BI Norwegian Business School - campus Oslo

GRA 19703

Master Thesis

Thesis Master of Science

Selecting Future Leaders - A validation study of the selection process for military leadership education in the Norwegian Armed Forces

Navn:	Marie Victoria Ødegaard, Kristoffer Tiller

Start:	15.01.2020 09.00
Finish:	01.09.2020 12.00

Master Thesis

MSc. Leadership and Organizational Psychology

Selecting Future Leaders
-
A validation study of the selection process for military leadership education in the Norwegian Armed Forces

Supervisor:
Christian Winther Farstad

Acknowledgements

During these two years as students, we have been introduced to a versatile education with many exciting subject areas, and gained an increased interest in the complex interplay between management, human resources and psychological factors. This thesis marks the end of our two-year long journey at BI Norwegian Business School. In that regard, we would like to thank our supervisor Christian Winther Farstad for his invaluable guidance during the process of writing this Master's thesis. We highly appreciate the time and effort spent over these past two years in sharing knowledge and insights about the research process.

We would also like to thank Henrik Sørlie and Thomas Hol Fosse at the Norwegian Defense University College (NDUC) for helping us get access to data, and for providing us with valuable inputs on the topics of military selection and education. We are especially grateful for the extra effort Henrik Sørlie has spent in helping us understand the dataset and guiding us through the analysis of this data.

We also express our gratitude to Stein Garang at the Norwegian Armed Forces HR and Conscription Centre for providing us with documents facilitating our knowledge and understanding of the Joint Admission and Selection process.

Finally, we would like to thank Michael Baas Bottenvik-Hartmann for providing feedback on this thesis.

The process of writing this thesis has given us valuable insights to the selection process and its ability to predict future performance in both education and training, in addition to performance in the role as a leader. These are insights we are thankful that the abovementioned people have helped us gain, and which we hope to bring along in our future work upon graduation from BI Norwegian Business School.

With the sincerest gratitude,
Kristoffer Tiller & Marie Victoria Ødegaard

Summary

The ability to predict future performance is crucial for sustainable selection procedures. While several selection methods individually predict both performance in job training programs and subsequent job performance, of particular interest is the ability of a system of methods to predict performance in related yet different domains.

This study investigates the extent to which the selection system for the Norwegian Armed Forces' Officer Candidate School is able to predict performance during education and training, in addition to subsequent performance in the role as a military leader. Furthermore, the relationship between selection, education and subsequent leadership performance is investigated through assessing the contribution of education and training in predicting leadership performance when controlling for the effect of the selection system.

While the selection system is highly predictive of academic performance in the education's theoretical aspects, it is far less predictive of performance in practical aspects and of subsequent performance in the role as a military leader. Performance during education and training is, expectedly, highly predictive of subsequent leadership performance. Paradoxically, it is performance in the education's practical aspects that contributes to subsequent leadership performance, while academic performance in the education's theoretical aspects appears to be negatively (though not significantly) related to subsequent leadership performance.

The study concludes that the skills, competencies and abilities required for learning in academic environments are not necessarily important for mastering the practical aspects of military leadership. In such, a tailormade rather than generic competency framework based on identification of specific skills, competencies and abilities directly relevant to a specific job role and its context would facilitate accurate definition of selection criteria, optimization of their use, and ultimately greater accuracy in predicting subsequent on-the-job performance.

Table of Contents

1. Introduction

Building and maintaining a capable workforce through successful selection procedures constitute a source of organizational competitive advantage (Salas, Tannenbaum, Kraiger & Smith-Jentsch, 2012; Viswesvaran & Ones, 2018). The most important property of any selection method is its predictive validity, that is, its ability to predict future performance (Schmidt & Hunter, 1998). The predictive validity of any selection method is directly proportional to its practical economic value (Schmidt, Hunter, McKenzie & Muldrow, 1979).

According to The Norwegian Department of Defense, the most valuable resource of the Norwegian military is its personnel (Forsvarsdepartementet, 2013). Identifying valid performance predictors for personnel selection has always been an important challenge for military organizations (Fosse, Buch, Säfvenbom, & Martinussen, 2015), and the ability to attract, select, develop and retain the right personnel is also said to be one of the most important strategic challenges of the Norwegian Armed Forces (NAF) (Forsvarsdepartementet, 2012).

Substantial resources are placed into personnel selection in the NAF, thus from cost perspective, this selection should entail valid methods. Considering the number of candidates evaluated and selected for military leadership education in the NAF, even small increases in the selection process' accuracy can have a substantial impact. However, in a military context, economic gains are not the main incentive for accurate selection procedures. Rather, increased operational capability and impact, and importantly, avoidance of loss of human lives motivates accurate selection (Forsvarets personell- og vernepliktssenter, 2019a).

The NAF is a hierarchical organization where military leaders are recruited through an education system that qualifies them for leadership positions at various levels (Lang-Ree & Martinussen, 2019). In such, the NAF does not primarily select candidates for military education, but for the profession (Forsvarets personell- og vernepliktssenter, 2019b). This requires leadership potential to be assessed and identified even at admission to lower-level educational programmes, which together with an increasing differentiation and specialization of the military profession (Forsvarets personell- og vernepliktssenter, 2019b), poses substantial requirements for personnel selection.

1.1 Research Question

The NAF is a particularly interesting organization within which to study personnel selection. Given that candidates are selected for a profession, and may come to fill positions across a range of hierarchical levels, branches and functions, identifying selection criteria able to predict performance in such a range of positions is a core challenge. Admission to these positions is regulated through military leadership education, which adds to the requirements of the selection process, necessitating its ability to also predict the likelihood of performance in and successful completion of education. The NAFs' military branches' need for competence constitutes the basis for initiating the selection process for education (Forsvarets personell- og vernepliktssenter, 2019a). A prerequisite for a sustainable education system is that resources for selection and training are an investment for a lifelong employment relationship. A successful selection process will therefore both result in lower levels of dropout during education and training, and higher productivity as a result of high long-term work performance.

Despite the varying requirements in terms of knowledge, skills and competencies associated with the range of positions military personnel may fill throughout their careers, a common requirement for all leader positions within the NAF, regardless of hierarchical level and function, is executing leadership in line with that specified in "*the Chief of Defence's perspective on leadership*" (*Forsvarssjefens grunnsyn på ledelse*) (Forsvarsstaben, 2012). Efforts to operationalize these specifications have resulted in a framework of *balanced leadership behaviour*, and a tool for measuring this is developed (Martinsen, Fosse, Johansen & Venemyr, unpublished). "*The Chief of Defence's perspective on leadership*" is also implemented into personnel selection and training, and governs which criteria and competencies are sought during the selection process aimed at evaluating military leadership potential.

However, this potential is not necessarily sufficient for direct entry into lower-level leader positions within the NAF. Military leadership requires mastery of specific military competencies achieved through education and practice in the military profession (Grebstad & Johnsen, 2019). In such, performance in military leadership education and training is likely to cultivate leadership potential and also be predictive of subsequent on-the-job leadership behaviour.

In order to document the predictive validity of a selection system, a covariance must be demonstrated between selection methods results and some

measure of performance. Typical performance criteria in validation studies include measures of job performance, performance in job training programs (Schmidt & Hunter, 1998), or educational performance (Lang-Ree & Martinussen, 2019).

Our case is the selection process for admission to the NAFs' Officer Candidate School (OCS), which has traditionally been the first step in the NAFs' three-leveled education system (Forsvarsdepartementet, 2013). This system was however recently restructured following the implementation of a new organizational rank-structure (St.prp. 111 LS, 2014-2015), dividing the structural model into two complementary career systems facilitating development of officers with broad competencies and specialists with in-depth expertise in their field. The selection system as described in this study however remains largely unchanged.

The traditional selection process for admission to the OCSs' success hinges on its ability to predict both performance in education and training, and subsequent performance in the role as a military leader. Through a longitudinal design, we thus investigate the ability of this selection process to predict both performance during military leadership education and training and subsequent *balanced leadership behaviour* (BLB). While BLB is not a measure of job performance per se, but rather a measure of specific leadership behaviours essential for effective military leadership, it is a highly relevant criterion to validate the selection process against for several reasons. Firstly, it represents a set of valued behaviours in the organization. Secondly, while the predictive validity of selection processes in the NAF has been evaluated against supervisory performance ratings criteria (Kjenstadbakk, 2012; Norrøne, 2016; Vik, 2013), the extent to which these are able to predict valued military leadership behaviours remains unknown. Lastly, research conducted in the process of validating the BLB tool suggests that BLB scores predict supervisory ratings of job performance (Martinsen et al., unpublished). Thus, balanced leadership behaviour is closely related to job performance for military leaders in the NAF, and is therefore conceptualized as a measure of military leadership performance in this study.

Given that military leadership requires mastery of specific military competencies achieved through education and practice in the military profession (Grebstad & Johnsen, 2019), the contribution of military education and training in explaining variation in BLB evaluations when controlling for the effect of the selection system is also investigated. Figure 1.1 illustrates this study's research model.

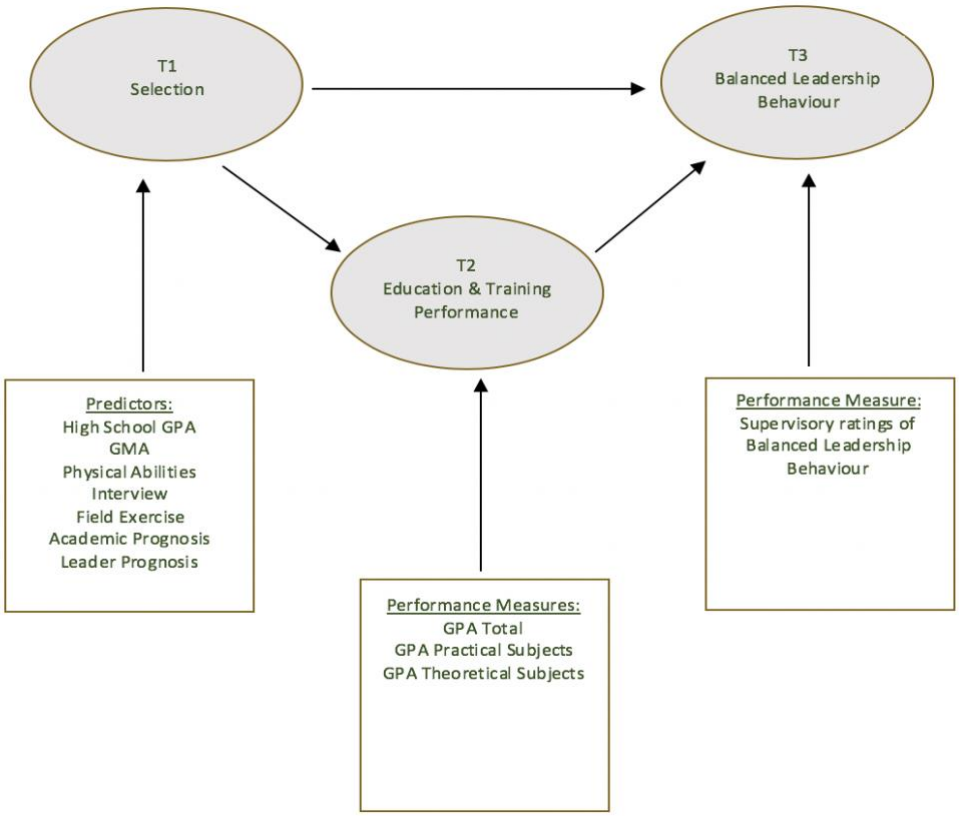

Figure1.1: Research model.

Thus, this study constitutes a validation study of the selection process for admission to the Norwegian Armed Forces' Officer Candidate School, aiming to answer the following research questions:

To what extent does the selection system for admission to military leadership education (T1) predict performance in education and training (T2), and subsequent performance in the role as a military leader (T3)?

To what extent does performance in military education and training (T2) predict subsequent performance in the role as a military leader (T3)?

2. Selection of Military Leaders in the Norwegian Armed Forces

Many young Norwegians aspiring to become leaders in the NAF apply for the Officer Candidate School, the first step in the NAFs' three-leveled education system (Forsvarsdepartementet, 2013). Candidates who are admitted to the OCS embark on a two-year education. The first year entails education and training in both theoretical and practical aspects of military leadership. The second year consists primarily of practice, where each officer candidate holds a leader role for a team of conscripts. The purpose of the OCS is to ensure the Armed Forces officer

candidates to occupy positions and execute relevant missions at the lowest level of leadership function, provide a basis for potential further military education, and to promote knowledge, skills and attitudes according to the NAFs' requirements and needs. Subsequent to completing the OCS, some graduates continue their military career as officer candidates, while others pursue further education through the Basic Officer Education (BAE)[1], which further qualifies for Advanced Officer Education (AO)[2]. These higher-level educations are required for rank ascendance within the organizational hierarchy and pursuing a life-long career. As the OCS is the first step of NAFs' effort to develop what may be their future military leaders and generals, it is considered a cornerstone in their education system (Forsvarets høyskole, 2013). Selecting the right personnel for admittance to the OCS is therefore of great importance, and should arguably be based on the best scientific evidence available.

The selection process for admission to OCS education and training has been conducted under Joint Admission and Selection ("Felles Opptak og Seleksjon"). This selection process is a four-step process with the aim and purpose to uncover candidates' fitness, abilities and motivation for leadership education in the NAF. The first step entails screening applicants based on high school grade point average (GPA), general mental ability (GMA) scores and criminal records. Qualified applicants are summoned for a two-week selection process. Upon arrival, applicants are checked for possession and use of illegal drugs, undergo a medical examination, and are tested in physical abilities. Applicants who do not pass the medical and/or physical requirements are sent home at this point, while qualified applicants undergo a structured interview assessing military leadership development potential. A leadership prognosis is derived based on interview scores, and an academic prognosis is calculated based on high school GPA and GMA scores.

Candidates who pass through to the next phase undergo a 5-7-day long field exercise consisting of various military simulation exercises. Candidates are divided into teams of 6-10 members, where the team leader role rotates among members. A trained observer follows each team and assesses each candidate's performance in the leader role. Based on field exercise scores, another leadership prognosis is set. Ultimately, the academic prognosis, an overall leader prognosis (derived by

[1] A 3-year undergraduate program at the Norwegian Defence University College that leads to a bachelor's degree in military studies, which specializes in leadership and military power, within the military branch of either the Air Force, Navy or the Army.

[2] A graduate program at the Norwegian Defence University College that leads to a master's degree in military studies, which specializes in military power, development of doctrines and concepts, joint military operations and staff methodology.

weighting and combining the prognoses derived through interview and field exercise scores), and results on physical ability tests constitute the basis for selection. Figure 2.1 depicts the successive step-wise process as a model. Candidates who do not meet the criteria at one stage are not tested further through more comprehensive and time-consuming methods.

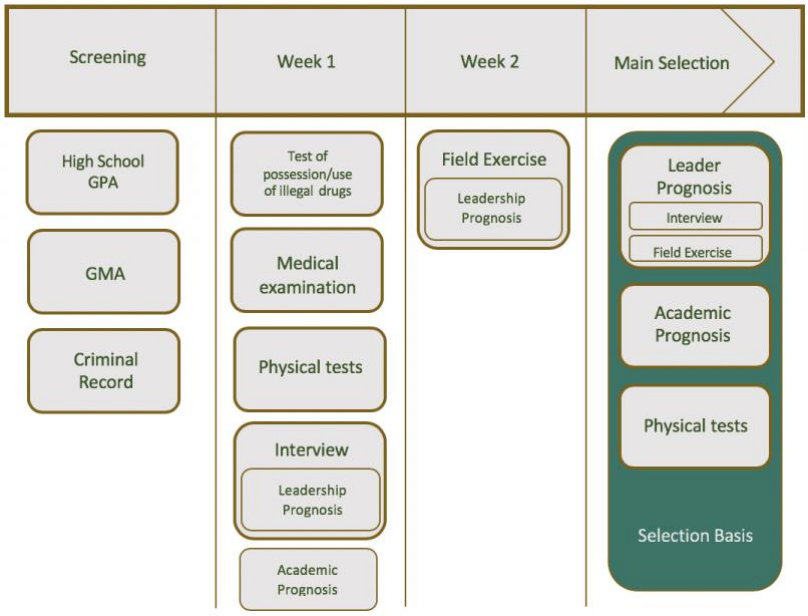

Figure 2.1: Model of the Joint Admission and Selection Process.

The distinctiveness of the NAFs' function and responsibility poses a core challenge to selection of candidates for military leadership education and training (Forsvarets personell- og vernepliktssenter, 2019a). Furthermore, a highly valid question is whether there exist robust generic selection criteria that embrace the range of positions across organizational levels and branches that candidates may come to hold after completing the OCS. With the military profession becoming ever more differentiated and specialized, it is almost a paradox that the selection process for lower-level education, training and leader positions is generic in nature.

«*The Chief of Defence's perspective on leadership*» states that "The selection of leaders through approved and scientific based instruments (tests), followed by leadership development measures that take into account the distinctiveness of the military leader's challenges, are important measures for ensuring that the Armed Forces at any given time have leaders with the required personal prerequisites" (Forsvarsstaben, 2012, p. 13). Thus, the utmost objective during selection is to identify candidates with the greatest potential for executing military leadership in line with the NAFs' distinctive guidelines.

3.Literature Review

The theoretical foundation for this study is twofold, drawing on theory from within the two academic fields of leadership and selection. Relevant leadership theories are presented and discussed against the context of the NAF, before the literature on selection methods is reviewed. Theories and research of various selection tools are presented and discussed in relation to the OCS selection process.

3.1 Theories of Leadership

Despite being extensively researched, the nature of leadership still remains a debated topic (Yukl, 2013). What most definitions of leadership have in common is that leadership involves a process of exerting influence over someone (Yukl, 2013) in order to achieve a goal (Avolio & Bass, 1994). For example, House and Javidan (2004) define leadership as "the ability of an individual to influence, motivate, and enable others to contribute toward the effectiveness and success of the organizations of which they are members of" (p. 15). One reason for the theoretical disagreements around the concept of leadership is that leadership has been studied from distinctive perspectives. Of these approaches, the leader-centred approach, which focuses on characteristics of the leader (Bass & Bass, 2008), and investigates the determinants of effective leadership in terms of traits, abilities, behaviours or aspects of the situation (Yukl, 2013), is most relevant to the OCS selection process, as candidates are admitted based on assessments of individual characteristics and portrayed behaviours.

3.1.1 Trait Theories of Leadership

A major focus within leadership research has been identifying the traits and skills that characterize successful leaders. While a trait – a «relatively stable disposition to behave in particular ways» (Yukl, 2013, pp.143) - can refer to various individual attributes such as personality aspects, temperament, needs, motives and values, much of the research within the trait approach has focused on personality traits (Yukl, 2013). Research suggests a relationship between aspects of personality and leadership, however, personality is found to have a stronger effect on leader emergence than on leader effectiveness (Judge, Bono, Ilies & Gerhardt, 2002). This distinction is important, as being perceived by others as a leader does not guarantee leadership effectiveness in terms of positive organizational, financial and relational

effects. This distinction is particularly important in relation to selection, where the aim must be to identify candidates who will actually be effective leaders as opposed to having the appearance of a leader.

Another focus within the trait approach has been to identify the skills predictive of leadership. Here, skills have been defined at various levels of abstraction, ranging from broad general abilities such as intelligence to more narrow and specific abilities (Yukl, 2013). Research suggests that the importance of cognitive skills increases along with the level of complexity of jobs (Schmidt & Hunter, 1998), while meta-analytic findings suggest that intelligence is moderately related to leadership effectiveness (Judge, Colbert & Ilies, 2004).

