UiT The Arctic University of Norway

Department of Technology and Safety

Good enough for government work?

A master's thesis on airmanship and military pilots' perception of competence

Mattis Ekeland Paulsen

MSc in Aviation, FLY-3930, June 2024

20685 words

Abstract

The purpose of this study was to examine the relationship between characteristics of the job as a pilot in the Royal Norwegian Air Force and how they affect the individual's perception of competence. The characteristics in question were *experience*, *currency*, *workload* and *complexity*. The empirical data used in the study (*n = 16)* was collected through a questionnaire distributed to three operational fixed-wing squadrons, flying differing aircraft with their own unique mission sets, with the intention of capturing trends prevalent across the whole organization. Linear regression analysis provided the basis for addressing the primary research question and a statistically significant final model ($R^2 = .70$) was obtained containing the predictor variables *experience (flight hours)* and *workload*. *Flight hours* displayed a strong positive relationship with *perception of competence*, while *workload* exhibited a strong negative association, and emerges as the most interesting result from the study. The data indicates that a majority of the pilots in the study are working close to a *workload redline*, with a particular disparity between the *time required* to execute task demands posed by the job, and the *time available* to complete them. As such, little tolerance is left for extra workload without incurring into an overload region where performance declines. Thus, the study provides information to stakeholders as how to foster a work environment in a 'Goldilocks zone' that stimulate for optimized competency and performance among pilots.

Table of contents

1 Introduction

1.1 Background

I have eight years of personal experience as a military pilot at an operational squadron in Norway. During that time, I managed to acquire ~2100 flight hours and completed all the pilot upgrades offered at the squadron level. When I left my tenure at the 333 Squadron I was qualified as Patrol Plane Commander, Instructor Pilot, Functional Check Flight Pilot, Simulator Instructor and Instrument Evaluator on the P-3 Orion. In addition to the qualifications in the aircraft I was also part of the leadership at the squadron and head of one of three *Flights*[1]. This is not a plug for my CV, but an example of what a typical career looks like for an intermediately experienced pilot in the RNoAF. For every new qualification acquired it felt like another layer of *complexity* was added to my job, and subsequent additional effort and time was required to stay *current* and *competent*. However, this increase in *workload* was not compensated for by more time available to study or practice my aviation skills. Quite the contrary, as I ascended the qualification-ziggurat, my non-flight related workload increased as well, resulting in less time available for flight training. This seemed like a paradox, and a failure of the organization to recognize the complexity involved in maintaining these qualifications and to fully acknowledge the implications of the *opportunity cost*-effect (Rønningstad, 2024). As a result, the workload at times felt overwhelming, and a dissatisfying sensation of flying without feeling sufficiently competent emerged. This is part of the motivation behind the study, namely, to investigate deeper if my experience was an outlier or a common feeling among pilots across squadrons in the RNoAF.

1.2 Actualization

Safety is and always has been one of the top priorities of aviation since its inception in 1903 (Perrow, 1999, p. 124). At the core of safe flight operations lies competent performance employed by the humans operating the aircraft, with roughly 80% of all modern aviation mishaps having been attributed to human error (Mendonca, Keller, & Dillman, 2019, p. 3). As such, the overall competence of pilots, or what pilots themselves call good airmanship, is a key component to aviation safety (Nergård, 2014). This has been acknowledged by major aviation

[1] Squadrons in the RNoAF are typically divided into *Wings* or *Flights*

institutions in switching to training systems where pilot competencies are at the forefront of what's being trained to and used as a performance marker (IATA, 2023; ICAO, 2013).

