Downloaded from http://www.efpa.eu/professional-development

Gubata, M. E., Oetting, A. A., Weber, N. S., Feng, X., Cowan, D. N., & Niebuhr D. W. (2012). A noncognitive temperament test to predict risk of mental disorders and attrition in US Army recruits. *Military Medicine, 177*, 374–379.

Guion, R. M., & Gottier, R. F. (1965). Validity of personality measures in personnel selection. *Personnel Psychology, 18*, 135–164.

Hansen, I. (2006). *Bidrag til psykologitjenestens historie i Forsvaret fra 1946–2006* [Contributions to the history of military psychology in the Armed Forces 1946–2006]. Militærpsykologiske meddelser, nr. 25. Oslo: Forsvarets skolesenter.

Held, J. D., Carretta, T. R., & Rumsey, M. G. (2014). Evaluation of tests of perceptual speed/accuracy and spatial ability for use in military occupational classification. *Military Psychology*, *26*, 199–220. doi:10.1037/mil0000043

Køber, P. K. (2015). *Velger Forsvaret de rette ungdommene? – analyse av seleksjon, gjennomføring og frafall i førstegangstjeneste 2010–2014*. [Are the Norwegian Armed Forces selecting the right soldiers? – analysis of selection, completion and drop-out in the basic military service 2010–2014]. (Norwegian Defence Research Establishment Report No. 2014/02174). Retrievable from http://www.ffi.no/no/Rapporter/14-02174.pdf

Martinussen, M., & Torjussen, T. M. (2004). Initial validation of a computer-based assessment battery for pilot selection in the Norwegian Air Force. *Human Factors and Aerospace Safety, 4*, 233–244.

The National Service Administration (2007). *Internet – en sparbössa för försvaret: Utvecklingen av Pliktverkets system för lämplighetsundersökning före mönstring* [Internet – a cost saver for the military: The development of a screening method before conscription by the National Service Administration]. Downloaded from: http://arkiv.edelegationen.se/verva/upload/verksamhetsstod/externa-konferenser/Kvalitetsmassan2007/128.Internet-sparbossa.pdf

Niebuhr, D. W., Gubata, M. E., Oetting, A. A., Weber, N. S., Feng, X., & Cowan, D. N. (2013). Personality Assessment Questionnaire as a pre-accession screen for risk of mental disorders and early attrition in US Army recruits. *Psychological Services*, *10*, 378–385. doi:10.1037/a0032783

Niebuhr, D. W., Scott, C. T., Powers, T. E., Li, Y., Han, W., Millikan, A. M., & Krauss, M. R. (2008). Assessment of recruit motivation and strength study: Pre-accession physical fitness assessment predicts early attrition. *Military Medicine*, *173*, 555–562.

Roberts, H. E., & Skinner, J. (1996). Gender and racial equity of the Air Force Officer Qualifying Test in officer training school selection decisions. *Military Psychology*, *8*, 95–113. doi:10.1207/s15327876mp0802_4

Rumsey, M. G. (2012). Military selection and classification in the United States. In J. H. Laurence & M. D. Matthews (Eds.), *The Oxford Handbook of Military Psychology* (pp. 129–147). New York: Oxford University Press.

Rumsey, M. G., & Arabian, J. M. (2014). Military enlistment selection and classification: Moving forward. *Military Psychology*, *26*, 221-251. http://dx.doi.org/10.1037/mil0000040

Salgado, J. F., & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European community. *European Journal of Work and Organizational Psychology*, *12*, 1–17.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities as predictors of work behaviors: A European meta-analysis. *Personnel Psychology*, *56*, 573–605.

Salo, M. (2008). Determinants of military adjustment and attrition during Finnish conscript service. Doctoral dissertation. Faculty of Education, University of Tampere, Finland. Downloaded from https://tampub.uta.fi/bitstream/handle/10024/67908/978-951-44-7470-5.pdf?sequence=1

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.