According to Yukl (2013), the relative importance of different skills varies depending on level of management. A taxonomy of skills related to leader effectiveness includes *technical skills* (knowledge about methods, processes, procedures and techniques), *interpersonal skills* (knowledge about human behaviour and interpersonal processes, empathy, communicative abilities) and *conceptual skills* (general analytical ability, logical thinking, inductive and deductive reasoning) (Yukl, 2013). Due to differing requirements at different levels of management, technical, interpersonal and conceptual skills may not be of equal importance to all leaders. While the importance of technical skills decreases as level of management increases, the importance of conceptual and interpersonal skills increases with the ascendance to higher levels (Yukl, 2013). Supportive of this is the finding that the importance of cognitive skills increases along with the level of complexity of jobs (Schmidt and Hunter, 1998).

3.1.2 Leadership Behaviours

Another approach to studying leadership has focused on the behaviours, which are more adaptable than traits, related to leader effectiveness. While certain traits may be positively related to leadership effectiveness, in order to understand exactly what effective leaders do to influence subordinates, their actual behaviours must be the unit of inquiry.

Summarizing research on leadership behaviour over the past half century, Yukl (2012) proposes three major types of leadership behaviour, distinguishable by their task, relations and change orientation. Task-oriented behaviours are those that contribute to the accomplishment of tasks efficiently and reliably, while relations-oriented behaviours increase mutual trust, cooperation and job satisfaction through

the support of subordinates (Yukl, 2012). These two types of behaviour are based on the meta-categories *initiating structure* and *consideration*, respectively, from early works on identifying categories of leadership behaviour (e.g. Fleishman, 1953). Change-oriented behaviours concern understanding the environment, adapting to it innovatively, and implementing major changes (Yukl, 2012).

Research on the effects of leadership behaviour on indicators of leadership effectiveness such as subordinate satisfaction, subordinate performance and superiors' ratings of leader effectiveness has generated ambiguous results. A meta-analysis interpreting and summarizing these mixed results found both initiating structure and consideration to be related to leadership outcomes, but consideration more strongly so (Judge, Piccolo & Ilies, 2004). Consideration relates more to follower satisfaction, motivation and leader effectiveness, while initiating structure relates more to leader job performance and group-organization performance (Judge, Piccolo & Ilies, 2004). Investigating the relative validities of different leader behaviours, Piccolo and colleagues (2012) found that when compared directly to other leadership styles, initiating structure hardly had any effect on organizational outcomes. With different leadership styles relating differently to various organizational outcomes, effective leadership may require both task, relations, and change-oriented behaviours (Martinsen, Fosse & Johansen, 2019; Yukl, 2013).

3.1.3 Leadership versus Military Leadership

While aspects of the NAFs' doctrinaire approach to leadership is based on general leadership theory, a legitimate question is whether military leadership is a unique concept, or if leadership is a generic concept valid across the military-civil distinction. If leadership in a military context deviates from leadership in civil organizations, general leadership theories may be inapplicable to a military context.

A review and comparison of military leadership doctrines and general leadership theories concluded that western military doctrines to great extent tend to correspond to the contents of established leadership theory (Chan, Soh & Ramaya, 2012). Particularly, definitions of leadership from general theory hold that leadership entails a process of influence. Influence is also a central component of *mission command*, which is a common military leadership philosophy (Parrington & Findlay, 2013). However, although military leadership is defined similarly to general definitions of leadership, the purpose and authority of military organizations renders military leadership distinguishable to some extent from

leadership in civil organizations. According to Hannah & Snowden (2013), what differentiates military from general leadership is the context in which it is executed. The context in which military leadership is executed is at times characterized as dangerous and extreme, which may be considered the core condition for performing military leadership (Hannah et al., 2013). This means that military personnel are faced with demanding tasks and environments that require hardiness, resilience and robustness (Bartone, Eid, Johnsen, Laberg, & Snook, 2009; Maddi, 2007), and more structure and professionalism than what is needed in many other industrial and professional contexts (Fosse, Skogstad, Einarsen & Martinussen, 2019). The notion of context further raises the question of whether military leadership can meaningfully be conceptualized as one coherent form of leadership.

The NAF is a highly complex organization, consisting of numerous hierarchical levels and sub-organizations. Furthermore, the organization contains different functional segments, reflecting the operative and administrative aspects of the military profession. This is reflected in the introduction of "*the Chief of Defence's perspective on leadership*" (Forsvarsstaben, 2012), in which it is stated that the NAFs' leadership philosophy lies at the intersection between operative and administrative principles. Military leaders must be able to master both the operative and administrative contexts, that is to integrate leadership and management (Johansen, Fosse & Boe, 2019). Additionally, the range of the NAFs' mission portfolio, including a peace, crisis and war perspective, adds to the contextual variation within which military leadership is executed.

Contingency theories of effective leadership hold that the traits or behaviours required for effective leadership vary for different situations (Yukl, 2013). Whereas the trait and behaviour approaches assume that certain leader traits or broadly defined behaviours are positively related to leadership effectiveness in all situations, the contingency perspective describes how situational variables may influence these relationships (Yukl, 2013). Given the multiple contexts within which a military leader may need to practice his or her leadership, it is likely that he or she must possess a range of different traits and behaviours in order to be an effective leader in the various contexts.

3.1.4 Requirements and Characteristics of the Military Leader

The nature of military organizations poses substantial requirements to military leaders. Military leaders must master a range of tasks, such as formulating

and implementing strategies, ensuring an effective organization with efficient communication flows, creating a sense of unity and an understanding of the mission within the section, ensuring the competency of individuals and teams, and, most importantly, lead under extreme conditions in crisis and combat situations. These complex tasks require gathering, analysing and combination of information, at times in unpredictable situations (Lang-Ree & Martinussen, 2019). Despite these task requirements, there is no definite answer to what effective military leadership should entail (Martinsen et al., 2019).

The NAFs' official leadership philosophy is *mission command*, which is considered the most effective leadership philosophy in military operations (Ben-Shalom & Shamir, 2011; Parrington & Findlay, 2013). At the core of this philosophy lies the notion that any mission must be understood in terms of its intention (Forsvarsstaben, 2014). Compliance is to be tied to the intention of the mission as opposed to the direct wording of a specific command (Forsvarsstaben, 2012). Mission command permits decentralized decision making based on the intention of the mission, allowing the military leader to make independent assessments for action based on his or her situational understanding (Forsvarsstaben, 2012). This philosophy allows initiative to be taken at all organizational levels (Forsvarsstaben, 2014), encouraging all soldiers to lead peers and subordinates at one point or another during their term of service. However, mission command can in certain instances also call for detailed orders and control (Forsvarsstaben, 2014). Thus, mission command entails shifting between strict command and decentralization, which requires high levels of trust (Forsvarsstaben, 2014). In such, the NAFs' leadership philosophy calls for both task-oriented (solving the mission) and relations-oriented (creating mutual trust) behaviours, which is well-aligned with the general leadership literature (Yukl, 2013).

"*The Chief of Defence's perspective on leadership*" (Forsvarsstaben, 2012) elaborates the requirements of leaders in the NAF further. The intention of this piece of work is to concretize how leaders in the NAF should practice leadership, and thus emphasizes leadership behaviours. *Mission-oriented*, *interaction-oriented* and *development-oriented* leadership behaviours are highlighted as essential for effective military leadership (Forsvarsstaben, 2012). Due to its importance in the military context, *role model*-orientation is added as a fourth category, and the importance of balancing these behaviours is stressed (Forsvarsstaben, 2012).

In recent years, initiative has been taken to further operationalize these leadership behaviours. Based on Yukl's (2012) taxonomy of leader behaviours, a context-specific theory of military leadership has been developed (Martinsen et al., unpublished). *Balanced leadership behaviour* has been introduced as an operationalization of individual prerequisites for practicing mission command (Johansen et.al., 2019). The adaption of Yukl's (2012) taxonomy for the NAFs' context includes the three meta-categories *mission focus*, *development focus* and *role model*. Mission focus reflects task orientation, development focus reflects dimensions of both relations and change orientation, while the role model category was added based on its important status in the NAFs' practice and culture (Martinsen et al., 2019). Each of these meta-categories consist of four facets, which together make up balanced leadership behaviour (Table 3.1).

Table 3.1

Military Leadership: Three main categories with facets (Martinsen et al, 2019)

Mission focus	Development focus	Role model
Following up	Stimulate independence	Organization orientation
Goal orientation	Communicate mastery	Willingness to learn
Clarifying roles	Stimulate innovative thinking	Consideration
Action orientation	Stimulate development	Integrity

Where the trait approach to leadership assumes that the endowment of certain traits makes some individuals natural leaders (Yukl, 2013), a behaviour approach implies that leadership can be trained and developed, as behaviours are more adaptable than traits (Cooper, 2010). Balanced leadership behaviour is indeed regarded a skill that can be trained and developed (Johansen et al., 2019), and constitutes a framework for leadership development within the NAF (Hæren, 2018). The conceptualization of BLB as a trainable skill aligns with the notion that military leadership requires mastery of specific military competencies achieved through education and practice in the military profession (Grebstad & Johnsen, 2019). Meta-analytic findings suggest that leadership training, given certain facilitative conditions, has a positive effect on several evaluation criteria (Lacerenza, Reyes, Marlow, Joseph & Salas, 2017). As the OCS is a cornerstone in the NAFs' efforts to train and develop future military leaders, one would expect this effort to promote the knowledge, skills and attitudes required for practicing effective military leadership. Followingly, we hypothesize that:

H1: Higher levels of performance in military leadership education and training is facilitative of higher levels of military leadership performance.

The operationalization of the specific leadership behaviours required for military leadership has facilitated the practical use of the framework in selection. Five specific leader competencies derived from "*the Chief of Defence's perspective on leadership*" are implemented as criteria for assessing leadership potential during the Joint Admission and Selection process. These include *role model*, *solving mission*, *interaction* and *development*. A fifth construct, *mental robustness*, has been identified and added as important to assess candidates on (Forsvarets personell- og vernepliktssenter, 2019a). Taken together, these leader competencies may be viewed as a competency profile for the role selected for, and through the use of selection methods, candidates are assessed against this profile.

3.2 Selection Methods

The practice of personnel selection is underpinned by the assumptions that there are individual differences between people in terms of skills, abilities and other personal characteristics, hence people are not equally qualified for all jobs, and that it is possible, to some extent, to predict future behaviour and estimate future job performance (Arnold et al., 2016). The purpose of selection methods is to measure individual characteristics to provide estimates of likely future job performance. The ability to predict future job performance and job-related learning, *predictive validity*, is the most important property of any selection method, and is directly proportional to its practical economical value (Schmidt & Hunter, 1998). In order to assess a method's predictive validity, the relationship between the construct measured by the method (the predictor) and subsequent work behaviour indicated by performance measures (the criterion) must be analysed (Arnold et al., 2016).

A selection process starts with the identification of the individual characteristics likely related to job performance. This requires information about the job role and information about what is valued – that is, criteria used to judge performance on the job (Viswesvaran & Ones, 2018). Conducting a job analysis generates systematic information about the job or role in question, while a competency analysis helps translate this information into describing observable behaviours underpinning the tasks and responsibilities of the role (Arnold et al., 2016). Competencies may be defined as "sets of behaviours that are instrumental in

the delivery of desired results or outcomes" (Kurz & Bartram, 2002, pp. 229). Underpinning these behaviours are individual characteristics – personality traits, abilities, motivation and knowledge – that form a competency potential (Kurz & Bartram, 2002). Given facilitative situational factors, this competency potential translates into observable behaviours essential for reaching desired results or outcomes. In selection settings, direct access to observing work behaviours is constrained, and therefore the underlying individual characteristics facilitative of desired behaviours is targeted instead (Skorstad, 2015).

Once the individual characteristics required to perform the tasks of a role are identified, these may be used as selection criteria. These individual characteristics must be measured, for which different selection methods may be used (Salgado, Viswesvaran & Ones, 2001). The choice of selection method and composition of selection tools will depend on both the organization's competency needs, resource constraints, and the methods' predictive validity (Lai, 2010).

There is a wide range of different methods that can be used in selection (Lai, 2010; Martinsen, 2009; Schmidt & Hunter, 1998; Skorstad, 2015). No method alone seems to be perfect (Schmidt & Hunter, 1998), but by combining different selection methods into a system, accuracy in selecting the candidates most likely to achieve the highest levels of job performance increases (Lai, 2010). Performance in any given role is likely to require several competencies with potentially differing underlying individual characteristics, which prompts the use of several selection methods. When several predictors are combined in a selection system, a key question is how much each additional predictor increases the predictive validity for job performance. The *incremental validity*, and ultimately increase in utility, will depend on not only the predictive validity of added predictors, but also on the correlation between the various predictors (Schmidt & Hunter, 1998). In order to assemble a selection system able to predict performance with the greatest possible accuracy, selection tools should be combined so as to tap into and measure the various dimensions that make up performance.

In the selection process for admission to the OCS, candidates are not selected for a specific job, but rather a general role as lower-rank leader, which makes it difficult to conduct specific job analyses to base selection on. Instead, the selection process is based on the analysis of competencies required of any leader in the NAF to practice mission command, as specified in "*the Chief of Defence's perspective on leadership*".

Following the four-step selection model previously presented, the following sections review the literature of the selection methods included in this selection system. Specifically, we focus on GPA, GMA, the selection interview and the assessment centre as predictors performance in education and training and subsequent job performance. The literature on physical ability tests is not reviewed, as these are included in the selection process merely to ensure that candidates satisfy the physical requirements for being able to complete education and training and being fit for the officer role (Forsvarets personell- og vernepliktssenter, 2016).

3.2.1 Grade Point Average

Employers have generally believed that grades help understand who will perform a job well (Campion, 1978; Zikmund, Hitt, & Pickens, 1978), arguing that grades are useful predictors because they reflect intelligence, motivation, and other abilities applicable to the job (Baird, 1985). What grades *do* reflect is academic performance, outcomes of successful behaviours in an educational system, and capture outcomes of learning in academic environments (Viswesvaran & Ones, 2018). In terms of individual characteristics, grades are primarily determined by cognitive ability (Kuncel & Hezlett, 2007), and to some extent personality factors such as conscientiousness and openness (Connelly & Ones, 2010). While school grades are not an intelligence measure, previous research has found strong correlations between school performance and intelligence (Roth et al., 2015).

Meta-analytic findings suggest that grade point average (GPA) to some extent predicts job performance as measured by supervisory ratings, with reported validity coefficients of .11 (O'Leary, 1980, as cited in Hunter & Hunter, 1984, pp. 85), .14 (Reilly & Chao, 1982), .16 (.33 when corrected for range restriction) (Roth, BeVier, Switzer & Schippman, 1996) and .34 (Schmidt, Oh & Shaffer, 2016). Performance in training is also to some extent predicted by GPA, though this has been less extensively studied than the relationship between GPA and job performance. Meta-analytic findings report validity coefficients of .30 (O'Leary, 1980, as cited in Hunter & Hunter, 1984, pp. 85).

Research from within a military context also suggests a relationship between GPA and various performance measures. Studies from the US Defence Department suggest that the best single predictor for adapting to military training is high school diplomas (Farr & Tippins, 2010). The same studies also show a strong link between higher education and military training (Farr & Tippins, 2010). Secondary education

is an established predictor of officer training, and academic performance in military education in particular has been shown to correlate with this (Alf, Neumann & Mattson, 1988). Studies conducted using samples from the NAF suggest that high school GPA is predictive of academic performance in military education and training (Kjenstadbakk, 2012; Norrøne, 2016; Vik, 2012), but not supervisory leadership ratings (Kjenstadbakk, 2012; Vik, 2012).

While GPA has been found to be a fairly good predictor of performance in work and training, some issues are related to the use of GPA for selection purposes. Comparing the GPA of candidates may not be fair, as there are likely to be differences in the harshness of different graders and course difficulty across learning institutions (Viswesvaran & Ones, 2018). Increasing levels of grade inflation over time (Kostal, Kuncel & Sackett, 2016) further add to the potential unfairness of ranking and comparing candidates based on GPA. Grades have also been found to show a simplex pattern of validity (Lievens, Ones & Dilchert, 2009), meaning that validity decreases as the temporal distance of criterion measurement increases (Roth et al., 1996). Most candidates applying for the OCS are quite young and have recently graduated high school, thus the issue of simplex validity patterns may not be of great concern in this context. Moreover, as many candidates lack previous work experience, GPA may be the only indication of previous performance available for assessment.

In the Joint Admission and Selection process, high school GPA is used to rank applicants. While there is no minimum required score, applicants are ranked based on high school GPA (weighted 0.7). Thus, the higher a candidate's GPA, the more likely (s)he is to proceed in the process. The rationale for using high school GPA as a selection criterion is that, although the selection process mainly aims to assess leadership potential, criteria must also be included to predict and ensure that admitted candidates will succeed academically during education. High school GPA is also used (in addition to GMA scores) to calculate an academic prognosis which constitutes part of the basis for final selection decisions. Based on these previous findings concerning GPA as a predictor of job and training performance, the following hypotheses are postulated:

H2: High school GPA is a significant predictor of military leadership performance.

As the nature of the OCSs' theoretical aspects is more closely aligned to traditional academic environments, we hypothesize that:

H3: High school GPA is a stronger predictor of academic performance in theoretical subjects than performance in practical subjects during military education and training.

3.2.2 Intelligence and General Mental Ability

Intuitively, it is hard to imagine any job where some extent of cognitive ability is not required in order to perform work tasks. Broadly defined, cognitive ability refers to the capacity to process, understand, reason with and remember information (Dilchert, 2017). A large number of different but highly related constructs constitute cognitive ability, and the commonality of these are commonly termed intelligence (Dilchert, 2017). Hierarchical models of intelligence posit the existence of a single general factor *g* (Carroll, 1993), describing one's *general mental ability* (Spearman, 1904). This general factor is collectively defined by different specific aptitudes, such as verbal aptitude, spatial aptitude, and numerical aptitude. Specific aptitude theory hypothesizes that performance in different jobs requires different cognitive aptitudes and, therefore, regression equations computed for each job incorporating measures of several specific aptitudes will optimize the prediction of performance on the job and in training (Schmidt & Hunter, 2004).

Murphy (1996) argued that hierarchical models suggest that general versus specific ability constructs can be used for different purposes. However, general mental ability is found to be the most valid predictor of supervisory ratings, correlating above .50 with later occupational level, performance in job training programs, and job performance (Schmidt & Hunter, 2004). Whether additional cognitive factors provide incremental validity is a function of how broadly or narrowly one defines cognitive ability and job performance (Schmitt, Cortina, Ingerick, & Wiechmann, 2003).

General mental ability (GMA) is measured through cognitive ability tests, which seek to uncover candidates' intelligence, including the ability to record, store and use information (Skorstad, 2015). Measures of GMA are widely used for selection across professions and countries, and several large meta-analyses have documented the predictive validity of GMA for job performance, with validity coefficients ranging from .45 to .70 (Hunter & Hunter, 1984; Salgado & Anderson, 2003; Salgado, Anderson, Moscoso, Bertua, De Fruyt, 2003; Schmidt & Hunter, 1998, Schmidt et al., 2016). Schmidt and Hunter (1998) found that cognitive ability measures are among the most valid predictors of job performance across all job

situations. Cognitive ability measures have also been shown to be the best available predictor for job-related learning, that is, acquisition of job knowledge on the job (Schmidt & Hunter, 1992; Schmidt, Hunter & Outerbridge, 1986).