Many studies have been conducted trying to measure the output of competence, i.e. competent performance (Baker & Dismukes, 2002). Few, if any studies, have focused on asking the pilots themselves how they regard their own competence and if this is at a satisfactory level to meet the performance criteria demanded by their job. This seems like a missed opportunity, as it appears reasonable to assume that there is a strong correlation between the perception of competence, and actual competence. While objective indicators are regarded as the most reliable and valid indicators of competency and performance, some studies have found concordance between self-report data and objective performance markers (Adler, Thomas, & Castro, 2005; Johansen, Laberg, & Martinussen, 2014, p. 527). Moreover, high self-awareness is a highly desirable personality trait that pilots are being selected for (Nergård, Hatlevik, Martinussen, & Lervåg, 2012), so a complete disconnect between perceived and actual competence would represent the opposite of self-awareness. Aviation writer Dr. Tony Kern (1997, p. xxxii) gets to the essence of this study with his quote:

> *"The airman is still the single largest variable on any aircraft, and no institutional training or evaluation system can ever approach the capability of the internal barometer that lies within each of us for assessing our personal state of competency."*

Thus, it seems logical that by gauging the average perception of competence among pilots in an organization can serve as an indicator of the overall performance level in that group. If a population of pilots report feelings of low competence this information could serve as a canary in a coal mine to an organization striving for safety and efficiency.

Furthermore, the feeling of being competent is also important from a job satisfaction perspective. As previous studies have summarized, there is a positive relationship between job performance and organizational commitment, well-being and lower turnover rates (Johansen et al., 2014, p. 536; W. Li, Lee, & Solmon, 2005). A meta-analysis conducted by Judge and Bono (2001) highlighted self-efficacy as one of the best predictors of job satisfaction. High self-efficacy will likely lead to high job satisfaction, in turn leading to stronger organizational commitment and lower turnover rates. All desirable characteristic an organization with highly specialized and difficult-to-replace personnel should seek to maximize. Thus, measuring the perception of competence will also be important from an organizational stability and efficiency standpoint. As such, it seems important for an aviation organization to have pilots that are feeling competent *both* from a performance standpoint *and* from a job satisfaction standpoint.

This thesis will delve into these issues and try to uncover what dimensions of the job as a military pilot contribute to variation in reported competence level. These characteristics are *experience, currency, workload,* and *complexity*. The variables have been chosen from a theoretical basis inspired by Hoffmann (1999) and an empirical framework based on what measures are typically encountered in aviation. The constructs of interest and how they related to competence are visualized in italic text in Figure 1.

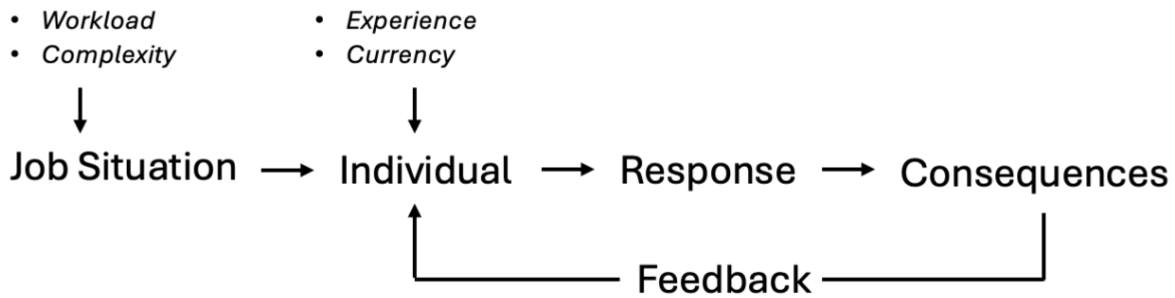

Figure 1: Situation-specific model of human performance. Adapted from Hoffmann (1999, p. 277)

From this model, *workload* and *complexity* can be seen as constructs of the job situation and contribute to regulate the job demands that the individual must cope with. The individual needs underlying attributes, such as knowledge, skills, attitudes and other factors (KSAO)[2], to meet the demands (ibid). These attributes of the individual are closely coupled with feedback from prior events (*experience*) and recency in training (*currency*) and drive variation in the response to a given situation. The consequence from the response informs the individual of the correctness of the action and gives feedback to the individual regarding their level of achievement. The level of achievement drives variation in the feedback and hence influence the *perception of competence* for the individual. This is how at its most basic level the characteristics of interest in the study influence variation in the feeling of competence.