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

Sundet, J. M., Tambs, K., Magnus, P., & Berg, K. (1988). On the question of secular trends in the heritability of IQ test scores: A study of Norwegian twins. *Intelligence, 12,* 47–59.

Taanila, H., Hemminki, A., Suni, J., Pihlajamäki, H., Parkkari, J. (2011). Low physical fitness is a strong predictor of health problems among young men: A follow-up study of 1,411 male conscripts. *BMC Public Health*, *11*, 590. doi:10.1186/1471-2458-11-590

Thomassen, Å. G., Hystad, S. W., Johnsen, B. H., Johnsen, G., Laberg, J. C., & Eid, J. (2015). The combined influence of hardiness and cohesion on mental health in a military

peacekeeping mission: A prospective study. *Scandinavian Journal of Psychology, 56*, 560–565. doi: 10.1111/sjop.12235

Torjussen, T. M., & Hansen, I. (1999). Forsvaret: Best i test? Bruk av psykologiske tester i Forsvaret, med spesiell vekt på flygerseleksjon [The Armed Forces: Best in testing? The use of psychological testing with special focus on pilot selection]. *Journal of The Norwegian Psychological Association, 36*, 772–779.

Trone, D. W., Cipriani, D. J., Raman, R., Wingard, D. L., Shaffer, R. A., & Macera, C. A. (2013). The association of self-reported measures with poor training outcomes among male and female US Navy recruits. *Military Medicine, 178*, 43–49.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology*, *81*, 557–574. doi:10.1037/0021-9010.81.5.557

Table 1

Bivariate Correlations and Reliability Estimates for Men (N =1785 to 2684)

Variables	*M*	*SD*	1	2	3	4	5	6	7
Mental ability									
1. GMA	5.8	1.5	–						
Self-reported variables									
2. Self-Perceived Physical Fitness	8.2	1.9	-.10**	.67					
3. Social and Life Skills	47.5	3.4	-.01	.30**	.74				
4. Self-Perceived Military Fit	7.0	1.3	-.12**	.36**	.41**	.73			
Measured variables									
5. Tested Physical Fitness	11.6	2.4	.00	.56**	.15**	.17**	–		
6. Officer-Rated Suitability	3.5	0.5	.12**	.14**	.14**	.14**	.23**	–	
Outcome variable									
7. Military performance	2.6	0.7	.13**	.11**	.10**	.05*	.19**	.13**	–

Note. *$p < .05$. **$p < .01$ (two-tailed).

Table 2

Bivariate Correlations and Reliability Estimates for Women (N = 350 to 592)

Variables	*M*	*SD*	1	2	3	4	5	6	7
Mental ability									
1. GMA	5.0	1.3	–						
Self-reported variables									
2. Self-Perceived Physical Fitness	7.7	1.8	-.01	.70					
3. Social and Life Skills	47.6	3.4	.07	.24**	.74				
4. Self-Perceived Military Fit	6.6	1.5	-.02	.36**	.38**	.75			
Measured variables									
5. Tested Physical Fitness	6.6	1.8	.06	.45**	.08	.12**	–		
6. Officer-Rated Suitability	3.3	0.5	.16**	.23**	.10*	.10*	.28**	–	
Outcome variable									
7. Military performance	2.6	0.6	.17**	.09	.12*	.03	.19**	.26**	–

Note. *$p < .05$. **$p < .01$ (two-tailed).

Table 3

Hierarchical Regression Analysis Results for the Prediction of Military Performance

	Military performance	
Predictor	ΔR^2	B^a
Step 1: Mental ability	.02***	
GMA		0.05***
Step 2: Self-reported variables	.02***	
Self-Perceived Physical Fitness		0.00
Social and Life Skills		0.02**
Self-Perceived Military Fit		0.00
Step 3: Measured variables	.02***	
Tested Physical Fitness		0.05***
Officer-Rated Suitability		0.10***
Step 4: Gender	.01***	
Gender (0 = male, 1 = female)		0.29**
Step 5: Interaction variables	.00	
Gender × Officer-Rated Suitability		0.17*
Total R^2	.07***	
N	2113	

Note. [a] The *B*'s (unstandardized coefficients) are from the final regression model. Only significant interaction terms were listed.