Theories about the determinants of job performance hold that the major direct causal impact of GMA is on the acquisition of job knowledge (Schmidt & Hunter, 1992; Schmidt et al., 2016). According to Hunter (1986), intelligent people have higher levels of job performance because they acquire more job knowledge more rapidly. The knowledge of how to perform the job is what causes higher levels of job performance (Hunter, 1986). Hence, the measurement of GMA is of particular utility for positions requiring rapid learning and high change capacity (Hunter, 1983; Schmidt, 2002; Schmidt, Hunter & Outerbridge, 1986).

Research findings further suggest that the relationship between GMA and job performance is greatest for complex and cognitively demanding tasks (Bertua, Anderson & Salgado, 2011; Farr & Tippins, 2010; Salgado et al., 2003; Schmidt, Hunter & Outerbridge, 1986). The reasoning behind this is that inconsistent tasks are complex and continue to draw on cognitive resources and require cognitive information processing, which has shown to have a large correlation with GMA over time (Ackerman, 1986, 1987, 1992).

In line with this, it is assumed that the relationship between intelligence and leadership is greater than that between intelligence and non-leaders, which might be due to the fact that leaders are required to solve poorly defined problems (Schmidt & Hunter, 1992). However, meta-analytic findings suggest that intelligence is only moderately related to leadership (Judge, Colbert & Ilies, 2004). Haslam, Reicher and Platow (2011) emphasize that the importance of the leader's abilities is situational, provided that different abilities are required for different leadership situations, which aligns with Yukl's (2013) proposition that the requirements for leadership effectiveness is contingent on leadership level. GMA is also found to predict movement in the job hierarchy, where individuals move into higher-complexity jobs if their GMA exceeds the complexity level of their current jobs (Wilk, Desmarais & Sacket, 1995). This is particularly relevant in the NAF, characterized by frequent job rotation with following job management requirements, rapid development and frequent organizational changes.

Research findings suggest that GMA also predicts performance and learning in job training programs (Hunter, 1986, Hunter & Hunter, 1984; Ree & Earles, 1992; Schmidt, Shaffer & Oh, 2008), and academic achievement (Richardson,

Abraham, Bond & Hinshaw, 2012). As with the relationship between GMA and job performance, GMA predicts performance in these domains through its role in knowledge acquisition and learning processes (Dilchert, 2017).

Research conducted with military samples report similar findings, indicating that GMA is a strong predictor of performance in military work and training (Carretta et al., 2014; Hunter, 1986; McHenry, Hough, Toquam, Hanson & Ashworth, 1990; Ree & Earles, 1991). Studies within the NAF have reported mixed findings of the relationship between GMA and various performance criteria during OCS education and training. Both Kjenstadbakk (2012) and Norrøne (2016) report that GMA is a significant predictor of OCS GPA, while only Kjenstadbakk (2012) finds GMA to be predictive of leadership assessments.

The concept of GMA has however received criticism for having minimal utility (Guilford, 1988). Extensive research conducted in the US Army highlights the importance of special factors beyond general intelligence (Farr & Tippins, 2010). These studies concluded that the importance of a general intelligence factor remains strong, but that it cannot rule out the need for more specific predictors (Farr & Tippins, 2010). However, multiple studies have indicated that weighted combinations of specific aptitudes (e.g., verbal, spatial, or quantitative aptitude) tailored to individual jobs do not predict job performance better than GMA measures alone, disconfirming specific aptitude theory (Hunter, 1986; Jensen, 1986; Sacket & Wilk, 1994; Schmidt, Ones & Hunter, 1992; Thorndike, 1986).

The NAF uses its own developed GMA measure ("Alminnelig Evnenivå"), as a measure of general learning ability, numerical and general reasoning, and general conceptual understanding (Skoglund, Martinussen & Lang-Ree, 2014). This measure has been used in the NAF for a long time and is well-documented as a cost-effective, reliable and valid predictor of job performance (Sundet, Barlaug & Torjussen, 2004). In the Joint Admission and Selection process, GMA scores are used to rank and screen applicants, based on the rationale that intelligence is considered the strongest predictor of scholastic achievement (Roth et al., 2015) and job-related learning (Schmidt & Hunter, 1998). The minimum admission criterion is set to GMA >/=5 (Forsvarssjefen, 2008), which equals an IQ of 100 (Forsvarets Sanitet, 2018), because this is shown to be the required level of cognitive ability in order to be able to follow the progression and complexity of both education, training and work (Forsvarets personell- og vernepliktssenter, 2019a). GMA scores are also used (in addition to high school GPA) to calculate an academic prognosis which

constitutes part of the basis for final selection decisions. Based on these previous findings concerning GMA as a predictor of job and training performance, the following hypotheses are postulated:

H4: GMA is a significant predictor of military leadership performance.

H5: GMA is a significant predictor of performance in military education and training.

3.2.3 Selection Interviews

Interviews are one of the most frequently used selection procedures (Macan, 2009; McDaniel, Whetzel, Schmidt & Maurer, 1994; Ryan & Ployhart, 2014), designed to predict future job performance based on applicants' oral responses to oral inquires. However, the selection interview is not a universal concept. In an extensive meta-analysis of the predictive validity of employment interviews, McDaniel and colleagues (1994) distinguish among interviews by the dimensions *interview content*, *how the interview is conducted*, and *the nature of the criterion*, and find that the interview's predictive validity depends on these various factors.

Interview content refers to what kind of questions are asked during an interview. Behavioural interviews involve asking interviewees to describe job-relevant behaviour in past situations, and is based on the premise that past behaviour predicts future behaviour (Janz, 1989). Thus, interviewers search for evidence of previously demonstrated desired behaviours, suggesting capability of similar behaviours in future job situations (Arnold et al., 2016). In contrast, situational interviews are based on goal-setting theory (Locke & Latham, 1990), the underlying assumption being that behavioural intentions predict future behaviour (Latham & Saari, 1984). Candidates are asked to indicate how they would behave in hypothetical situations, and behaviourally anchored rating scales are used to rate and compare responses with those provided by others (Arnold et al., 2016).

Meta-analytic results suggest that situational interviews yield high predictive validity for job performance (McDaniel et al., 1994). However, a key problem with situational interviewing is that it does not account for different levels of experience. Experienced applicants may have a better understanding of what is required in the setting, and thus have an advantage over inexperienced applicants (Arnold et al., 2016). Job complexity has been found to influence the validity of situational but not behavioural interviews (Huffcutt, Conway, Roth & Klehe, 2004),

where only behavioural interviews significantly predict job performance for complex jobs (Krajewski, Goffin, McCarthy, Rothstein, & Johnston, 2006). Nevertheless, both situational and behavioural interviews are found to have good criterion-related validity, meaning that interview performance is predictive of job performance (Klehe & Latham, 2006). While both yield validity, behavioural interviews may be slightly better for more complex jobs (Levashina, Hartwell, Morgeson, & Campion, 2014).

The manner of information collection in an interview relates to the extent of standardization across candidates. Interview structure is defined as "any enhancement of the interview that is intended to increase psychometric properties by increasing standardization or otherwise assisting the interviewer in determining what questions to ask or how to evaluate responses" (Campion, Palmer, & Campion, 1997, p. 656). One can distinguish between unstructured and structured interviews, where structured interviews typically are based on a job description, follow a set format with pre-determined questions, ask the same set of questions in the same order to all candidates, and note and score responses following a formalized scoring guide. Unstructured interviews tend to lack these characteristics, potentially reducing measurement reliability (McDaniel et al., 1994).

For predicting job performance, meta-analytic findings indicate that the structured interview (regardless of content) is more valid than the unstructured interview (McDaniel et al., 1994; Schmidt & Hunter, 1998). The reason for this is that standardization ensures consistency in questions across interviewers and interviewees and in the set of criteria used to evaluate responses (Arnold, et al., 2016). Wiesner and Cronshaw (1988) found a corrected validity of r = .31 for unstructured interviews and r = .62 for structured interviews. Schmidt & Hunter (1998) report similar coefficients of r = .38 for the unstructured and r =. 51 for the structured interview. For predicting training performance, meta-analytic findings indicate similar validity coefficients for both the structured (r=.34) and unstructured (r=.36) interview (McDaniel et al., 1994). Similarly, Schmidt & Hunter (1998) report a validity coefficient of .35 for both structured and unstructured interviews.

The interview differs somewhat from other selection methods in that its social nature may allow for human bias to affect interview scores, potentially reducing predictive validity. However, by increasing interview structure, the influence of biases may be reduced, ultimately increasing the interview's predictive validity (Posthuma, Morgeson & Campion, 2002). Barrick, Shaffer, & DeGrassi

(2009) found that the relationship between verbal impression management and interview ratings was lower for high-structure than for low structure interviews, suggesting that structure narrows and directs focus to job-relevant information.

Contrary to other selection methods such as GMA tests, the interview is not a measure of a single psychological construct. Noting that it remains unknown exactly what combination of constructs the interview measures, Schmidt and colleagues (2016) suggest that it probably measures a combination of previous experience, mental ability, personality traits, and specific job-related skills and behaviour patterns. Meta-analytic findings support that interviews to some extent measure mental ability, reporting corrected average correlations between GMA and unstructured interviews (r=.50) and structured interviews (r=.35) (Huffcutt, Roth & McDaniel, 1996). Schmidt and Hunter's (1998) findings of differing incremental validity for the structured and unstructured interview in combination with GMA measures (24% vs. 8%) further reflects such a difference.

In the Joint Admission and Selection process, candidates undergo a structured situational interview, in which hypothetical questions are posed to assess five personal characteristics essential for leadership in the NAF: *role model, solving mission, mental robustness, interaction* and *development*. Responses are assessed according to a scoring guide, and an overall score is used (together with field exercise scores) to set a leader prognosis, indicating expectations of military leadership development potential. The choice of hypothetical questioning is based on the fact that applicants are young with limited previous experience to reference. Research from within the NAF indicates that the interview significantly predicts subsequent performance measures, also when controlling for GMA (Isaksen, 2014, as cited in Forsvarets personell- og vernepliktssenter, 2019a; Kjenstadbakk, 2012; Norrøne, 2016; Vik, 2013). Based on the review of findings concerning the validity of the selection interview, the following hypotheses are postulated:

H6: The interview is a significant predictor of military leadership performance.

Given the interview's emphasis on behaviours, likely to be more relevant in mastery of practical aspects of education and training, we hypothesize that:

H7: The interview is more predictive of performance in the practical aspects of education and training than in theoretical aspects.

3.2.4 Assessment Centre/Field Exercises

An assessment centre (AC) consists of a standardized evaluation of behaviour based on multiple sources (Rupp, et al., 2015), where an organization can profile applicants' ability across a range of competencies and highly job-related contexts (Arnold, et al., 2016). The design of an AC reflects the need to assess the extent to which applicants can demonstrate a range of competencies which define the key knowledge, skills and abilities required to perform the role in question. ACs bundle various simulations to elicit behaviour required in the actual content of a role that can be evaluated, scored, and used in the prediction of work-relevant criteria. The goal is to have candidates show that they can perform appropriately in a variety of realistic job-related situations, whilst trained assessors independently observe and rate candidate behaviours across different exercises (Viswesvaran & Ones, 2018). A systemic procedure is used for recording and rating specific behaviours as they occur, and independent assessor ratings and reports are then brought together to form an overall rating for each candidate in a moderation session where the candidates' performance is discussed (Arnold, et al., 2016).

The appeal of ACs lies in their generally good levels of criterion-related and face validity (Hough & Ones, 2001) and ability to provide insight into relevant dimensions that prove hard to evaluate through other selection methods (Borman, 1982; Meriac, Hoffman, Woehr & Fleisher, 2008). Although the predictive validity of ACs has been much debated, meta-analytic findings indicate that the average validity of AC studies is very good, because assessment is based upon direct observation of job-relevant behaviours (Arnold, et al., 2016). However, a common finding in the AC literature relates to the problem of convergent versus discriminant validity (Lance, 2008). Cross-situational consistency across exercises rather than within exercises indicates discriminant validity, meaning that different competencies are evaluated separately in each exercise. However, in most ACs, assessors are more likely to provide similar ratings for an individual across different dimensions within the same exercise, rather than for the same competency across exercises (Robertson & Smith, 2001). This represents *convergent validity* and suggests that exercises, not dimensions, are the important construct behind ratings.

This poses a problem to whether ACs measure what they intend to. Kuncel and Sackett (2014) argue that the construct validity issue may not actually be that problematic, as the competency scores for each exercise are only a *step* towards a final overall AC rating for each competency. They argue that if focus is shifted to

this overall AC competency rating, exercise-specific effects are no longer the dominant source of final AC ratings when ratings for each competency across all of the AC exercises are combined (Kuncel & Sackett, 2014).

Several meta-analyses have reported the validity of overall AC ratings (Viswesvaran & Ones, 2018). Schmitt, Gooding, Noe, and Kirsch (1984) found that the AC was a good predictor of job performance ratings (r = .43) and grades (r = .31), whilst Hunter and Hunter (1984) report that AC also predicts promotion (r = .63). Similar findings are also found in analyses conducted in a military setting (Dobson & Williams, 1989; Melchers & Annen, 2010) and within the NAF (Isaksen, 2014, as cited in Forsvarets personell- og vernepliktssenter, 2019a; Kjenstadbakk, 2012; Norrøne, 2016).

During the Joint Admission and Selection process, candidates undergo a field exercise consisting of multiple evaluation methods intended to obtain information about candidates' prerequisites and development potential, where trained assessors independently observe and rate candidate behaviour. The field exercise may thus be described as a form of assessment centre. During the field exercise, candidates are assessed on the same five personal characteristics essential for leadership in the NAF as during the interview: *role model*, *solving mission*, *mental robustness*, *interaction* and *development.* Field exercise scores are used (together with interview scores) to set a leader prognosis, indicating expectations of military leadership development potential. While in the interview scores are given based on candidates' reports of how they would behave in hypothetical situations, the field exercise generates ratings of actual displayed behaviours.

The rationale for measuring these same leadership characteristics through field observation is that self-reports of behaviours may not be completely accurate. For example, while most candidates report that they will master long walks with heavy equipment and the use of armed weapons, 10-20% of candidates resign from the selection process after exposure to this during the first night of the field exercise (Forsvarets personell- og vernepliktssenter, 2019a). As these tasks are absolute minimum requirements for what a candidate should master, their assessment is vital before embarking on a long and costly military leadership education. Based on the abovementioned findings concerning the predictive validity of assessment centres, the following hypotheses are postulated:

H8: The field exercise is a significant predictor of military leadership performance.

Given the closer resemblance in context between the field exercise and practical aspects of education and training, we hypothesize that:

H9: The field exercise is more predictive of performance in the practical aspects of education and training than in theoretical aspects.

4. Research Methodology and Design

This chapter explains and describes the study's methodological considerations. The basis for the data sample is introduced, along with data sampling procedures. The study's variables are presented, and the measurement of these is explained. Finally, the research design is elaborated on, followed by a description of the study's statistical analysis.

4.1 Data Sample and Data Collection

Our study is based on data from the NAFs' "Leader Candidate Study 2015-2020" (Forsvarets lederkandidatstudie 2015-2020) (Stabsskole, 2015), a study conducted to further develop the selection and education of officer candidates in the NAF and enhance competencies revolving leadership, military psychology and selection. The "Leader Candidate Study 2015-2020" is approved by Norwegian Centre for Research Data (Norsk samfunnsvitenskapelig datatjeneste). The study monitors candidates who pursue the OCS selection process throughout selection, education and training, with the aim of controlling the selection process' predictive validity. The study is based on a survey offered to every candidate at the outset of the selection process. Applicants answer a questionnaire including consent that answers can be linked to information regarding the selection tests and future performance measures from education and practice. Through informants at the *Norwegian Defence University College* (NDUC), we gained access to anonymized selection data from the "Leader Candidate Study" for candidates admitted to the OCS in 2016 and 2017. The choice to include two cohorts significantly increased the sample size, which serves to improve the reliability of results.

The selection data were coupled by our informants at NDUC to data from another research project within the NAF aimed at developing a context specific theory and a new measure of military leadership (Martinsen et al., unpublished). In this project, following an initial conceptual identification of military leadership, an instrument was developed to measure balanced leadership behaviour (BLB). In the

validation process of this instrument, a total of 5,374 military personnel has participated by responding to the instrument's items, providing evaluations of subordinates', peers', superiors' or own leadership behaviours.

We chose as our sample the 259 subjects for whom selection data and BLB scores could be coupled. For 125 of these subjects, we were able to acquire data from OCS exam protocols consisting of grades obtained during the education's first year. These data were anonymized and coupled to selection data and BLB scores by our informants at NDUC. The sample includes 106 candidates admitted to the OCS in 2016 and 153 candidates admitted in 2017, from each of the three military branches, the Army (n=207), the Navy (n=19), and the Air Force (n=33).

The data for our sample were generated at three points in time (T1, T2 and T3). Candidates' scores from the selection process were generated in advance of the candidates' admission to the OCS (T1). Academic results were generated during the candidates' first year of education and training (T2). BLB evaluations were generated at the completion of the education's second year of practice (T3). Given the inclusion of two cohorts, data is generated over a period of three years in total.

4.2 Variables and Measurement Tools

The dataset contains 12 variables. The variables containing data collected during the selection process (T1) are referred to as this study's predictors, as this data is collected with the aim of predicting future performance. These variables include *1. High School GPA, 2. GMA, 3. Physical Strength, 4. Physical Endurance, 5. Interview, 6. Field Exercise, 7. Academic Prognosis, 8. Leader Prognosis*.

The variables containing grades obtained during the OCSs' first year of education, *9. GPA Total, 10. GPA Practical Subjects* and *11. GPA Theoretical Subjects,* constitute one set of criteria against which the selection process' methods are validated. In analysing to which extent the selection system predicts BLB when controlling for the effect of performance in education and training, *10. GPA Practical Subjects and 11. GPA Theoretical Subjects* serve as predictor variables.

The selection methods included in the selection process are hypothesized to not only predict academic performance, but also leadership potential and performance in the role as a lower-level military leader. The variable containing BLB scores generated at the completion of the education's second year of practice, *12. Balanced Leadership Behaviour,* constitutes this study's second criterion. Table 4.1 lists the study's variables.

Table 4.1

Overview of the study's variables

Variable	Explanation
1.GPA (Scale 1-66)	Grade point average from high school diploma
2.GMA (Scale 1-9)	Measure of general learning ability, numerical and general reasoning, and general conceptual understanding.
3.Physical Strength (Scale 1-9)	Average score of 3 strength tests. Push-ups, sit-ups and pull-ups (2016) / Medicine ball throw, standing long jump and pull-ups (2017).
4.Physical Endurance (Scale 1-9)	3000-meter track run
5.Interview (Scale 1-9)	Measure of leadership potential defined by 5 leadership competencies: role model, solving mission, mental robustness, interaction and development. Average of scores across these 5 dimensions.
6.Field Exercise (Scale 1-9)	Measure of leadership potential defined by 5 leadership competencies: role model, solving mission, mental robustness, interaction and development. Average of scores across these 5 dimensions.
7.Academic Prognosis (Scale 1-9)	The candidate's presumed potential for succeeding academically during the education. Derived by weighting and combining GPA and GMA.
8.Leader Prognosis (Scale 1-9)	The candidate's presumed potential for succeeding as a military leader after completing education and training. Derived by weighting and combining leadership potential scores assessed through interview and field exercise.
9.GPA Total (Scale 0-6)	Grade point average of all subjects undergone during education and training.
10.GPA Practical Subjects (Scale 0-6)	Grade point average of all practical subjects undergone during education and training.
11.GPA Theoretical Subjects (Scale 0-6)	Grade point average of all theoretical subjects undergone during education and training.
12.Balanced Leadership Behaviour (Scale 1-7)	Measure of leadership behaviour. Average of scores across three sub-dimensions: role model, mission focus and development focus.