More specifically for aviation, *Experience* is closely recorded for all professional aviators in the form of flight hours and seniority. *Currency* is imposed through stringent annual training

[2] KSAO is an acronym used to refer to the attributes required to do a job well (Damos, 2011).

requirements and flights are meticulously logged to keep track of pauses in flying before recurrent training measures are triggered. *Complexity* typically increases with experience as aviators get qualified in more advanced aircraft types or acquire additional qualifications on their existing airframe. *Workload* in this study is regarded as the subjective feeling of the amount of effort required to complete the cumulative sum of all duties demanded by the job and the training and self-study necessary to remain competent. All these variables are assumed to cause variation in the perception of competence, and expanding the knowledge base of this contribution can help organizations drive closer towards a "Goldilocks-zone" where the working environment is optimized for competent and satisfied pilots. This provides the basis for the primary and secondary research questions.

1.3 Research Questions

1.3.1 Primary Research Question

How do characteristics of the profession as a military aviator cause variation in the perception of competence?

The characteristics under question in this project are *experience, complexity, workload,* and *currency.* These characteristics will serve as the predictor variables for the study. *Competence* will serve as the outcome variable, and the construct is considered with respect to airmanship, i.e. competencies relevant for executing the tasks as a pilot well[3]. This involves both a flight safety aspect and for military aviators a tactical aspect, what the Air Operations Doctrine calls *airmindedness* (Haga & Maaø, 2018, p. 138).

1.3.2 Secondary Research Question

Are pilots in the RNoAF satisfied with their own level of competence?

In plain words, do pilots feel good enough for their job as Norwegian Air Force pilots?

[3] The terms «predictor» and «outcome» variables are chosen in this thesis as the selected phrases for describing the *independent* and *dependent* variables. These expressions are selected on the basis of a more intuitive understanding of which variables contribute in predicting (independent) the outcome (dependent) as recommended by Field, Miles, and Field (2012, p. 7).

1.4 Research Data

The empirical basis for this study is the data collected from the questionnaire sent out to military pilots actively flying at three operational squadrons in the RNoAF. The data is then analyzed, and a multiple regression analysis is performed and will form the principal basis for answering the primary research question. Satisfaction with competence is also measured in the survey and a plot of Likert data and a paired t-test is produced to address the secondary research question.

1.5 Definitions

Competence, competency and competencies are terminologies that are frequently being used in the scientific literature in a wide variety of topics such as psychology, management, education, politics, and not the least aviation (Hoffmann, 1999; Moore, Cheng, & Dainty, 2002). However, the terms have often been used somewhat imprecisely and interchangeably, allowing confusion to emerge regarding their exact meaning. This thesis will adopt the definitions suggested by Arifin (2021, p. 761) as visualized in Figure 2.

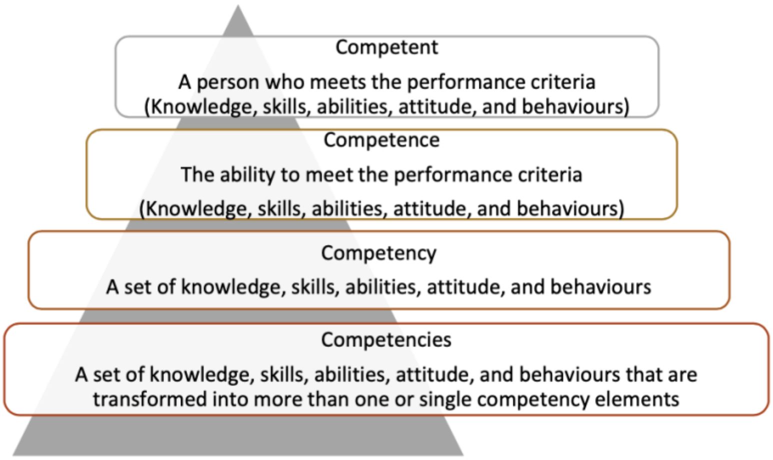

Figure 2: Definitions of Competency

From this figure, it is the feeling of *competence*, or the state of feeling *competent*, that is the construct of interest in this study. To enable these feelings thus requires underlying *competencies*, or *competency*, resulting in the ability to execute competent performance. As such, there are subtle, but important distinctions between the terms and it is the goal to employ them with precision throughout the thesis.