*$p < .05$. **$p < .01$. ***$p < .001$.

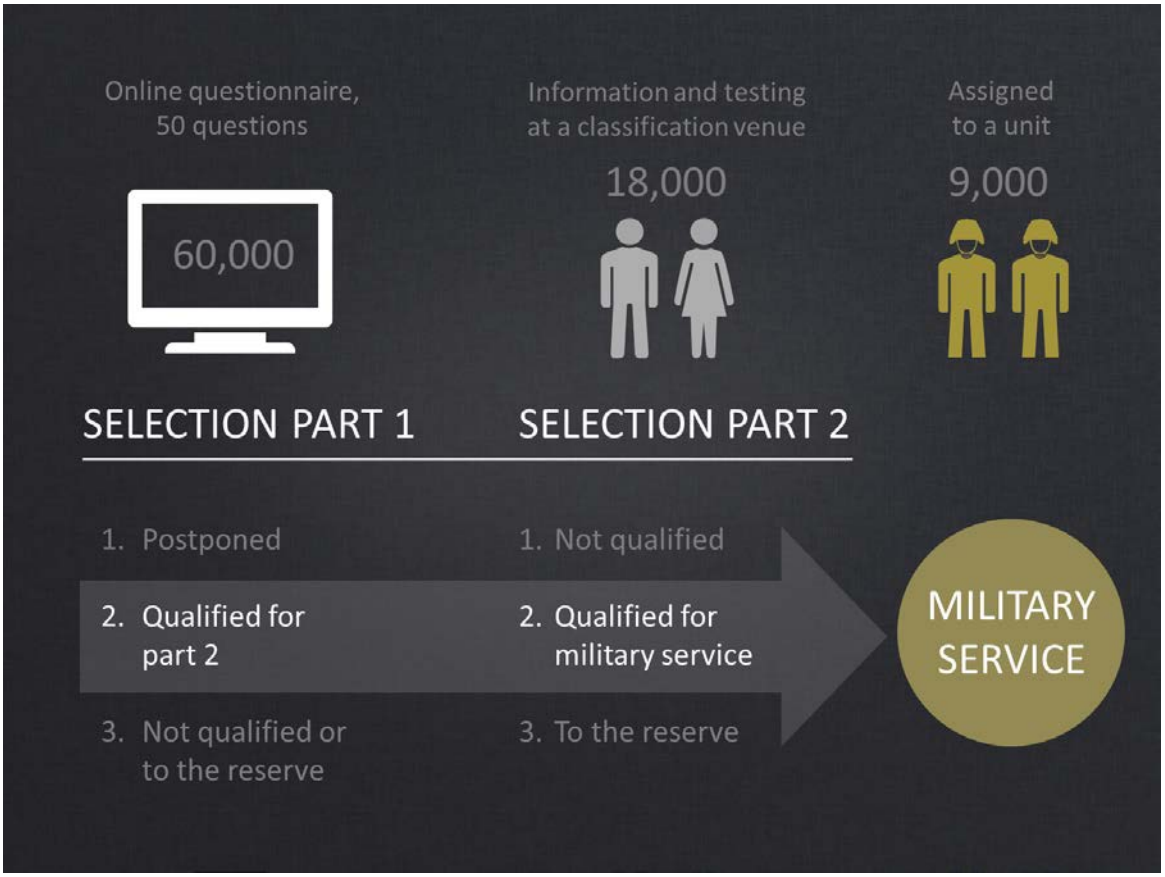

Figure 1 The Two-Step Selection System for Conscripts in the Norwegian Armed Forces. (Copyright: The Norwegian Armed Forces' Personnel and Conscription Centre (FPVS))