Most of the study's variables are discrete numeric variables at the ordinal level. However, for the purpose of performing the statistical analyses required to answer the problem formulation, all variables are treated at the interval level. Variables not initially at the ordinal level have been recoded to meet the criteria for inclusion in the analyses.

1. High School GPA is measured on a continuous scale ranging from 1 to 66 and is calculated by multiplying high school GPA (1-6) by 10 and adding up to 6 potential additional credits (up to 4 STEM credits and 2 additional credits, e.g. military service points).

2. GMA is measured on a 9-point scale where 1 is the lowest score and 9 is the highest. This score is generated by calculating the average of scores on three tests (also measured on 9-point scales) measuring aspects of intelligence: *Raven, Numerical Reasoning ("Tallrekker"),* in the form of number series sequence tests, and *Verbal Ability ("Ordforståelse")* assessing knowledge of Norwegian and English language. The admission criterion for GMA is a score of 5,0 or higher, although in certain cases candidates with a lower score may be considered. These candidates must re-sit the cognitive ability tests and achieve a minimum average score of 3,0 and a minimum score of 2,0 on each of the three sub-tests.

3. Physical Strength is also measured on 9-point scales, where 1 is the lowest score and 9 is the highest. The physical strength score is generated by calculating the average score of three tests of physical strength. In the 2016 selection process these included push-ups, sit-ups and pull-ups, and the admission criterion was an average score of 2,0 or higher. In the 2017 selection process physical strength was tested through medicine ball throw, standing long jump and pull-ups, and the admission criterion was set at an average score of 5,0 or higher.

4. Physical Endurance was measured by means of a 3000-meter track run both years, also on a 9-point scale, where 1 is the lowest score and 9 is the highest.

5. Interview is also measured on a 9-point scale where 1 is the lowest score and 9 is the highest. The interviews conducted during the selection process are structured situational interviews, where interviewers follow an interview guide for posing questions and a scoring manual for evaluating responses. The purpose of the interview is to generate information about the candidate's leadership potential. Based on "*the Chief of Defence's perspective on leadership"*, questions are derived to measure 5 personal characteristics essential for leadership in the NAF: *role model*, *solving mission*, *mental robustness*, *interaction* and *development*. The *role model* construct involves exhibiting core values of respect, responsibility and courage, and showing integrity in the way leadership is conducted. *Solving mission* involves demonstrating decision making skills and prioritizing and handling various sets of challenges. *Mental robustness* involves mastery of uncertainty and performance under physically and psychologically challenging environments. The *interaction* construct concerns ability to care for others, create trust and delegate tasks to enable role clarity, while *development* involves ability to stimulate and recognize independent thinking in followers. Candidates receive scores on each of these 5 leadership competencies, based on which a leader prognosis is calculated,

indicating how the candidate is expected to perform as a leader at a basic officer candidate level. The leader prognosis derived from the interview is later used as a component in calculating the total *8. Leader Prognosis*. In the data we received, only scores on the individual interview components were included. A total interview score - *5. Interview* - was calculated by averaging the scores of the 5 individual components. Thus, while the individual components are discrete numerical variables, the total interview score is continuous with two decimals. Our data included some observations with missing values on 1-2 of the individual interview components (n=50). Due to time limitations, *solving mission* and *interaction* are sometimes not prioritized in the interview, which explains these missing values. For these subjects, the average was calculated based on the remaining individual component scores.

6. Field Exercise is measured on a continuous 9-point scale where 1 is the lowest score and 9 is the highest. The field exercise score is generated by calculating the average of scores on 5 personal characteristics essential for leadership in the NAF as observed in the field: *role model, solving mission, mental robustness*, *interaction* and *development*. While scores on these sub-components take discrete values ranging from 1-9, the average total score holds two decimals. The field exercise aims to reveal candidates' potential leader characteristics in a field setting. Based on candidates' performance during the field exercise, qualitative evaluations are made by trained observers regarding the 5 various aspects of leadership behaviour. These evaluations are later translated into numeric scores following a scoring guide to set a leader prognosis, which is ultimately combined with the leader prognosis derived in the interview to set the final *8. Leader Prognosis*.

The variable *7. Academic Prognosis* is also measured on a 9-point scale where 1 is the lowest score and 9 is the highest. The academic prognosis is derived by weighting and combining high school GPA and GMA scores, and is a numerical expression of how the candidate is expected to perform academically.

8. Leader Prognosis is derived by weighting and combining leader potential scores as assessed through the interview and the field exercise. The prognosis is measured on a 9-point scale where 1 is the lowest score and 9 is the highest, and constitutes a numerical expression of how the candidate, given education and training, is expected to perform as a leader at a lower-rank level. This prognosis, combined with the academic prognosis constitute the final basis for selection

decisions, and are weighted and combined to derive a final score ranking the remaining candidates at the end of the selection process.

The variables measuring academic performance during the education's first year, *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects*, were calculated based on grades in exam protocols received from representatives at the Army's, Air Force's and Navy's OCSs, while *9. GPA Total* was extracted directly from these protocols. In the Army and Navy, grades are given on a 6-point scale A-F where A is the highest grade, E is the lowest possible pass grade, and F equals fail. These grades were recoded into numeric values where A=5, B=4, C=3, D=2, E=1 and F=0. The Air Force operates with a percentage scale for grading, where 0-39,99% equals grade F, 40-49,99% equals grade E, 50-59,99% equals grade D, 60-79,99% equals grade C, 80-89,99% equals grade B, and 90-100% equals grade A. In order to make grades across all three military branches comparable, percentage grades were recoded first from the interval level (1-100) to the ordinal level (A-F), and then into numeric values where A=5, B=4, C=3, D=2, E=1 and F=0.

The various subjects included in the education vary to some extent between the three military branches. Some subjects are general and included in the education programme regardless of branch, while others are branch-specific. Despite these differences, we consider a GPA score obtained during the education's first year to be a representative measure of performance, comparable across branches. The variable *9. GPA Total* is calculated by averaging all individual subject grades for each student. *10. GPA Practical Subjects* was calculated by averaging all grades in the practical subjects undergone by each student, while *11. GPA Theoretical Subjects* was calculated by averaging all grades in the theoretical subjects undergone by each student. These variables are thus measured on a continuous 6-point scale with scores holding two decimals.

12. Balanced Leadership Behaviour is measured by administering the instrument to candidates' superiors, asking them to retrospectively assess leadership behaviour for the total period for which the candidate had functioned as the assessor's subordinate (a period of 6 or 12 months, depending on military branch). The BLB tool consists of 12 items measured on a 7-point scale where 1 is the lowest score and 7 is the highest – with the exception of item 2 and item 4 which are reversed. Items 1-4 measure the sub-component *role model*, items 5-8 measure the sub-component *mission focus*, and items 9-12 measure the sub-component *development focus*. The data we received included only the raw scores on each of

these 12 items. The total BLB score was calculated by averaging the scores of the 12 items (items 2 and 4 were first reversed). The balanced leadership behaviour variable is hence a continuous variable with two decimals. This variable constitutes the main criterion against which to validate leadership development potential.

4.2.1 Missing Values

Methodological considerations include decisions about how to treat potential missing values in the dataset. Missing values effectively reduce sample size, and ultimately statistical efficiency (Stehlik-Barry & Babinec, 2017). The main concern with missing values is whether they are missing at random or if some factor is systematically causing missing values for a given variable, potentially introducing bias into the data (Stehlik-Barry & Babinec, 2017).

Our dataset has a fair amount of missing values. Variable *2. GMA* has 1 missing value. This is likely due to an error in the data entering phase, and the observation is included in the dataset. Variable *7. Academic Prognosis* has 154 missing values. The reason for this is likely differences between branches and across cohorts in registering this prognosis. In the period 2015-2017, the calculation of this prognosis was changed, with some branches calculating the academic prognosis directly into the final combined academic/leader prognosis score. Because these observations hold valid values across other variables, because the raw scores that would have been entered into the prognosis (high school GPA and GMA) are present, and because deleting them from the dataset would more than halve the sample size (and hence decrease the accuracy of all analyses not including these variables), these observations are included in the dataset.

The T2 variables *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects* have 134 missing values, while *9. GPA Total* has 144 missing values. The reason for these missing values is that in the work with gathering academic results to couple with the existing selection- and balanced leadership behaviour data, only data for 125 students could successfully be coupled. Academic results were coupled (by our informants) by means of matching ID numbers. A likely explanation for the low number of matches is that hand-written ID numbers may be distorted when scanned for digital registration (i.e. 1 is transformed to 7 and vice versa, 6 is transformed to 0, etc.), or differences in ID registration (full ID number versus birth date only). The 10 additional missing values for *9. GPA Total* are likely due to some error or slack in the update and registration of the exam protocols. For these 10

observations, grades in every individual subject were registered in the protocols, but for some reason a final GPA was not listed. Because 125 is the total number of observations for which it was possible to obtain and link academic results to selection data and balanced leadership behaviour scores, these observations are included in the dataset. The 10 observations with missing values on *9. GPA Total* are also included in the dataset, as these hold valid values across other variables, and importantly, for *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects*.

12. Balanced Leadership Behaviour has 2 missing values. This is likely due to the intentional or unintentional skipping of items in the instrument. The observations are included in the dataset.

4.2.2 Reliability and Validity

The purpose of selection methods is to measure individual characteristics in order to provide estimates of likely future job performance. In order to confidently make such predictions, it is vital that the measurement instrument produces *reliable* results (Skorstad, 2015). Reliability concerns the accuracy of measurement scores, and the extent to which these are influenced by random errors of measurement (Cooper, 2010). While the reliability of the NAFs' instrument for measuring GMA has been documented (Sundet, Barlaug & Torjussen, 2004), the reliability of interview and field exercise scores may be more questionable, as these rely on subjective judgements. Cooper (2010) points out that because behavioural ratings often involve observing behaviour only for a short time, and in one or two situations, they may not be highly accurate measures.

Joint Admission and Selection interviewers and field exercise observers are trained to classify behaviours according to a particular checklist, and follow structured guidelines, which may serve to increase the reliability of assessments. Moreover, during the field exercise, behaviour is observed for an extended amount of time across multiple situations, potentially increasing the accuracy of evaluations. The use of two methods for assessing leadership potential (interview and field exercise) may also increase the accuracy of leadership potential scores, as measuring constructs through several instruments tends to cancel out the systematic errors associated with each instrument (Cooper, 2010).

Furthermore, it is also vital that an instrument's measurements reflect the construct the instrument claims to measure (Cooper, 2010). The *validity* of an instrument concerns the extent to which inferences drawn based on scores are valid

(Skorstad, 2015). Of greatest concern in relation to selection is construct validity – the extent to which an instrument measures the construct it claims to measure (Cooper, 2010) -, and predictive validity – the extent to which the instrument may be used to predict a future criterion (Lang-Ree & Martinussen, 2019).

One inherent challenge in estimating the predictive validity of a selection method is caused by the fact that performance measures can only be obtained for individuals who pass through a selection process and are hired. This effect of a sample selection process, resulting in an observed sample not representative of the population of interest, is referred to as range restriction (Sackett & Yang, 2000). When predictor scores are used as selection criteria, the resulting workforce may not provide the full theoretically possible range of scores on the predictor and/or criterion measures, resulting in underestimated validity coefficients (Arnold et al., 2016). Figure 4.1 illustrates this phenomenon.

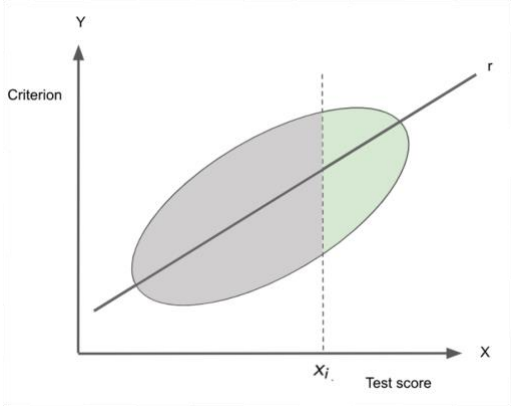

Figure 4.1: Model of the range restriction phenomenon.

Scores falling below the cut-off at x_i constitute "missing data" in analyses. Several of this study's predictors have cut-off values, hence range restriction may result in underestimated validity coefficients. This may be particularly prominent for GMA measures, where the cut-off value is set at 5 on a 1-9 scale. Corrections for range restriction are not conducted in this study.

4.3 Research Design and Statistical Analysis

The aim of this study is to assess the predictive validity of the selection process for admission to the NAFs' Officer Candidate School. This selection system is assembled with the objective of identifying the candidates with the highest potential for becoming military leaders, and prospering during education and training. Thus, the predictive validity of the selection process is assessed

against both education and training performance criteria, and subsequent military leadership evaluations. Furthermore, given that military leadership requires mastery of unique military competencies attainable through education and training in the military profession (Grebstad & Johnsen, 2019), we assess the relationship between performance in education and training and subsequent leadership evaluations.

In such, this study's design is longitudinal, examining the predictive validity of variables collected at one point (T1) for performance measures collected at later points in time (T2 and T3). The predictor variables include all scores collected during the selection process. Academic results and BLB assessments collected after the first and second year of the educational programme, respectively, constitute the performance criteria. Figure 4.2 visualizes the timeline of data generation and the postulated hypotheses in the study's research model.

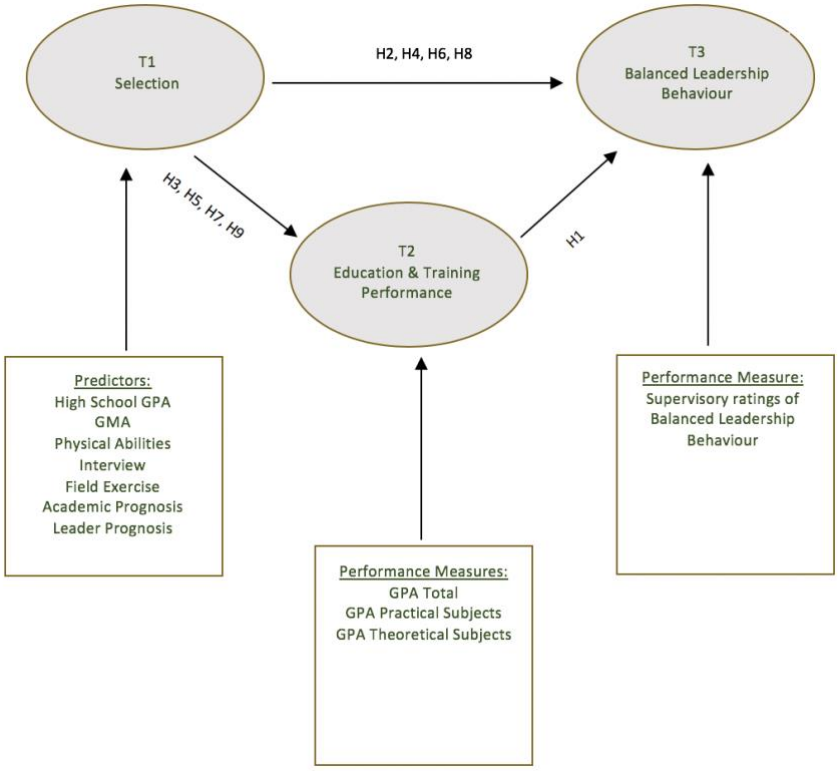

Figure 4.2: Research model with hypotheses.

In order to investigate the predictive validity of the selection process' methods and test hypotheses, we conducted both correlational analyses and multiple hierarchical regression analyses, to examine to what extent the predictors (selection methods) explain the variance in the criteria (performance measures). Correlation analyses show the strength of the correlation between predictors and criteria. Multiple hierarchical regression analyses illustrate to what extent a set of predictors predict a given criterion, and how much its variance can be explained by

the dependent variables (Martinussen & Hunter, 2008). Multiple hierarchical regression analyses also allow for evaluating the predictive validity of each predictor while controlling for the effect of the other (Agresti & Finlay, 2009).

The dataset was first explored visually and through descriptive statistics, in order to provide an overview and detect relationships between the independent and dependent variables. Correlational analyses were then conducted, in order to obtain numerical expressions of the relationships between variables. This provided a basis for conducting subsequent regression analyses. Pearson's correlation coefficient (*r*) indicates the strength and direction of a linear relationship between two variables. However, it does not imply causality, meaning that inferences may not be made about the causal relationship between the two variables. For this reason, and because correlation can only be computed for two variables at a time, correlational analyses were not sufficient in order to answer this study's problem formulation.

Multiple linear regression analyses were conducted in order to understand how the output (criteria) changes as a result of changes to the input (predictor) variables (Jank, 2011). By conducting multiple hierarchical regression analyses, we were able to insert different sets of predictors through several steps of analyses. The multiple regression coefficient (R2) expresses how much of the variation in the criteria are explained by variation in the various predictors. Thus, through several steps of analysis, we assessed to what extent the various selection methods and the selection battery as a whole predict 1) performance during military education and training, 2) military leadership evaluations and 3) the contribution of military education and training in explaining variation in military leadership evaluations when controlling for the effect of the selection system.

5. Results

In this chapter we present the results of descriptive statistics and correlation analyses, before presenting the results from the regression analyses. Descriptive statistics are presented to provide an overview over the dataset. Results from correlational analyses indicate the qualitative nature of relationships between variables, such as direction and strength. Finally, regression analysis results indicate to what extent predictors actually predict the various performance criteria.

5.1 Descriptive Statistics and Correlation Analyses

The dataset was first explored through descriptive statistics and visualizations such as histograms and plots to provide an overview of the distribution of the various variables. These variable distributions are presented in Table 5.1, with minimum, maximum and mean values, and standard deviations.

Table 5.1

Descriptive statistics of the study's variables

Variable	n	Minimum	Maximum	M	SD
1.High School GPA (Scale 1-66)	259	28,30	61,70	46,48	5,68
2.GMA (Scale 1-9)	258	3,00	9,00	6,48	1,27
3.Physical Strength (Scale 1-9)	259	3,67	9,00	6,88	1,43
4.Physical Endurance (Scale 1-9)	259	3,67	9,00	7,44	1,35
5.Interview (Scale 1-9)	259	2,60	9,00	6,48	1,34
6.Field Exercise (Scale 1-9)	259	2,60	8,80	6,18	1,37
7.Academic Prognosis (Scale 1-9)	105	3,15	8,35	6,64	0,77
8.Leader Prognosis (Scale 1-9)	259	2,77	8,77	6,24	1,15
9.GPA Total (Scale 0-6)	115	1,83	4,26	3,21	0,51
10.GPA Practical Subjects (Scale 0-6)	125	1,80	4,50	3,45	0,51
11.GPA Theoretical Subjects (Scale 0-6)	125	0,33	4,75	3,06	0,84
12.Balanced Leadership Behaviour (Scale 1-7)	257	2,50	6,83	5,23	0,68

The predictor variables all show a relatively narrow range, with exception of *1. High School GPA* (SD=5,68), which has a much wider measurement scale. The range is particularly low for *7. Academic Prognosis* (SD=0,77). This may be a result of the large number of missing variables (n=105), and the range would possibly be larger had the missing values been present. The range restriction in predictor variables is however expected, as only candidates scoring above cut-off values progress in the selection process. We further note that the distributions of the predictor variables are all slightly skewed, with mean scores falling above the mean of the scales. This is most notable for *4. Physical Endurance* ($\bar{x}$=7,44).