1.6 Delimitations

The competencies addressed in the study are the ones relevant for carrying out the job as a military aviator only. The role as an officer and pilot in the military also involves competencies related to other tasks not related to flight operations, but these are outside the scope of this thesis to properly maintain focus on the theme of aviation. Maintaining this focus is also relevant based on the assumption that military pilots apply because of their motivation to primarily become a pilot, not an officer (Grindheim, p.23). Furthermore, the study only regards the pilot's subjective perception of the variables in question and assumes there is correlation between *perceived* and *actual* conditions. No effort is made to uncover whether the self-reported data correspond to the *true* state of conditions.

The project is an observational study only with no control or experimental groups for comparison. No longitudinal data is collected; thus, the results will only provide a snapshot of how the participating pilots feel about their job situation and competence at a specific point in time.

2 The Royal Norwegian Air Force

2.1 Selection and Training

The RNoAF is a small, but highly competent organization employing a modern fleet of various fixed- and rotary wing aircraft. As a numerically inferior force highly reliant upon cooperation with allied forces a high level of competence must permeate the organization to compensate for what it lacks in numbers (Haga & Maaø, 2018, p. 151). The small size results in a low volume of yearly candidates selected for pilot training and a small headcount of active pilots employed at any given time. The low volume of candidates makes initial pilot training prohibitively expensive to run internally, and therefore all the foundational flight training up until type-rating on operational aircraft are conducted in the USA, training with either the US Navy, Air Force or Army (Grindheim, 2023, p. 3). This requires the Norwegian Defense Sector to pay upfront for all training slots bought from the US Department of Defense with no options for a refund if a candidate were to fail the training in the US. This results in Norwegian pilot candidates going through a rigorous selection process prior to being sent stateside and these student pilots have typically performed well at their respective American training commands, with a very low attrition rate (Harlem, 2016, p. 3). As such, the RNoAF is privileged in receiving high performing pilots returning for duty in Norway.

For *ab initio* pilots arriving at the squadron the primary focus is to get qualified in the specific airframe type flown in Norway, which usually have slight modifications from the aircraft flown in the US. Attention is given mainly to flight duties with the goal of passing the squadron specific qualification program which is a prerequisite to become an operational pilot. The syllabus includes elements like learning the national specific aircraft procedures, squadron specific SOP, operating in Norwegian air space with winter operations, mountainous terrain, ICAO procedures and so on. National tactical procedures are introduced as well. As such, there is a lot to learn the first year at the squadron and non-flying related duties are usually kept to a minimum.

After some years, pilots can expect to study and qualify for new ratings which are typically platform dependent. Fighter pilots upgrade from wingman to formation lead, big wing pilots from co-pilot to pilot-in-command and so forth. This can occur at a relatively young age, as compared to their civilian counterparts. Upgrading to pilot-in-command represents a leap in responsibility as the aircraft commander has the highest authority on board and is responsible for mission execution, maneuvering and navigating the aircraft, alongside with all other aspects pertaining to flight safety, regardless of other crew or passengers rank (Luftforsvaret, 2017, p. 22). This highlights a unique characteristic for the Air Force as a branch of the military, namely the reliance on expert operators regardless of rank. In the air, competence trumps rank (Haga & Maaø, 2018, p. 142).

Alas, with seniority follows more duties and responsibilities on the ground as well, as pilots typically hold leadership positions at the squadrons, and throughout the Air Force in general. This is related to the authority held by experienced pilots, as they, alongside with certain navigators, are the only personnel with the legal power to authorize flight operations (Luftforsvaret, 2017, p. 15).

2.2 Situation

The RNoAF of 2024 is affected by recent major acquisition programs like the introduction of the F-35, P-8 and AW101. The new platforms are yet to become fully operational, and the massive amount of work partaken in achieving FOC[4] is likely to continue until 2030, putting extra stress on those involved and restrictions on operational capabilities (Rjaanes et al., 2020, p. 3). This also entail that the personnel working on these platforms must learn to operate and understand novel cutting-edge technology, which is why the need for increased technological knowledge in general has been stated as an important consequence of the implementation of these aircraft (ibid, p.14). Moreover, as a reaction to rising geopolitical tensions a historical defense budget increase was recently proposed by the Norwegian government, adding new capabilities to the Air Force like unmanned airborne systems and new helicopters (Prop 87 S, 2024), further increasing the need to quickly adapt, learn, and utilize new competencies. This trait is stated as one of the hallmarks of airpower and requires a concept of life-long learning (Haga & Maaø, 2018, p. 133).