Through visualisations of the distributions of variables *9. GPA Total*, *10. GPA Practical Subjects*, *11. GPA Theoretical Subjects*, these variables appear to be more normally distributed than the predictor variables. The mean scores of these variables fall closer to the mean of the scales: *9. GPA Total* ($\bar{x}$=3,21), *10. GPA Practical Subjects* ($\bar{x}$=3,45), *11. GPA Theoretical Subjects* ($\bar{x}$=3,06). Scores on these variables tend to fall quite close (on average) to the mean, reflecting a somewhat narrow range: *9. GPA Total* (SD=0,51), *10. GPA Practical Subjects* (SD=0,51) and *11. GPA Theoretical Subjects* (SD=0,84).

For the criterion *12. Balanced Leadership Behaviour*, scores are distributed in a slightly skewed manner, with the mean score ($\bar{x}$=5,23) falling above the mean of the scale. On average, scores on this variable tend to fall somewhat less close to the mean (SD=0,68) than do scores for the academic performance criteria variables. The study's table of correlations is presented in Table 5.2.

Table 5.2

Table of correlations

Variable	1	2	3	4	5	6	7	8	9	10	11	12
1.High School GPA	1											
2.GMA	.365**	1										
3.Physical Strength	-.205**	-.091	1									
4.Physical Endurance	-.084	-.080	.501**	1								
5.Interview	.183**	-.016	.069	.018	1							
6.Field Exercise	-.147*	-.067	.229**	.186**	.103	1						
7.Academic Prognosis	.790**	.679**	.060	.162	.167	.030	1					
8.Leader Prognosis	-.099	-.066	.232**	.174**	.328**	.972**	.063	1				
9.GPA Total	.195*	.023	.066	.048	.086	.275**	.324**	.284**	1			
10.GPA Practical Subjects	.056	-.030	.201*	.155	-.036	.291**	.186	.273**	.845**	1		
11.GPA Theoretical Subjects	.486**	.317**	.073	-.047	.332**	.174	.609**	.241**	.402**	.229**	1	
12.Balanced Leadership Behaviour	-.007	-.070	-.006	-.043	.125*	.160*	-.086	.182**	.301**	.321**	-.057	1

Note. N=259. *. Correlation is significant at the 0.05 level (2-tailed), **. Correlation is significant at the 0.01 level (2-talied).

5.2 Regression Analyses

The predictor variables were entered as models into stepwise hierarchical regression analyses against the study's criteria variables. Predictor variables were entered stepwise in the order they are derived during the selection process. The first model contains *1. High School GPA* and *2. GMA*. *3. Physical Strength* and *4. Physical Endurance* were entered in the second step, *5. Interview* in the third step, and finally *6. Field Exercise* was entered in the fourth and last step. Due to issues of multicollinearity, *7. Academic Prognosis* and *8. Leader Prognosis* were entered into a separate model to assess whether these add any value in explaining more of the variance in the criteria than do the individual selection tools. Table 5.3 presents the results for the analysis with *9. GPA Total* as the criterion.

Table 5.3

Multiple hierarchical regression analysis with criterion 9. GPA Total as dependent variable

	Model 1	Model 2	Model 3	Model 4
1.High School GPA	.217*	.228*	.218*	.251*
2.GMA	-.058	-.057	-.053	-.041
3.Physical Strength		.087	.081	.014
4.Physical Endurance		.010	.016	.006
5.Interview			.051	.005
6.Field Exercise				.299**
R2	.041	.049	.052	.132
Adjusted R2	.024	.015	.008	.084
R2 Change	.041	.008	.002	.080
F-value	2.396	1.430	1.193	2.739*

Note: N=115. Standardized Beta coefficients are reported. *p<.05, **p<.01.

1.High School GPA is the only significant variable of the first (β = .217*), second (β = .228*), and third (β = .218*), model, but none of these models as a whole are significant. In the fourth and only significant model, *1. High School GPA* is still a significant predictor (β = .251*) after controlling for *6. Field Exercise*, however *6. Field Exercise* is the model's most significant variable (β = .299**). Hence, we find no support for the hypothesis that GMA is a significant predictor of performance in military education and training (*H5*). Adding *6. Field Exercise* resulted in a substantial increase in Adjusted R Square to .084, meaning that the selection system as a whole is able to explain 8,4% of the variation in *9. GPA Total*.

A separate analysis was performed to assess the predictive validity of *7. Academic Prognosis* and *8. Leader Prognosis* for *9. GPA Total*. Table 5.4 presents the results of this analysis.

Table 5.4

Regression analysis of variables 7. Academic Prognosis and 8. Leader Prognosis with criterion 9. GPA Total

7.Academic Prognosis	.300**
8. Leader Prognosis	.348**
R2	.226
Adjusted R2	.208
F-value	12.686**

Note: N=90. Standardized Beta coefficients are reported, *p<.05. **p<.01.

Both *7. Academic Prognosis* (β = .300**) and *8. Leader Prognosis* (β = .348**) are significant variables in this model. Adjusted R Square of the model is .208, meaning that the academic and leader prognoses as a model can account for

more than double the amount (20,8%) of the variation in *9. GPA Total* than can the model including the entire battery of selection tools.

Correlational analyses suggest that practical and theoretical subjects are not strongly related, and thus that performance in the one may be predicted by different factors than the other. Hence, the same multiple hierarchical analyses were performed with *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects* as criteria, as with *9. GPA Total*. Table 5.5 presents the results for the analysis with *10. GPA Practical Subjects* as the criterion.

Table 5.5

Multiple hierarchical regression analysis with criterion 10. GPA Practical Subjects as dependent variable

	Model 1	Model 2	Model 3	Model 4
1.High School GPA	.084	.108	.119	.155
2.GMA	-.060	-.056	-.059	-.042
3.Physical Strength		.166	.171	.123
4.Physical Endurance		.057	.054	.042
5.Interview			-.057	-.093
6.Field Exercise				.275**
R2	.007	.047	.050	.118
Adjusted R2	-.009	.015	.009	.073
R2 Change	.007	.040	.003	.069
F-value	.429	1.453	1.234	2.619*

Note: N=124. Standardized Beta coefficients are reported. *p<.05, **p<.01.

Neither *1. High School GPA*, *2. GMA*, *3. Physical Strength*, *4. Physical Endurance* or *5. Interview* are significant in the first three models, nor are these models as a whole. The fourth model on the other hand is significant, though *6. Field Exercise* is the only significant variable of this model (β = .275**). Adding *6. Field Exercise* in the fourth model, Adjusted R Square is significantly increased to .073, meaning that the battery of selection tools as a whole is able to explain 7,3% of the variation in *10. GPA Practical Subjects*. We note that while *1. High School GPA* was a significant predictor of *9. GPA Total*, it is not for *10. GPA Practical Subjects*, which lends initial support to the hypothesis that high school GPA is a stronger predictor of academic performance in theoretical subjects than performance in practical subjects during military education and training *(H3)*. Furthermore, these results provisionally disconfirm the hypothesis that the interview is more predictive of performance in the practical aspects of education and training than in theoretical aspects *(H7)*, but lend provisional support to the

hypothesis that the field exercise is more predictive of performance in the practical aspects of education and training than in theoretical aspects *(H9).*

Again, a separate analysis was performed to assess the predictive validity of *7. Academic Prognosis* and *8. Leader Prognosis* for the criterion *10. GPA Practical Subjects*. Table 5.6 presents the results of this analysis.

Table 5.6

Regression analysis of variables 7. Academic Prognosis and 8. Leader Prognosis with criterion 10. GPA Practical Subjects

7.Academic Prognosis	.162
8. Leader Prognosis	.347**
R2	.154
Adjusted R2	.135
F-value	7.941**

Note: N=90. Standardized Beta coefficients are reported. *p<.05, **p<.01.

8. Leader Prognosis is the only significant variable of this model (β = .347**). The model as a whole is significant, with an Adjusted R Square of .135. Thus, the academic and leader prognoses as a model can account for almost double the amount (13,5%) of the variation in *10. GPA Practical Subjects* than can the model of the entire battery of selection tools. Note that while both *7. Academic Prognosis* and *8. Leader Prognosis* were significant predictors of *9. GPA Total*, only *8. Leader Prognosis* is a significant predictor of *10. GPA Practical Subjects*.

To assess the predictive validity of the battery of selection tools for theoretical academic performance, the predictor variables were entered into the same analyses, this time with *11. GPA Theoretical Subjects* as the criterion. Table 5.7 presents the results for the analysis with criterion *11. GPA Theoretical Subjects*.

Table 5.7

Multiple hierarchical regression analysis with criterion 11. GPA Theoretical Subjects as dependent variable

	Model 1	Model 2	Model 3	Model 4
1.High School GPA	.427**	.453**	.405**	.435**
2.GMA	.162	.161	.172*	.185*
3.Physical Strength		.220*	.196*	.156
4.Physical Endurance		-.129	-.118	-.129
5.Interview			.249**	.218**
6.Field Exercise				.229**
R2	.259	.294	.354	.402
Adjusted R2	.246	.271	.326	.371
R2 Change	.259	.036	.060	.048
F-value	21.107**	12.405**	12.924**	13.087**

Note: N=124. Standardized Beta coefficients are reported. *p<.05, **p<.01.

1. High School GPA is the first model's only significant variable (β = .427**). The model as a whole is significant, with an Adjusted R Square of .246. Hence, these two predictors alone are able to explain 24,6% of the variance in *11. GPA Theoretical Subjects*. *1. High School GPA* is still a significant variable (β = .453**) when controlling for *3. Physical Strength* and *4. Physical Endurance*, while *3. Physical Strength* is also a significant variable of the second model (β = .220*). The model as a whole is also significant. By adding *3. Physical Strength* and *4. Physical Endurance* in the second model, Adjusted R Square increases to .271, significantly improving the model. *5. Interview* was added in the third model, resulting in Adjusted R Square increasing yet again to .326. *1. High School GPA* (β = .405**) and *3. Physical Strength* (β = .196*) remain significant variables when controlling for *5. Interview*. In the third model, *2. GMA* (β = .172*) and *5. Interview* (β = .249**) are also significant variables. The third model as a whole is also significant, and can account for 32,6% of the variance in *11. GPA Theoretical Subjects*. In the fourth and final step, *6. Field Exercise* was added, increasing Adjusted R Square further to .371. *1. High School GPA* (β = .435**), *2. GMA* (β = .185*), and *5. Interview* (β = .218**) are still significant variables when controlling for *6. Field Exercise*. *3. Physical Strength* is however no longer significant after controlling for *6. Field Exercise*, while *6. Field Exercise* is a significant variable of the fourth model (β = .229**). The model as a whole is also significant, and is able to account for 37,1% of the variance in the criterion *11. GPA Theoretical Subjects*.

These findings, taken together with those from Table 5.5, lend support to the hypothesis that high school GPA is a stronger predictor of academic performance in theoretical subjects than performance in practical subjects during military education and training *(H3).* Our findings indicate that the interview is a highly significant predictor of performance in the education's theoretical, but not practical aspects. Hence support is not found for the hypothesis that the interview is more predictive of performance in the practical aspects of education and training than in theoretical aspects *(H7).* Support is however found for the hypothesis that the field exercise is more predictive of performance in the practical aspects of education and training than in theoretical aspects *(H9)*, although the difference in marginal (β = .275** versus β = .229**).

Once again, a separate analysis was performed to assess the predictive validity of *7. Academic Prognosis* and *8. Leader Prognosis* for *11. GPA Theoretical Subjects*. Table 5.8 presents the results of this analysis.

Table 5.8

Regression analysis of variables 7. Academic Prognosis and 8. Leader Prognosis with criterion 11. GPA Theoretical Subjects

7.Academic Prognosis	.597**
8. Leader Prognosis	.174*
R2	.401
Adjusted R2	.388
F-value	29.171**

Note: N=90. Standardized Beta coefficients are reported. *p<.05, **p<.01.

Both *7. Academic Prognosis* (β = .597**) and *8. Leader Prognosis* (β = .174*) are significant variables, and so is the model as a whole. With an Adjusted R Square of .388, the two prognoses as a model are able to explain 38,8% of the variance in *11. GPA Theoretical Subjects*, which is only a slight improvement from the model containing the selection variables as predictors.

As the next step of analysis, the predictor variables were once again entered into the same hierarchical regression analysis to assess their predictive validity for the leadership criterion. Table 5.9 presents the results for the analysis with *12. Balanced Leadership Behaviour* as the criterion.

Table 5.9

Multiple hierarchical regression analysis with criterion 12. Balanced Leadership Behaviour as dependent variable

	Model 1	Model 2	Model 3	Model 4
1.High School GPA	.021	.022	-.009	.013
2.GMA	-.078	-.080	-.069	-.070
3.Physical Strength		.021	.005	-.018
4.Physical Endurance		-.056	-.052	-.068
5.Interview			.126	.107
6.Field Exercise				.164*
R2	.005	.008	.023	.048
Adjusted R2	-.003	-.008	.003	.025
R2 Change	.005	.002	.015	.025
F-value	.676	.492	1.167	2.080

Note. N=256. Standardized Beta coefficients are reported. *p<.05, **p<.01.

Neither *1. High School GPA, 2. GMA, 3. Physical Strength, 4. Physical Endurance* or *5. Interview* are significant variables in the first three models, nor are

these models significant. *6. Field Exercise* is the only significant variable of the fourth model (β = .164*), but the model as a whole is not significant. Adding *6. Field Exercise* in the fourth model increased Adjusted R Square to .025, meaning that the model representing the battery of selection methods as a whole is only able to explain 2,5% of the variation in the criterion balanced leadership behaviour.

Through these findings, we find no support for our hypotheses that high school GPA, GMA and the interview are significant predictors of military leadership performance (*H2*, *H4*, *H6*). However, these findings do lend support to the hypothesis that the field exercise is a significant predictor of supervisory ratings of military leadership performance *(H8).*

A separate analysis was performed to assess the predictive validity of *7. Academic Prognosis* and *8. Leader Prognosis* for criterion *12. Balanced Leadership Behaviour*. Table 5.10 presents the results of this analysis.

Table 5.10

Regression analysis of variables 7. Academic Prognosis and 8. Leader Prognosis with criterion 12. Balanced Leadership Behaviour

7.Academic Prognosis	-.103
8. Leader Prognosis	.245*
R2	.067
Adjusted R2	.049
F-value	3.636*

Note: N=104. Standardized Beta coefficients are reported. *p<.05, **p<.01.

8. Leader Prognosis is the only significant variable in this model (β = .245*). The model as a whole is however significant. Adjusted R Square of this model is .049, meaning that the academic and leader prognoses as a model can account for almost double the amount (4,9%) of the variation in balanced leadership behaviour scores than can the model including the battery of selection tools.

The final analysis to be performed was with the aim of assessing to what extent academic performance in education and training predicts balanced leadership behaviour when controlling for the selection system. *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects* were entered as predictors into the hierarchical regression. Table 5.11 presents the results of the analysis including *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects* as predictors of the criterion of *12. Balanced Leadership Behaviour*.

Table 5.11

Multiple hierarchical regression analysis with criterion 12. Balanced Leadership Behaviour as dependent variable

1.High School GPA	.027
2.GMA	-.013
3.Physical Strength	-.057
4.Physical Endurance	-.202*
5.Interview	.103
6.Field Exercise	.115
10.GPA Practical Subjects	.387**
11.GPA Theoretical Subjects	−.218 †
R2	.197
Adjusted R2	.141
F-value	3.496**

Note: N=123. Standardized Beta coefficients are reported. *p<.05, **p<.01.

The results of this analysis indicate that, when controlling for the various selection predictors, *10. GPA Practical Subjects* is a significant predictor of *12. Balanced Leadership Behaviour* (β = .387**). (*11. GPA Theoretical Subjects* is nearly significant (β = -.218) at the .05 level p=.051). While *6. Field Exercise* was a significant predictor of BLB evaluations (β = .164*) in previous analyses (Table 5.9), it is no longer a significant predictor controlling for the effect of education and training. (Note however that the number of observations included in this analysis is only 125, compared to 289 in previous analyses). *4. Physical Endurance* is the only significant selection predictor (β = -.202*) when controlling for the effect of education and training, with a negative effect on BLB evaluations. Accounting for variables *10. GPA Practical Subjects* and *11. GPA Theoretical Subjects* results in an Adjusted R Square of .141. Thus, when accounting for both the battery of selection tools and performance during education and training, 14,1% of the variance in *12. Balanced Leadership Behaviour* can be explained. These findings partially support the hypothesis that higher levels of performance in military leadership education and training is facilitative of higher levels of military leadership performance *(H1).*

6. Discussion

In this section, we discuss the study's results and findings against theories and findings from the literature review. The findings related to predicting performance in education and training are discussed first, followed by a discussion of the findings related to predicting balanced leadership behaviour. The section

culminates with a general discussion of the relationship between selection, education and training and subsequent military leadership performance.

6.1 Predicting Performance in Education and Training

The NAFs' military branches' need for competence constitutes the basis for initiating the selection process for education in the Armed Forces (Forsvarets personell- og vernepliktssenter, 2019a). This adds to the requirements of the selection process, necessitating its ability to predict the likelihood of performance in and successful completion of education and training.

6.1.1 High School GPA as a Predictor of Academic Performance

By assessing the predictive validity of high school GPA as a predictor for academic performance, we found support our hypothesis that high school GPA is a stronger predictor of academic performance in theoretical subjects than performance in practical subjects during military education and training (*H3*).

We initially found that high school GPA was the best single predictor for academic performance in military training, and that the issue of simplex validity patterns is in fact not of great concern in this context (Lievens, Ones & Dilchert, 2009; Roth et al., 1996). However, we were somewhat surprised when controlling for the effect of the other predictors. While high school GPA is still a significant predictor, the field exercise is an even more significant predictor of overall performance in education and training. This finding exemplifies how the total GPA may be a somewhat imprecise measure of performance in quite different areas, requiring different skills and abilities. Distinguishing between practical and theoretical subjects reveals that field exercise scores are more strongly related to performance in the education's practical aspects, while high school GPA is more strongly related to performance in theoretical aspects.

These findings were further strengthened by regression analysis with GPA in theoretical subjects as the criterion. Here, we find that not only is high school GPA a significant variable controlling for other predictors, it is also the predictor with the strongest positive effect on academic results in theoretical subjects throughout the four models tested. Ultimately, these findings indicate that high school GPA might not be the best single predictor for adapting to military training *as a whole*, as previous US studies have indicated (Farr & Tippins, 2010). However, our findings suggest that high school GPA is a valid predictor of officer training

and academic performance in military education (Alf et al., 1988; Kjenstadbakk, 2012; Norrøne, 2016; Vik, 2012) and that it does capture outcomes of learning in academic environments (Viswesvaran & Ones, 2018).

In terms of individual characteristics, grades are primarily determined by cognitive ability (Kunzel & Hezlett, 2007), and significant correlations between school performance and intelligence have been found (Roth et al., 2015). Significant correlations were also found among our predictor variables (Table 5.2), where high school GPA correlates substantially with GMA, indicating that measures of cognitive ability and grades are related constructs. This implies that both high school GPA and GMA are important selection tools for assessing candidates' presumed potential for succeeding academically during education.

6.1.2 GMA as a Predictor of Academic Performance

While we find high school GPA to be a significant predictor of total OCS GPA, we find it somewhat counterintuitive that no support is found for our hypothesis that GMA is a significant predictor of performance in military education and training (*H5*), and that it has a negative effect on total GPA. Similar results are found for the criterion GPA in practical subjects. Thus, it is further emphasized that using traditional predictors for academic performance is not expedient for predicting performance in the OCSs' practical aspects. The main plausible reason for this is that different skills and abilities might be required to perform well in these aspects of education and training, as opposed to traditional academic environments (Viswesvaran & Ones, 2018) where performance is primarily determined by cognitive ability (Kunzel & Hezlett, 2007). This is supported by the fact that we found that GMA is a significant predictor of GPA in theoretical subjects, lending some support to our hypothesis *(H5).* Consequently, the NAFs' GMA measures of verbal, numerical and general reasoning might be a better suited predictor for performance in theoretical subjects than in practical subjects. However, high school GPA is the single best predictor of theoretical academic performance, indicating that previous academic results are a stronger predictor than the ability to process, understand, reason with and remember information. In such, measures of GMA cannot rule out the need for more specific predictors (Farr & Tippins, 2010) of performance in a context that is specific for military leadership.

The *true* validity of GMA as a predictor of performance in military training and education might not the captured in the Joint Admission and Selection process,

as performance measures can only be obtained for individuals who pass through a selection process and are hired. The minimum requirement is GMA >/= 5, equalling an IQ of 100 (Forsvarets Sanitet, 2018), meaning that about 40% of the population are opted out. Hence, the majority of possible performance measures are never documented, in that the largest part of the method's explanatory power lies within the 40% of the population which is cut off. This range restriction (Sackett & Yang, 2000) is likely to result in underestimated validity coefficients (Arnold et al., 2016).

The academic prognosis, a weighted and combined measure of high school GPA and GMA, is used as a basis for final selection decisions. This prognosis is of practical utility to the extent that it is a better predictor of academic achievement than are its individual components. Our results reveal that the academic prognosis significantly predicts performance in the education's theoretical aspects, but not in practical aspects. It is also a significant predictor of overall performance in education and training, with the model containing both the academic and leader prognosis being able to account for more than double the amount of the variation in total GPA than the model including the entire battery of selection tools. This indicates that the effect of underestimated validity coefficients is reduced and that the contribution of the broad performance measure is higher for overall performance. Hence, these findings indicate that the weighted combination of high school GPA and GMA scores provides a better measure of academic performance than the two individual measures, and support the use of the academic prognosis as collective predictor set, as a basis for final selection decisions.

6.1.3 Interview as a Predictor of Academic Performance

The hypothesis that the interview is more predictive of performance in the practical aspects of education and training than in theoretical aspects (*H7*) is not supported by this study's findings. Quite contrary, we find that the interview is significantly predictive of performance in theoretical aspects, but not practical aspects of education and training. This finding is somewhat surprising, given the interview's emphasis on leadership behaviours, which intuitively should be more strongly related to behavioural patterns in practice than skills and abilities required for performance in traditional academic environments.

Correlational analyses (Table 5.2) reveal a slight significant correlation between high school GPA and interview scores (r=.183**), indicating that the interview captures some construct also captured by high school GPA. Meta-analytic

findings suggest that interviews to some extent measure mental ability (Huffcutt et al., 1996), which has also been found to be a determinant of grades (Kuncel & Hezlett, 2007). No such correlation between GMA measures and interview scores was found in this study, suggesting that the interview captures some other aspect related to high school academic performance. Our analyses reveal that high school GPA significantly correlates with OCS GPA in theoretical subjects (r=.486**), while only marginally with total GPA (r=.195*) and not at all with GPA in practical subjects. It thus seems that there is some shared construct captured by both high school GPA and the interview that is related to academic performance in theoretical subjects, but not performance in practical subjects. Although intended to measure leadership potential, the interview may also capture constructs such as previous experience, personality traits and behaviour patterns – exactly what combination of constructs is measured by the interview remains unknown (Schmidt et al., 2016).

The effect of pre-interview impressions may be particularly relevant here, as part of the preparation for the interview includes interviewers reviewing candidates' application papers and high school diploma. Meta-analytic findings suggest that interviewers' access to cognitive ability test scores decreases the interviews predictive validity (McDaniel et al., 1994). The correlation between high school GPA and interview scores and the finding that the interview is a better predictor of GPA in theoretical than practical subjects could suggest that access to measures of previous performance in the form of grades might have a similar effect.

6.1.4 Field Exercise as a Predictor of Academic Performance

The hypothesis that the field exercise is more predictive of performance in the practical aspects of education and training than in theoretical aspects (*H9*) is supported by the findings of this study, albeit only marginally. While the OCS education and training contains theoretical subjects for which skills and abilities required for achievement in traditional academic environments undoubtedly are important, it also contains training in the practical aspects of military leadership. In such, the field exercise which tests candidates in practical aspects of military leadership in military-specific contexts provides added value to the selection process in predicting performance in the education and training's practical subjects.

The finding that the field exercise is also a significant predictor of academic performance in the theoretical aspects of the education, even when controlling for predictors such as high school GPA and GMA, is somewhat surprising. This is

particularly interesting given the slight negative correlation between field exercise scores and high school GPA (*r*=-.147*). A potential explanation for this could be that there is some motivational factor required to overcome the challenging and demanding nature of the field exercise that also contributes to endurance and performance in the education's academic challenges.

6.1.5 The Predictive Validity of the Entire Selection System

The OCS total GPA represents performance during military leadership education and training, and reflects acquisition of job-related knowledge, a major determinant of subsequent job performance (Schmidt & Hunter, 1992; Schmidt et al., 2016). Given that candidates are educated and trained in both academic and practical aspects of military leadership, overall performance is expected to require several different skills, abilities and characteristics. Hence, accurate prediction of overall performance may require the combination of several measurement methods.

Our results indicate that the selection system as a whole is able to explain 8,4% of the variation in the total GPA criterion. Of all selection tools included in the selection system, only high school GPA and field exercise scores are predictive of total GPA. Thus, while the selection system to some extent predicts performance in education and training as measured by total GPA, a vast proportion of variation in total GPA remains unaccounted for by the system of selection methods. As pointed out, because performance in theoretical and practical subjects are likely to require quite different skills and abilities, the total GPA may be a somewhat imprecise criterion. In such, differentiating between GPA in practical and theoretical subjects may better capture the relationships between the various predictors and aspects of performance in education and training.

The entire selection system as a model is able to account for 7,3% of the variation in GPA in practical subjects. Of all the selection methods, the field exercise is the only significant predictor of performance in the education's practical aspects. Together, these findings imply that the system as a whole is not particularly predictive of performance in the practical aspects of the education and training, though the prognoses used for final selection decisions are somewhat valid predictors. Still, much of the variation in this aspect of performance is attributed to factors other than those assessed during the selection process.

However, as a whole, the selection system appears to be fairly predictive of performance in the theoretical aspects of the OCS education. The entire selection

system as a model is able to account for a substantial 37,1% of the variation in the criterion GPA in theoretical subjects. While high school GPA and GMA, included in the selection system to ensure academic performance, alone can account for 24,6% of the variation in GPA in theoretical subjects, the inclusion of the remaining selection predictors results in considerable incremental validity.

Out of these three performance criteria, the selection system is best able to predict academic performance in the OCSs' theoretical subjects. In such, the selection system's success in predicting performance in education and training depends on how this is conceptualized. Overall, the system is partly successful in predicting the likelihood of performance in and successful completion of education and training, containing predictors important for ensuring performance in both theoretical and practical aspects of military education and training.

6.2 Predicting Balanced Leadership Behaviour

The NAF does not primarily select candidates for military education, but for the profession (Forsvarets personell- og vernepliktssenter, 2019b), which necessitates the identification of candidates with the greatest potential for executing military leadership during selection. Based on the requirements for effective leadership within the NAF, substantial resources are placed into assessing candidates' leadership development potential.

6.2.1 High School GPA as a Predictor of Balanced Leadership Behaviour

Our findings do not support the hypothesis that high school GPA is a significant predictor of military leadership performance (*H2*). This suggests that behaviours resulting in successful outcomes of learning in academic environments (Viswesvaran & Ones, 2018) are not the same as those required for effective military leadership. This finding is somewhat surprising, as grades are primarily determined by individual characteristics such as cognitive ability (Kuncel & Hezlett, 2007) and to some extent personality factors (Connelly & Ones, 2010), which are also thought to be related to leadership (Schmidt & Hunter, 1998; Judge et al., 2002). However, meta-analytic findings suggest that GPA is more predictive of performance in training than job performance (O'Leary, 1980, as cited in Hunter & Hunter), because of the strong resemblance between training programs and classroom demands (Schmidt et al., 2016). Our findings reflect this notion, with high school GPA having a positive significant effect on performance in military

education and training. Thus, while not a directly relevant predictor of leadership performance, it may provide utility in predicting performance in education and training, which in turn is likely to be related to subsequent leadership performance.

6.2.2 GMA as a Predictor of Balanced Leadership Behaviour

The results of this study do not support the hypothesis that GMA is a significant predictor of military leadership performance (*H4*). This is a highly surprising finding, given that measures of GMA are found to be among the most valid predictors of job performance across all job situations (Schmidt and Hunter, 1998). Furthermore, meta-analytic findings suggest that the relationship between intelligence and leadership is stronger than between intelligence and non-leaders (Schmidt & Hunter, 1998), which may be explained by the fact that leaders are required to solve problems poorly defined problems (Schmidt & Hunter, 1992).

Military leaders are faced with complex tasks, at times in unpredictable situations (Lang-Ree & Martinussen, 2019) in a context at times characterized as dangerous and extreme, which is considered the core condition for performing military leadership (Hannah et al., 2013). In such, one would expect intelligence to be important in executing military leadership. However, this description is not necessarily particularly fitting of the context within which OCS candidates practice military leadership during the education's second year of practice.

Balanced leadership behaviour evaluations indicate the extent to which candidates demonstrate 1) task-oriented behaviours such as goal setting, role clarification, monitoring progress towards goals and action-orientation (*mission focus*), 2) development-oriented behaviours aimed at developing the competence of subordinates to support their goal-attainment, such as supporting autonomy, promoting novelty-seeking, supporting mastery and encouraging subordinate personal development (*development focus*), and 3) *role model* behaviours such as organization-orientation (versus self-centered behaviours), willingness for personal development, consideration and integrity.

Seen in relation to Yukl's (2013) contingency theory of leadership skills, role model and development-oriented behaviours, both having relational components (Martinsen et al., unpublished), may require what Yukl (2013) defines as interpersonal skills, as these behaviours require knowledge about human behaviour and interpersonal processes. Mission focus on the other hand may to a greater extent require more technical and conceptual skills. However, this study's

subjects receive BLB evaluations after 6- or 12-month period of lower-level leadership practice. Given that the importance of conceptual skills, reflecting GMA, increases in line with ascendance to higher organizational levels (Yukl, 2013), mission focus scores may be less affected by GMA, rather reflecting technical aspects. The "missions" or tasks that candidates are met with during the OCS practice year might not be that complex or draw on cognitive resources requiring cognitive information processing, which has shown to have a large correlation with GMA over time (Ackerman, 1986, 1987, 1992; Bertua, Anderson & Salgado, 2011; Farr & Tippins, 2010; Salgado et al., 2003; Schmidt, Hunter & Outerbridge, 1986).

Given the increasing degree of specialization in the NAF (Forsvarets personell- og vernepliktssenter, 2019b), one can assume that requirements for the leader's characteristics will vary widely in relation to different disciplines. Thus, it seems plausible that a minimum GMA requirement is important, so that leaders who are selected have the prerequisites for being promoted within the organization (Schmidt & Hunter, 2004; Wilk, Desmarais and Sacket, 1995). However, the importance of intelligence might be less pronounced at lower-rank levels within the Armed Forces' organizational hierarchy, but increase along with rank ascendance as one is exposed to increasing demands of conceptual skills requiring higher levels of intelligence (Yukl, 2013). In such, the requirements for intelligence may change in parallel with level of leadership and thus not be expressed during the OCS year of practice, but rather after a few years of service as an officer.

With this assumption, ability measurement will probably be an important predictor for selection for the OCS, even if this study's analyses do not find as strong correlations between GMA and the criterion for leadership performance as expected. With behaviours being to some extent determined by personality (Cooper, 2010), evaluations of leadership behaviours at lower-rank levels may be less influenced by GMA, and more so by other individual characteristics more directly related to observable behaviours. However, being perceived by others as a leader does not guarantee long term leadership effectiveness in terms of positive organizational, financial and relational effects. In such, the true validity of GMA as a predictor of leadership effectiveness may not be captured in this study.

The lack of predictive validity of GMA for BLB evaluations may also be attributed to issues of range restriction as previously described, where performance measures for 40% of the population are lacking, leading to range restriction (Sackett & Yang, 2000) and ultimately underestimated validity coefficients (Arnold et al.,

2016). However, this finding can probably be partly explained also by the fact that intelligence consists of several facets, and that the «g-factor» may not provide optimal predictive validity in relation to, for example, leadership (Skorstad, 2015). Hence, this finding supports theories that leadership is situation-dependent and adapted to the situation in which leadership is exercised, and thus different abilities are required for different leadership situations (Haslam, Reicher & Platow, 2011).

6.2.3 Interview as a Predictor of Balanced Leadership Behaviour

Given the interview's behavioural emphasis, aiming to measure intentions of behaviours vital for effective military leadership, we hypothesized that the interview is a significant predictor of military leadership performance (*H6*). However, this study's findings lend no support to this hypothesis. The interview's lack of ability to predict subsequent leadership behaviours may suggest poor construct validity, meaning the interview may not measure what it intends to (Cooper, 2010). Several sources of error may obscure the measurement of leadership potential in the interview. Although the interview is structured, interviewers are given leeway to improvise follow-up questions, which may reduce standardization across candidates and ultimately validity (Posthuma et al., 2002).

Furthermore, although interviewers undergo interview training and follow an interview guide and template for scoring responses, the interview's social context may render interviewer evaluations vulnerable to various biases, such as the "halo effect" (Skorstad, 2015), the "similar-to-me-effect" (interviewer-applicant similarity), and the effect of first impressions and pre-interview impressions (Posthuma et al., 2002; Skorstad, 2015). As discussed, the effect of pre-interview impressions may be particularly relevant here. Our findings may suggest that access to measures of previous performance in the form of grades might distort the measurement of potential for leadership development.

Alternatively, candidate attempts at impression management or poor self-knowledge may reduce the accuracy of responses and ultimately evaluations (Cooper, 2010). Research suggests that the relationship between verbal impression management and interview ratings is lower for high-structure than low structure interviews (Barrick et al., 2009). Yet, a discrepancy is observed between self reports of and actual mastery of essential tasks (Forsvarets personell- og vernepliktssenter, 2019a), which may indicate an effect of social desirability or poor self-knowledge on responses to interview questions. Alternatively, as personality traits, abilities,

motivation and knowledge, in addition to facilitative situational factors are important for behaviours essential for reaching desired outcomes (Kurz & Bartram, 2002), behavioural intentions are likely insufficient for predicting future behaviour.

While both situational and behavioural interviews are predictive of job performance (Klehe & Latham, 2006), a core issue with situational interviewing is that it does not account for different levels of experience (Arnold et al., 2016). Some applicants who apply for the OCS have some previous military experience from compulsory military service, while others have no previous military experience at all. Candidates familiar with the NAF and military context may therefore have an advantage over inexperienced candidates by having a better understanding of what is required and desired in described settings (Arnold et al., 2016). If so, the interview could be measuring levels knowledge of and familiarity with the military context instead of military leadership potential. However, precisely because many candidates do not have any previous work experience, situational interviewing avoids discriminating against candidates with no previous work-related behaviours to draw on as required in behavioural interviewing (McDaniel et al., 1994). While the behavioural interview may be slightly favourable for predicting job performance in complex jobs (Krajewski et al., 2006; Levashina et al., 2014), the situational interview may be fairer and more effective for entry-level positions.

It is worth noting that in contrast to other selection tools, the interview serves several other purposes beyond measurement of personal characteristics, i.e. evaluating the extent of fit between the candidate and the organization (Skorstad, 2015). Given its social nature, the interview may capture important elements otherwise hard to uncover (Cooper, 2010), such as divergent attitudes and perceptions or psychological vulnerabilities. The flexibility of follow-up questions may hence be highly valuable in order to detect such critical factors. Ideally, the interview should produce valid measures of leadership potential predictive of subsequent leadership evaluations, yet some of its value might lie in identifying individuals not suited for the military profession in terms of attitudes or psychological vulnerabilities. Given the abovementioned challenges related to self-reports of behaviour in hypothetical situations, leadership potential might be more accurately measured through direct observation.

6.2.4 Field Exercise as a Predictor of Balanced Leadership Behaviour

The results of our analyses support the hypothesis that the field exercise is a significant predictor of military leadership performance (*H8*). During the field exercise, behaviour in military-specific contexts is observed over an extended period of time, allowing for assessment of performance in a variety of relevant situations across different exercises (Arnold et al., 2016). Particularly, candidates are evaluated on the extent to which they exhibit *role model*, *solving mission*, *mental robustness*, *interaction* and *development* behaviours. This resembles the measurement of BLB, where supervisors assess the extent to which subordinates over an extended period of time have demonstrated various mission focused-, development focused-, and role model-oriented behaviours. Assessment based on direct observation of job-relevant behaviours (Arnold et al., 2016) thus appears to result in fairly good criterion-related validity (Hough & Ones, 2001).

Our results reveal that the field exercise is the only significant predictor of BLB, suggesting that important characteristics such as leadership quality or social skills are difficult to measure through other tools than behavioural ratings (Cooper, 2010). The field exercise thus appears to provide insight into relevant dimensions that prove hard to evaluate through other selection methods (Borman, 1982; Meriac et al., 2008). Given the distinctiveness of the military context (Hannah & Snowden, 2013), it seems that observing behaviours in context-specific and critical situations is essential for predicting future military leadership behaviours.

While we find the field exercise to be a significant predictor of BLB, a stronger effect was expected given the similarity of constructs measured in the field exercise and BLB evaluations. Several sources of error may potentially obscure filed exercise measures, resulting in an apparently weaker relationship between leadership potential estimates and subsequent leadership behaviour. A common source of error in AC behavioural ratings concerns assessors being more likely to provide similar ratings across different dimensions within the same exercise, rather than for the same competency across exercises (Robertson & Smith, 2001). Concerns may thus be raised to whether the field exercise actually measures the five leadership dimensions intended. Yet, as scores for each exercise are only a step towards an overall final rating of each leadership dimension, exercise-specific effects may not be the dominant source of final ratings when ratings for each leadership dimension across all exercises are combined (Kuncel & Sackett, 2014).

Despite the use of systemic procedures for rating specific behaviours as they occur, assessors' subjectivity may still influence behavioural ratings through the

"halo effect", so that candidates scoring high on one particular leadership dimension also receive high scores on the other dimensions. Another potential source of error concerns the distinction between leadership and leadership emergence (Lord, De Vader & Alliger, 1986). Implicit shared beliefs about leader behaviours and traits could potentially favour candidates who fit a certain military leader prototype. Such implicit beliefs might affect assessors' perceptions directly or, alternatively, have an indirect effect on scores through followers' reactions. Followers tend to allow others to lead when they match the followers' perceptions of what good leaders should be (Lord et al., 1986). Thus, when holding the leader role, candidates who fit followers' leader perceptions may elicit higher levels of followership, which in turn could influence assessors' ratings.