The introduction of new platforms also poses challenges on safety as best practices has yet to be learned and the experience level is low. The Air Force has adopted the motto "*Mission First, Safety Always"* and has explicitly stated that safety shall permeate the organization in all operations, regardless of level of conflict (Haga & Maaø, 2018, p. 124). To promote safety, the Air Force has employed modern aviation safety tools like CRM/TEM, Operational Risk Management, adoption of the *just culture*-principles and thorough briefings and debrief before

[4] Full Operational Capability (FOC) marks the completion of a development effort.

and after flights. Emphasis has been put on the concept of holistic debrief, which is practiced in all squadrons to provide an arena for transaction of knowledge, assess the operational result of the mission and share emotional reactions (Moldjord, 2023). In this domain, the RNoAF is perceived as progressive and willing to adopt unorthodox practices in the effort of continuous improvement in efficiency and safety. Although safety is considered paramount and is reflected in the track record of the organization, there has still been a few high-profile accidents and incidents in recent years, with the Kebnekaise-accident in 2012 being the worst (SHK, 2013). The accident gave a stark reminder about the inherent risks of flying, especially in military operations where man and machine are pushed to the limits of their capabilities flying fast and low in a potentially hostile environment with little room for error (Haga & Maaø, 2018, p. 145; Perrow, 1999, p. 126). This poses stringent requirements for competent performance by the humans in the loop.

3 Theoretical Framework

Vast amounts of research exist on aviation in general, and on pilot competencies and performance in particular. It is outside the scope of this study to give a conclusive review on the complete body of literature on this subject. The intention of this chapter is to highlight relevant studies to create a theoretical framework to build upon for the remainder of the thesis.

3.1 Literature Review

Several studies on Norwegian military pilots have been conducted over the last few years, some of which are presented below. The most prevalent topic for these studies has been on pilot selection and their predictive validity (Grindheim, 2023; Svensson, 2013). These studies rely on objective data recorded during the selection process and compares them with subsequent performance during the *ab initio* pilot training. Findings suggest that there is concordance between the selection battery and performance in flight school. However, these studies stop short of discussing performance at the squadron level where the pilots are *de facto* performing the job they were selected to do. No such study has, to the authors knowledge, been conducted in Norway, for good reasons. Measures of actual performance at operational squadrons is hard to obtain, expensive and difficult to measure (Wickens & Dehais, 2019) and would likely be protected information, unsuitable for unclassified studies. Therefore, this study aims to do the second-best thing, which is indirectly measure performance by asking the pilots themselves how they regard their competence level, assumed to be an important predictor of performance.

Other recent and relevant studies on Norwegian military pilots have been the job analysis of Harlem (2016) examining the abilities and aptitudes necessary to become a good F-35 pilot. Findings indicate that there has been a shift from *physical* to *cognitive tasks* for pilots with the transition from F-16 to F-35, due to the introduction of new advanced sensors and an airframe that is easier to fly requiring less focus on traditional stick-and-rudder skills. The study of Knutsen (2023) examined aviation safety practices in the Norwegian fighter jet community with a focus on resilience engineering and Safety-II, mainly drawing on the work of Erik Hollnagel and Sidney Dekker. His findings suggest that the Norwegian fighter squadrons exhibit a solid foundation for resilience with an emphasize on the ability to learn and transfer knowledge due to their strong focus on practices such as brief and debrief. A study conducted by Wibe and Hanssen (2023) examined how the implementation of the F-35 affected the safety culture in the fighter community. Their findings indicate a healthy safety culture but notes that the community is characterized by high workload and a competitive work environment.

Finally, Stueland (2023) did an analysis of perception of risk and complacency among the same group of fighter pilots and his research indicate that there is a negative association between experience level and attitudes towards risk, i.e. more experienced pilots perceive certain situations as less risky than their inexperienced counterparts, which ultimately can lead to failures in risk perception.