Our results indicate a positive relationship between field exercise scores and physical strength (r=.229**) and physical endurance (r=.186**). Physical abilities undoubtedly facilitate mastery of the field exercise's physically challenging aspects, yet there is no intuitive reason for these abilities being related to for example the demonstration of decision-making skills (*solving mission*), the ability to care for others (*interaction*) or the ability to stimulate independent thinking in followers (*development*). These correlations could be interpreted as physical abilities affecting perceptions of leadership qualities, resulting in more physically able candidates receiving higher scores. As being perceived as a leader does not guarantee leadership effectiveness (Judge et al., 2002), such an effect might explain the low correlation between field exercise scores and subsequent BLB evaluations.

Evidently, there are challenges inherent to both the interview and the field exercise as measurement tools for assessing leadership development potential. In such, the use of two methods for assessing leadership potential is recommended, as measuring constructs through several instruments tends to cancel out the systematic errors associated with each instrument, resulting in more accurate measures (Cooper, 2010). We find that the leader prognosis, derived from interview and field exercise scores and used as a basis for selection decisions, is a better predictor of BLB than both interview and field exercise scores individually. This suggests that the weighted combination of interview scores and field exercise scores providing a better measure of leadership potential than the two individual measures, which supports the use of the leader prognosis as a basis for final selection decisions.

6.2.5 The Predictive Validity of the Entire Selection System

Because performance in any given role is likely to require several competencies with potentially differing underlying individual characteristics, and because no selection method alone seems to be perfect (Schmidt & Hunter, 1998), combining various selection methods into a system tends to increase predictive accuracy (Lai, 2010). The Joint Admission and Selection system is assembled with the intention to identify the candidates with the greatest potential for executing military leadership. Thus, of major interest is the extent to which the selection system as a whole is predictive of performance in the role as a military leader, and the incremental validity of each added predictor (Arnold et al., 2016).

Our results indicate that the selection system as a whole is able to explain a mere 2,5% of the variation in balanced leadership behaviour. Of all selection tools included in the selection system, the field exercise is the only significant predictor of subsequent balanced leadership behaviour. These findings suggest that the selection system as a whole rather poorly predicts future military leadership performance as measured by balanced leadership behaviour, leaving 97,5% of the variance in BLB evaluations unaccounted for. The prognoses based on which final selection decisions are made are slightly more valid predictors, yet also these leave a vast proportion of variation in balanced leadership behaviour evaluations unaccounted for (95,1%). Because it is very difficult to control extraneous factors when dealing with humans (Jank, 2011), all of the variance in criterion variables is rarely accounted for in the social sciences. Yet, the finding that nearly all variance in balanced leadership behaviour is explained by other factors than those included in the battery of selection tools assembled with the purpose of identifying and predicting leadership potential warrants a critical stance.

The preceding discussion of individual predictors has illuminated potential errors of measurement associated with each selection method that may decrease the reliability and, importantly, predictive validity of these measurement tools. However, reliability and validity in the measurement of criteria is equally important in validation processes (Arnold et al., 2016; Lang-Ree & Martinussen, 2019). An apparent lack of predictive validity between the system of selection tools and balanced leadership behaviour might just as well be ascribed to erroneous criterion measurement as predictor measurement.

Studies conducted in the process of validating the BLB tool have found satisfactory reliability measures (Martinsen et al., unpublished), suggesting that the tool generates accurate evaluations (Cooper, 2010). In addition to being reliably

measured, a good criterion must also contain variability in performance across individuals (Lang-Ree & Martinussen, 2019; Schmidt & Hunter, 1998), and performance scores should ideally be distributed across the entire measurement scale (Lang-Ree & Martinussen, 2019). Descriptive statistics (Table 5.1) show that BLB scores are not distributed across the entire scale of measurement, with a range of 4,33 and mean score of 5,2 on a 7-point scale. These statistics could suggest that supervisors are reluctant of giving low scores, potentially disturbing the true relationship between predictors and balanced leadership behaviour.

On the other hand, as candidates who receive BLB evaluations have been selected on multiple criteria and undergone two years of military education, training and practice, one might expect such a high average score. Research indeed suggests that performance variability would be larger if all candidates were hired or selected randomly (Schmidt et al., 1979). Furthermore, the possibility that the purpose of measurement affects evaluations cannot be entirely ruled out. It is possible that supervisors would assess subordinates' balanced leadership behaviour differently if evaluations were collected for practical use rather than research purposes.

It is worth noting that the entire selection system contains selection tools included, not because their measures are hypothesized to be predictive of leadership-related criteria, but for the sake of ensuring ability to master academic and physical aspects of education and training. While GMA and GPA are likely to be relevant predictors for academic performance, personality traits such as extraversion and conscientiousness may be more relevant for predicting leadership-related criteria (Lang-Ree & Martinussen, 2019). Meta-analytic findings indeed suggest there is a relationship between aspects of personality and leadership (Judge et al., 2002). One might therefore assume that a greater proportion of variance in BLB could be accounted for, were personality measures included in the selection system. In such, it is perhaps less surprising that the selection system does not account for a substantial proportion of variance in balanced leadership behaviour, as it both contains certain predictors not necessarily relevant to leadership criteria, and lacks personality measures which likely are relevant.

Although "*the Chief of Defence's perspective on leadership*" holds that effective military leadership requires balancing behaviours oriented towards mission, interaction, development and being a role model (Forsvarsstaben, 2012), contingency theories of effective leadership posit that the traits or behaviours required for effective leadership vary for different situations (Yukl, 2013). It is

plausible that the context(s) under which candidates execute leadership during the education's second year of practice call for one of these behaviour orientations to a greater extent than others. Hence, the use of the aggregated BLB criterion could potentially conceal relationships between selection predictors and sub-components of balanced leadership behaviour; mission focus reflecting task-oriented behaviours, development focus reflecting relations- and change-oriented behaviours, and role model-oriented behaviours, respectively.

Although there may be individual differences facilitative of performance in a leader role, military leadership requires mastery of specific military competencies achieved through education and practice in the military profession (Grebstad & Johnsen, 2019). In such, a more likely relationship between selection predictors and subsequent military leadership may be one where the selection system is predictive of potential for prospering during military leadership education and training, and the competencies achieved through the OCS in turn are causal of subsequent military leadership performance.

6.3 Performance in Education and Training as a Predictor of Balanced Leadership Behaviour

The leadership development potential sought during selection indicates how candidates, given education and training, are expected to perform as a leader at a lower-rank level. Assuming that specific military competencies acquired in education and practice cultivate leadership potential and contribute to military leadership performance, a final analysis was performed to assess the extent to which performance in military leadership education and training predicts subsequent military leadership performance, controlling for the effect of selection. Results of this analyses suggest that performance in OCS practical subjects is a highly significant predictor of BLB evaluations, while performance in theoretical subjects appears to have a negative, though not significant effect on BLB evaluations. Thus, we partially find support for the hypothesized relationship between performance in education and training and subsequent military leadership performance (*H1*).

These findings suggest that it is not academic performance in theoretical subjects, but rather performance in practical subjects that is predictive of subsequent BLB. In such, it is a paradox that the selection process is most successful in predicting theoretical academic performance, when this in turn does not predict, or is even negatively related to subsequent military leadership

performance. Furthermore, even after controlling for the effect of military leadership education and training, the vast majority of variance in BLB evaluations (85,9%) is still unaccounted for.

Identifying valid performance predictors for personnel selection has always been an important challenge for military organizations (Fosse et al., 2015), and a highly valid question is whether there exist robust generic selection criteria that embrace the range of positions across organizational levels and branches that candidates may come to hold. With the military profession becoming ever more differentiated and specialized, it is a paradox that the selection process for lower-level leader positions is generic in nature. Because candidates are not selected for a specific job, but rather a general role as lower-rank leader, it is difficult to conduct specific job analyses to base the selection process on. Instead, candidates are assessed against a competency profile consisting of leader competencies derived from "*the Chief of Defence's perspective on leadership*".

A challenge with generic competency frameworks is that they often lack sufficient behavioural specificity for many job roles (Arnold et al., 2016). Ideally, behavioural indicators in a competency model should be directly relevant to a specific job role and its context, so that assessment criteria may be more accurately defined and their use optimized in selection (Arnold et al., 2016). In such, a core challenge with the Joint Admission and Selection system may be that it is based on a generic competency profile. Furthermore, because the OCS has traditionally been a qualifying requirement for pursuing higher levels of military education, in turn qualifying for rank ascendance and the pursuit of a life-long career, it has been necessary for its selection process to ensure potential for mastering even more academically demanding environments.

As the NAFs' traditional vertical career system has been found generate an inexpedient rank structure, resulting in a competency structure that is not adequately adjusted the NAFs' need for in-depth competencies, a new rank-structure has recently been implemented to ensure fewer generalists and more specialists (St.prp. 111 LS, 2014-2015). This implementation reflects the military profession becoming ever more differentiated and specialized, and has required a restructuring of the NAFs' education system as well. The new education system is customized to educate for two different career paths, which in turn has implications for these educations' selection models. The selection for the new officer education must to a greater extent be based on the fact that officer candidates embark an

education where the requirements for analytical problem understanding, strategic insight and overall understanding are emphasized more than previously (Forsvarets personell- og vernepliktssenter, 2019a). The new specialist education on the other hand, being more experience-based, revolves around more fundamental leadership and discipline understanding (Forsvarets personell- og vernepliktssenter, 2019a), and will thus require a different set of selection criteria.

The findings of this study, validating the traditional selection system for admission to the former OCS, seem supportive of such a restructuring of selection for military education. While grades in theoretical subjects reflect academic performance and capture outcomes of learning in *academic environments* (Viswesvaran & Ones, 2018), performance in practical aspects of military leadership education and training is what appears to cultivate leadership potential and facilitate acquirement of specific military competencies required for mastery of military leadership for lower-rank leaders. Performance in theoretical subjects might be more important to provide a basis for potential further military education (BAE or AO) required for rank ascendance within the organizational hierarchy and to acquire competence to conduct planning, and management of operations required when pursuing a life-long career as an officer (Forsvarets høyskole, 2013). Thus, high school GPA, GMA and performance in theoretical subjects may be more important predictors for future officer education, while the field exercise and the traditional OCSs' practical aspects are likely more important for the future specialist education, as mastery of this aspect of education and training appears more closely related to the nature of work as a lower-rank military leader.

A distinction between two different career paths with respective educations further seems sensible in that the skills, competencies and abilities required for higher-level officer educations are not necessarily important for mastering practical aspects of military leadership at lower-rank levels. The distinction between officers and specialists narrows the range of possible positions individuals may come to fill across hierarchical levels, branches and functions, which in turn may facilitate the identification of specific skills, competencies and abilities to base selection for these different educations on. Selection for the new officer education may reflect the greater requirements for analytical problem understanding, strategic insight and overall understanding than previously. In such, selection may be based on a more tailormade competency model, so that assessment criteria may be more accurately defined and their use optimized in selection (Arnold et al., 2016).

The selection of military leaders revolves around identifying individuals with the potential to, given education and training, master the balancing of leader behaviours essential for effective military leadership. Balanced leadership behaviour is indeed regarded a skill that can be trained and developed (Johansen et al., 2019). However, a competency potential, formed by individual characteristics such as personality traits, abilities, motivation and knowledge, only translates into observable behaviours essential for reaching desired outcomes given facilitative situational factors (Kurz & Bartram, 2002). Our findings may indicate that situational factors during the period for which BLB is assessed are not facilitative of the potential assessed during selection translating into desirable behaviours.

However, the lack of predictive validity of the selection system in predicting balanced leadership behaviour in aspiring officer candidates does not necessarily mean that the selection system is not successful in identifying potential future leaders. Despite being extensively researched, the nature of leadership still remains a debated topic (Yukl, 2013), and there is no definite answer to what effective military leadership should entail (Martinsen et al., 2019). In such, the predictive validity of the selection system may be contingent on the conceptualization of leadership. Leadership behaviours relate differently to various indicators of leadership effectiveness such as subordinate satisfaction, subordinate performance and superiors' ratings of leader effectiveness. While BLB is found to predict supervisory ratings of job performance (Martinsen et al., unpublished), one cannot rule out that other factors also contribute to effective military leadership.

Finally, although BLB is regarded a skill that can be trained and developed, it is unlikely that this skill is fully developed after two years of lower-level military education and training. Thus, the true contribution of the individual characteristics measured during selection in predicting behaviours essential for leadership effectiveness may become more evident after sustained practice in the military profession, with exposure to a greater variety of tasks, challenges and demands, requiring a range of skills and behaviours.

7. Conclusion and Implications

The Joint Admission and Selection system's success hinges on its ability to predict both performance in education and training, and subsequent performance in the role as a military leader. The selection system has limited predictive validity for overall education and training performance, with only high school GPA and

field exercise scores being valid predictors. However, it is highly successful in predicting academic performance in the education's theoretical aspects. High school GPA and GMA, included in the selection system to ensure academic performance, can alone account for a substantial amount of variation in performance in theoretical aspects of education and training, yet the inclusion of the remaining selection predictors results in considerable incremental validity. The selection system is thus found to be partially successful in predicting performance in education and training.

The selection system holds limited predictive validity of performance in the role as a military leader at a lower-rank level, where the field exercise is the only of the system's six selection tools to be a valid predictor of this criterion. This suggests that important military leadership characteristics are best captured through behavioural ratings, and that the field exercise provides insight into relevant dimensions of the distinctive military context that prove hard to evaluate through other selection methods (Borman, 1982; Meriac et al., 2008). The selection interview is less able to capture such important leadership characteristics predictive of subsequent military leadership performance, yet some of its value might lie in identifying individuals not suited for the military profession in terms of attitudes or psychological vulnerabilities The finding that GMA, one of the most valid predictors of job performance across all job situations and strongly related to leadership (Schmidt & Hunter, 1998), is not predictive of performance in the role as a military leader at a lower-rank level is surprising. However, the importance of cognitive ability is likely to increase along with rank ascendance and exposure to increasing demands of conceptual skills (Yukl, 2013).

The selection system's limited validity for predicting military leadership performance is partly attributed to the fact that military leadership requires mastery of specific military competencies achieved through education and practice in the military profession (Grebstad & Johnsen, 2019). Paradoxically, it is performance in the OCS education's practical aspects, which the selection system is less successful in predicting, that contributes to subsequent military leadership performance. Different skills and abilities might be required to perform well in these aspects of education and training than in traditional academic environments (Viswesvaran & Ones, 2018). In such, traditional predictors of academic performance are inexpedient for predicting performance in practical subjects, and the need for more specific predictors of performance in a military-specific cannot be ruled out.

Being based on a generic competency framework, a core challenge with the Joint Admission and Selection system is a lack of sufficient behavioural specificity. The need to ensure that candidates admitted to lower-level education have potential for both for mastering military leadership at lower-rank levels, but also advancing in the organizational hierarchy through higher-level education may pose too challenging and perhaps even opposing requirements. A tailormade competency model based on identification of specific skills, competencies and abilities directly relevant to a specific job role and its context would facilitate accurate definition of selection criteria, optimization of their use, and ultimately greater accuracy in predicting subsequent on-the-job performance.

8. Limitations and Directions for Future Research

A main limitation of this study is the lack of corrections of range restriction, which is likely to result in underestimated validity coefficients for the various predictors (Arnold et al., 2016). Furthermore, the study's relatively small sample size warrants caution in the interpretation of results. As a larger sample size results in a smaller standard error (Stehlik-Barry & Babinec, 2017), the use of a larger sample would allow for higher confidence in the accuracy of results.

The use of only one source of leadership ratings may also be a limitation in this study, as source has been found to strongly and systematically affect leadership research results (Conway & Lance, 2010). The use of several rating sources may be needed for a valid approach to the study of leadership using inventories. Future research should include BLB-ratings provided by several sources such as subordinates and peers in addition to superordinates in order to obtain a more valid criterion measure. Relatedly, the use of BLB evaluations of leaders who have sustained practice in the military profession over a prolonged period as a criterion may give valuable insights in the selection process' predictive validity for performance in lifelong employment, which is considered a prerequisite for a sustainable selection system. Moreover, as the use of an aggregated BLB measure potentially may conceal relationships between selection predictors and sub-components of BLB, future research should also investigate the predictive validity of the selection system with each of these sub-components as criteria. Finally, future research is needed to assess whether competency models tailormade to the new military education system facilitate optimized use of selection criteria.

Literature

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10*, 101-139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. *Psychological Bullettin, 102*, 3-27.

Ackerman, P. L. (1992). Predicting individual differences in complex skill. *Journal of Applied Psychology acquisition: Dynamics of ability determinants*, 598–614.

Agresti, A., & Finlay, B. (2009). Statistical Methods for the Social Sciences (4th ed.). Upper Saddle Rives, N.J.: Pearson Prentice Hall.

Alf, E. F., Neumann, I. & Mattson, J. D. (1988). Revision of the United States Naval Academy Selection Composite. Navy Personnel Research and Development Center. SAN DIEGO CA. Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a200237.pdf

Arnold, J., Randall, R., Patterson, F., Silvester, J., Robertson, I., Harris, D., Cooper, C., Burnes, B, & Axtell, C. (2016). *Work Psychology: Understanding human behaviour in the workplace* (6th ed.). Harlow: Pearson Education.

Avolio, B. J., & Bass, B. M. (1994). *Improving Organizational Effectiveness through Transformational Leadership*. Thousand Oaks, CA: SAGE.

Baird, L. L. (1985). Do grades and tests predict adult accomplishment? *Research in Higher Education, 23*, 3-85.

Barrick, M. R., Shaffer, J. A., & DeGrassi, S. W. (2009). What you see may not be what you get: Relationships among self-presentation tactics and ratings of interview and job performance. *Journal of Applied Psychology, 94*(6), 1394-1411.

Bartone, P. T., Eid, J., Johnsen, B. H., Laberg, J. C., & Snook, S. A. (2009). Big five personality factors, hardiness, and social judgment as predictors of leader performance. *Leadership & Organization Development Journal*, *30*, 498–521.

Bass, B. M., & Bass, R. (2008). *The Bass Handbook of Leadership: Theory, Research and Managerial Applications*. New York: The Free Press.

Ben-Shalom, U., & Shamir, E. (2011). Mission Command between theory and practice: the case of the IDF. *Journal of Defence and Security Analysis, 27*(2), 101-117.

Bertua, C., Anderson, N., & Salgado, J. F. (2011). The predictive validity of cognitive ability tests: A UK meta-analysis. *Journal of Occupational and Organizational Psychology,* 78(3), 387-409.

Borman, W. C. (1982). Validity of Behavioral Assessment for Predicting Military Recruiter Performance. *Journal of Applied Psychology, 67*(1), 3-9.

Campion, M. A. (1978). Identification of variables most influential in determining interviewers' evaluations of applicants in a college placement center. *Psychological Reports, 47*, 947-952.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50*(3), 655-702.

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*(4), 247-264.

Carroll, J. B. (1993). *Human Cognitive Abilities: A Survey of Facotr-Analytic Studies*. New York: Cambridge University Press.

Chan, K.-Y., Soh, S., & Ramaya, R. (2012). *Military Leadership into the 21st Century. Science and Practice*. Singapore: Cengage Learning Asia.

Connelly, B. S., & Ones, D. S. (2010). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin, 136*(6), 1092-1122.

Conway, J. M., & Lance, C. E. (2010). What reviewers should expect from authors regarding common method bias in organizational research. *Journal of Business and Psychology, 25*, 325-334.

Cooper, C. (2010). *Individual Differences and Personality* (3rd ed.). London: Hodder Education.

Dilchert, S. (2017). Cognitive Ability. In C. Viswesvaran, H. K. Sinangil, D. S. Ones, & N. Anderson (Eds.), *The SAGE Handbook of Industrial, Work & Organizational Psychology: Personnel Psychology and Employee Performance* (Vol. 1, pp. 396-449). London: SAGE Publications.

Dobson, P., & Williams, A. (1989). The validation of the selection of male British Army officers. *Journal of Occupational Psychology, 62*, 313-325.

Farr, J. L., & Tippins, N. T. (2010). *Handbook of employee selection.* New York: Taylor & Francis Group.

Fleishman, E. A. (1953). The description of supervisory behaviour. *Personnel Psychology, 37*, 1-6.

Forsvarets høyskole. (2013). *Nasjonal rammeplan for grunnleggende befalsutdanning i Forsvaret (GBU).* Oslo: Forsvarets høyskole.

Forsvarets personell- og vernepliktssenter. (2016). *Forsvarets opptak og seleksjon sommer 2016 – erfaringsrapport*. Hamar: Forsvarets personell- og vernepliktssenter.

Forsvarets personell- og vernepliktssenter. (2019a). *Evaluering av Forsvarets seleksjonsprosess*. Hamar: Forsvarets personell- og vernepliktssenter.

Forsvarets personell- og vernepliktssenter. (2019b). *Forsvarets opptak og seleksjon – anbefaling etter innføring av ordning for militært tilsatte og utdanningsreformen.* Hamar: Forsvarets personell- og vernepliktssenter.

Forsvarets Sanitet. (2018). *Helse for stridsevne: Nøkkeltall og vurderinger fra Forsvarets helseregister*. Sessvollmoen: Forsvarets Sanitet.

Forsvarsdepartementet. (2012). *Prop. 73 s (2011-2012) et forsvar for vår tid.* Oslo: Forsvarsdepartementet.

Forsvarsdepartementet. (2013). *Meld. St. 14 (2012-2013) Kompetanse for en ny tid*. Oslo: Forsvarsdepartementet.

Forsvarssjefen. (2008). *Forsvarssjefens beslutningsnotat nr. 2/2008* . Oslo: Forsvaret.

Forsvarsstaben. (2012). *Forsvarssjefens grunnsyn på ledelse i Forsvaret*. Oslo: Forsvarsstaben.

Forsvarsstaben. (2014). *Forsvarets fellesoperative doktrine*. Oslo: Forsvarsstaben.

Fosse, T. H., Skogstad, A., Einarsen, S. V., & Martinussen, M. (2019). Active and passive forms of destructive leadership in a military context: a systematic review and meta-analysis. *European Journal of Work and Organizational Psychology*, *28*(5), 708-722.

Fosse, T. H., Buch, R., Säfvenbom, R., & Martinussen, M. (2015). The impact of personality and self-effiacacy on academic and military performance. The mediating role of self-efficacy. *Journal of Military Studies*, *6*(1), 24-37.

Grebstad, K., & Johnsen, Å. (2019). Forskjellen mellom militært lederskap og styring. In R. B. Johansen, T. H. Fosse & O. Boe (Eds.). *Militær Ledelse* (1st ed., pp. 39-61). Bergen: Fagbokforlaget.

Guilford, J. P. (1988). Some changes in the structure-of-intellect model. *Educational and Psychological Measurement, 48*(1), 1-4.

Hannah, S., John., M., Peng, A. C., Lord, R. G., Trevino, L. K., Kozlowski, S. W., . . . Doty, J. (2013). Joint influences of individual and work unit abusive supervision on ethical intentions and behaviors: A moderated mediation model. *Journal of Applied Psychology*, *98*, 579–592.

Hannah, S. T., & Snowden, W. J. (2013). Leadership in the profession of arms. In M. Rumsey (Ed.), *The Oxford Handbook of Leadership* (pp.291-310). New York: Oxford Press University.

Haslam, S. A., Reicher, S. D., & Platow, M. J. (2011). *The new psychology of leadership, identity, influence and power.* East Sussex: Psychology Press.

Hough, L. M., & Ones, D. S. (2001). The structure, measurement, validity and use of personality variables in industrial, work, and organisational psychology. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of Industrial, Work and Organisational Psychology* (1st ed., pp. 233-277). London: SAGE.

House, R. J., & Javidan, M. (2004). Overview of GLOBE. In R. J. House, P. J. Hanges, M. Javidan, P. W. Dorfman, & V. Gupta (Eds.). *Culture, Leadership, and Organizations. The GLOBE Study of 62 Societies* (pp. 9-28). Thousand Oaks, CA: SAGE.

Huffcutt, A. I., Conway, J. M., Roth, P. L., & Klehe, U. C. (2004). The impact of job complexity and study design on situational and behavior description interview validity. *International Journal of Selection and Assessment, 12*(3), 262-273.

Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81(*5), 459-473.

Hunter, J. E. (1983). A causal analysis of cognitive ability, job knowledge, job performance, and supervisor ratings. In F. Landy, S. Zedeck, & J. Cleveland, *Performance measurement and theory* (pp. 257-266). Hillsdale, NJ: Erlbaum.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*, 340–362.

Hunter, J. E., & Hunter, R. F. (1984). Validity and Utility of Alternative Predictors of Job Performance. *Psychological Bulletin, 96*(1), 72-98.

Hæren. (2018). *Konsept for leder- og ledelsesutvikling i Hæren*. Oslo: Forsvarets høgskole.

Jank, W. (2011). *Business Analytics for Managers*. New York: Springer.

Janz, T. (1989). The patterned behaviour description interview: The best prophet of the future is the past. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research and practice* (pp. 158-168). Thousand Oaks, CA: Sage.

Jensen, A. R. (1986). g: Artifact or reality? *Journal of Vocational Behavior, 29,* 301-331.

Johansen, R. B., Fosse, T. H., & Boe, O. (2019). Hva er militær ledelse? In R. B. Johansen, T. H. Fosse & O. Boe (Eds.). *Militær Ledelse* (1st ed., pp. 15-38). Bergen: Fagbokforlaget.

Judge, T. A., Bono, J. E., Ilies, R., & Gerhardt, M. W. (2002). Personality and leadership: a qualitative and quantitative review. *Journal of Applied Psychology, 87*(4), 765-780.

Judge, T. A., Colbert, A. E., & Ilies, R. (2004). Intelligence and Leadership: A Quantitative Review and Test of Theoretical Propositions. *Journal of Applied Psychology, 89*, 542- 552.

Judge, T. A., Piccolo, R. F., & Ilies, R. (2004). The forgotten ones? The validity of consideration and initiating structure in leadership research. *Journal of Applied Psychology, 89*(1), 36-51.

Kjenstadbakk, T. J. (2012) Seleksjon til befalsskolen. En evaluering av seleksjonssystemets prediktive validitet (Masters Thesis). Oslo: Forsvarets høgskole. Retrieved 07.01.20 from http://hdl.handle.net/11250/100096

Klehe, U. C., & Latham, G. (2006). What would you do – really or ideally? Constructs underlying the behavior description interview and the situational interview in predicting typical versus maximal performance. *Human Performance, 19*(4), 357-382.

Kostal, J. W., Kuncel, N. R., & Sackett, P. R. (2016). Grade inflation marches on: Grade increases from the 1990s to 2000s. *Educational Measurement: Issues and Practice, 35*(1), 11-20.

Krajewski, H. T., Goffin, R. D., McCarthy, J. M., Rothstein, M. G., & Johnston, N. (2006). Comparing the validity of structured interviews for managerial-level employees: Should we look to the past or focus on the future? *Journal of Applied Occupational and Organizational Psychology, 79*(3), 411-432.

Kuncel, N. R., & Hezlett, S. A. (2007). Standardized tests predict graduate students' success. *Science, 315*(5815), 1080-1081.

Kuncel, N. R., & Sackett, P. R. (2014). Resolving the assessment centre construct validity problem (as we know it). *Journal of Applied Psychology, 99*(1), 38-47.

Kurz, E., & Bartram, D. (2002). Competency and individual performance: Modelling the world of work. In I.T. Robertson, M. Callinan & D. Bartram (Eds.), *Organizational effectiveness: The role of psychology* (pp.227-255). Chichester: Wiley.

Lacerenza, C. N., Reyes, D. L., Marlow, S. L., Joseph, D. L., & Salas, E. (2017). Leadership training design, delivery and implementation: A meta-analysis. *Journal of Applied Psychology, 102*(12), 1686-1718.

Lai, L. (2010). *Strategisk kompetansestyring* (2 ed.). Bergen: Fagbokforlaget.

Lance, C. E. (2008). Why assessment centres do not work the way they are supposed to. *Industrial and Organisational Psychology, 1*, 84-97.

Lang-Ree, O. C., & Martinussen, M. (2019). Seleksjon av militære ledere. In R. B. Johansen, T. H. Fosse & O. Boe (Eds.). *Militær Ledelse* (1st ed., pp. 97-116). Bergen: Fagbokforlaget.

Latham, G. P., & Saari, L. M. (1984). Do People Do What They Say? Further Studies on the Situational Interview. *Journal of Applied Psychology, 69*(4), 569-573.

Levashina, J., Hartwell, C. J., Morgeson, F. P., & Campion, M. A. (2014). The structured employment interview: Narrative and quantitative review of the research literature. *Personnel Psychology, 67*, 241-293.

Lievens, F., Ones, D. S., & Dilchert, S. (2009). Personality scale validities increase throughout medical school. *Journal of Applies Psychology, 94*(6), 1514-1535.

Locke, E. A., & Latham, G. P. (1990). *A theory of goal setting and task performance.* Englewood Cliffs, NJ: Prentice Hall.

Lord, R. G., De Vader, C. L., & Alliger, G. M. (1986). A Meta-Analysis of the Relation Between Personality Traits and Leadership Perceptions: An Application of Validity Generalization Procedures. *Journal of Applied Psychology, 71*(3), 402-410.

Macan, T. (2009). The employment interview: A review of current studies and directions for future research. *Human Resource Management Review, 19*(3), 203-218.

Maddi, S. R. (2007). Relevance of hardiness assessment and training to the military context. *Military Psychology*, *19*, 61–70.

Martinsen, Ø. (2009). *Perspektiver på ledelse.* Oslo: Gyldendal Akademisk.

Martinsen, Ø. L., Fosse, T. H., & Johansen, R. B. (2019). Effektiv ledelse i militære organisasjoner. In R. B. Johansen, T. H. Fosse & O. Boe (Eds.). *Militær Ledelse* (1st ed., pp. 157-186). Bergen: Fagbokforlaget.

Martinsen, Ø., Fosse, T. H., Johansen, R. B., & Venemyr, G. O. (Unpublished). A New Perspective on Military Leadership.

Martinussen, M., & Hunter, D. (2008). *Luftfartspsykologi*. Bergen: Fagbokforlaget.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The Validity of Employment Interviews: A comprehensive Review and Meta-Analysis. *Journal of Applied Psychology, 79*(4), 599-616.

McHenry, J. J., Hough, L. M., Toquam, J. L., Hanson, M. L., & Ashworth, S. (1990). Project A validity results: The relationship between predictor and criterion domains. *Personnel Psychology, 43*, 335-354.

Melchers, K. G., & Annen, H. (2010). Officer selection for the Swiss Armed Forces: An evaluation of validity and fairness issues. *Swiss Journal of Psychology, 69*, 105-115.

Meriac, J. P., Hoffman, B. J., Woehr, D. J., & Fleisher, M. S. (2008). Further evidence for the validity of assessment center dimensions: A meta-analysis of the incremental criterion-related validity of dimension ratings. *Journal of Applied Psychology, 93*, 1042-1052.

Murphy, K. R. (1996). *Individual differences and behavior in organizations.* San Francisco: Jossey-Bass.

Norrøne, T. (2016). Utvelgelse av kandidater til Sjøforsvarets grunnleggende befalskurs: En undersøkelse av den prediktive validiteten til seleksjonssystemet. (Master's Thesis). Trondheim: Norges Teknisk-Naturvitenskapelige Universitet. Retrieved 24.05.20 from https://ntnuopen.ntnu.no/ntnu-xmlui/handle/11250/2402099

Parrington, J. & Findlay, M. (2013). Mission command – addressing challenges and sharing insights. *Joint Force Quarterly, 71*(4), 103-107.

Piccolo, R. F., Bono, J. E., Heinitz, K., Rowold, J., Duehr, E., & Judge, T. A. (2012). The relative impact of complementary leader behaviours: which matter most? *The Leadership Quarterly, 23*(3), 567-581.

Posthuma, R. A., Morgeson, F. P., & Campion, M. A. (2002). Beyond Employment Interview Validity: A Comprehensive Narrative Review of Recent Research and Trends Over Time. *Personnel Psychology, 55*, 1-81.

Ree, M. J., & Earles, J. A. (1991). Predicting training success: Not much more than *g*. *Personnel Psychology, 44*, 321-332.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1*, 86-89.

Reilly, R., & Chao, G. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 35*(1), 1-62.

Richardson, M., Abraham, C., Bond, R. & Hinshaw, S. P. (2012). Psychological Correlates of University Students' Academic Performance: A Systematic Review and Meta-Analysis. *Psychological bulletin, 138*(2), 353-387.

Robertson, I. T., & Smith, M. (2001). Personnel Selection. *Journal of Occupational and Organizational Psychology, 74*(4), 441-472.

Roth, B., Becker, N., Romeyke, S., Schäfer, S., Domnik, F., & Spinath, F. M. (2015). Intelligence and school grades: A meta-analysis. *Intelligence, 53*, 118-137.

Roth, P. L., BeVier, C. A., Switzer, F. S., & Schippmann, J. S. (1996). Meta-Analyzing the Relationship Between Grades and Job Performance. *Journal of Applied Psychology, 81*(5), 548-556.

Rupp, D. E., Hoffman, B. J., Bischof, D., Byham, W., Collins, L., Gibbons, A. M., … Simonenko, S. (2015). Guidelines and Ethical Considerations for Assessment Center Operations: International Taskforce on Assessment Center Guidelines. *Journal of Management, 41*(4), 243-253.

Ryan, A. M., & Ployhart, R. E. (2014). A century of selection. *Annual Review of Psychology, 65*, 693-717.

Sacket, P. R., & Wilk, S. L. (1994). Within-group norming and other forms of score adjustment in pre-employment testing. *American Psychologist, 49*, 929-954.

Sackett, P. R., & Yang, H. (2000). Correction for Range Restriction: An Expanded Typology. *Journal of Applied Psychology, 85*(1), 112-118.

Salas, E., Tannenbaum, S. I., Kraiger, K., & Smith-Jentsch, K. A. (2012). The science of training and development in organizations: What matters in practice. *Psychological Science in the Public Interest, 13*(2), 74-101.

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), 1-17.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & De Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European Community meta-analysis. *Personnel psychology, 56*(3), 573-605.

Salgado, J. F., Viswesvaran, C. & Ones, D. S. (2001). Predictors Used for Personnel Selection: An Overview of Constructs, Methods ans Techniques. In N. Anderson, D. S. Ones, H. K. Sinangil & C. Viswesvaran (Eds.), *Handbook of Industrial, Work & Organizational Psychology, Volume 1: Personnel Psychology* (pp. 165-199). London: SAGE Publications.

Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there can be no debate. *Human Performance, 15*, 187-210.

Schmidt, F. L., & Hunter, J. E. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science, 1*, 89-92.

Schmidt, F. L., & Hunter, J. E. (1998). The Validity and Utility of Selection Methods in Personnel Psychology: Practical and Theoretical Implications of 85 Years of Research Findings. *Psychological Bulletin, 124*(2), 262-274.

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.

Schmidt, F. L., Hunter, J. E., McKenzie, R. C., & Muldrow, T. W. (1979). The impact of valid selection procedures on work-force productivity. *Journal of Applied Psychology, 64*, 609-626.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). The impact of job experience and ability on job knowledge, work sample performance and supervisory ratings of job performance. *Journal of Applied Psychology, 71*, 432-439.

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) Retrieved 25.05.20 from https://testingtalent.com/wp-content/uploads/2017/04/2016-100-Yrs-Working-Paper-on-Selection-Methods_Schmidt-Mar-17.pdf

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43,* 627-670.

Schmidt, F. L., Shaffer, J. A., & Oh, I.-S. (2008). Increased accuracy of range restriction corrections: Implications for the role of personality and general mental ability in job and training performance. *Personnel Psychology, 61*, 827-868.

Schmitt, N., Cortina, J. M., Ingerick, M. J., & Wiechmann, D. (2003). Personnel Selection and Employee Performance. In W. C. Borman, R. J. Klimoski, D. R. Ilgen, & I. B. Weiner (Eds.), *Handbook of Psychology* (12 ed., pp. 77-106). Hoboken, New Jersey: John Wiley & Sons, Inc.

Schmitt, N., Gooding, R. Z., Noe, R. A., Kirsch, M. (1984). Meta-analyses of validity studies published between 1964 and 1982 and the investigation of study characteristics. *Personnel Psychology, 37*, 407-422.

Skoglund, T. H., Martinussen, M., & Lang-Ree, O. C. (2014). Papir versus pc. *Tidsskrift for Norsk Psykologforening, 51*(6), 450-452.

Skorstad, E. (2015). *Rett person på rett plass – Psykologiske metoder i rekruttering og lederutvikling* (2nd ed.). Oslo: Gyldendal Akademisk.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201-293.

St.prp.111 LS (2014-2015) *Ordningen for militært tilsatte og endringer i forsvarspersonelloven m.m. (militærordningen)*. Oslo: Forsvarsdepartementet.

Stabsskole, F. H. F. (2015). *Prosjektbeskrivelse: Lederkandidatstudien 2015-2020 En studie av kultur, profesjonsidentitet og individuelle faktorer i Forsvarets seleksjonsprosess og lederutdanning*.

Stehlik-Barry, K. & Babinec, A. J. (2017*). Data Analysis with IBM SPSS Statistics*. Birmingham: Packt Publishing, Limited.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence, 32*(4), 349-362.

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29*, 332–339.

Vik, J. S. (2013). Har seleksjon noen betydning? En studie av seleksjonens prediktive validitet (Master's Thesis). Tromsø: Universitetet i Tromsø. Retrieved 06.01.20 from https://munin.uit.no/handle/10037/5310.

Viswesvaran, C., & Ones, D. O. (2018). Non-Test Methods and Techniques Used in Employee Selection 1. In C. Viswesvaran, H. K. Sinangil, D. S. Ones, & N. Anderson (Eds.), *The SAGE Handbook of Industrial, Work & Organizational Psychology: Personnel Psychology and Employee Performance)* (Vol. 1, pp. 724-768). London: SAGE Publications.

Wiesner, W. H., & Cronshaw, S. F. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 61*(4), 275-290.

Wilk, S. L., Desmarais, L. B., & Sackett, P. R. (1995). Gravitation to jobs commensurate with ability: Longitudinal and cross-sectional tests. *Journal of Applied Psychology, 80*, 79-85.

Yukl, G. (2012). Effective leadership behaviour: what we know and what questions need more attention. *Academy of Management Perspectives, 26*(4), 66-85.

Yukl, G. (2013). *Leadership in Organizations* (8th ed.). Essex: Pearson Education Limited.

Zikmund, W. G., Hitt, M. A., & Pickens, B. A. (1978). Influence of sex and scholastic performance on reactions to job applicant resumés. *Journal of Applied Psychology, 63*, 252-254.