None of these studies address the topic of this study, and as such there seems to exist a local knowledge gap, at least in Norway, for the topic at hand, indicating that an opportunity exists to gain new insights into the field of competency and performance of Norwegian military pilots.

3.2 Variables

The following paragraphs will lay down the theoretical foundation for the variables of interest in the study and provide further insight as to why they were selected in the first place.

3.2.1 Competence

As described in paragraph 1.5, the terminology regarding competence has been used imprecisely and have seen various definitions from differing authors. No widely accepted single definition exist (Strebler, 1997). However, the common factor that unite the literature on the subject in attempts to come up with definitions is their intention to improve human performance at work (Hoffmann, 1999), which is also the objective of this research. From Figure 2, the definition of competence used in this thesis is *"the ability to meet the performance criteria"*. From this statement there are two dimensions emerging. (1) relates to the *abilities* of an individual/organization, while (2) is concerned with *performance criteria* posing a set of demands. This is similar to the definition of Hoffmann (1999) where he discuss the two-dimensionality of the term *competency*. This thesis will adopt a slight adaptation to the classification from Hoffmann (1999) in that *competence*, not *competency*, has two dimensions applicable to its definition, namely the *input* and *output of competence*.

The input of competence, the *abilities*, is the construct most commonly referred to as *competency* and relates to the underlying attributes, KSAO's, of an individual. An example is found in the Air Doctrine (Haga & Maaø, 2018, p. 133) with the following quote:

> *"Competency can be regarded as the sum of knowledge, skills and attitudes enabling a human and/or organization to solve tasks demanded by the job."*

As such, the first dimension (input) of competence is the *enabler* of the second dimension (output) of competence which is related to *solving tasks demanded by the job*. This is more related to competent performance, or the observable behavior displayed by an individual or organization. Thus, there is a *competency* (input) and *performance* (output) component to the concept of *competence*. The combination of these components will ultimately influence perceptions regarding competence, and *both* components need to be present to invoke a feeling of *being competent.* As often is observed in the literature the quote mixes between individuals and groups of individuals (organizations). An organization cannot *feel* anything in the traditional meaning of the word; accordingly, this study will only investigate competence on the level of individuals.

Definitions regarding the terminology around competence/competency are frequently found in the aviation literature as well with the International Air Transport Association (IATA, 2023, p. 4) defining competency as:

> *"A dimension of human performance that is used to reliably predict successful performance on the job. A competency is manifested and observed through behaviors that mobilize the relevant knowledge, skills, and attitudes to carry out activities or tasks under specified conditions."*

Again, we see the two-dimensionality of input/output emerge, leading to the terms of performance and competency having become almost inextricably intertwined in aviation speak. Examples of what this looks like in practice could be having the skills and know-how to fly an ILS approach down to minimums within parameters, putting weapons on target, leading a formation through a composite air operation or more generally executing a safe flight from start to end. ICAO (2013, p. xii) defines the outcome of these operations as the measurement of performance, and the abilities enabling the outcome as competency. As shown, these terms go hand in hand and have strong concordance with each other. A lack in competency will lead to a lack in performance and so on. Both dimensions are captured in the term *competence.* Therefore, a highly competent pilot is desirable as it will lead to higher performance, desirable outcomes and an increase in mission effectiveness and safety. This construct is often referred to in pilot lingo as *airmanship*.

3.2.1.1 Airmanship

Any two different aviators would likely give slightly different definitions of the term *airmanship* (Nergård, 2014). However, most pilots would agree when they see the absence of airmanship, manifested in operational errors and aviation mishaps frequently labeled as *human error*, with connotations of *"poor airmanship"* seeping through the accident reports (Kern, 1997, p. xxx). Examples of definitions of the term do exist, however, with the FAA (2021, pp. 1-1) providing a broad, but succinct characterization:

> *"A sound knowledge of and experience with the principles of flight, the ability to operate an airplane with competence and precision both on the ground and in the air, and the application of sound judgement that results in optimal operational safety and efficiency."*

Hence, airmanship is deeply intertwined with both a pilot's competency and performance. It has been stated that flyers have a separate identity beyond their official rank or aeronautical rating, and this identity is closely tied up with the concept of airmanship (Flight Safety Australia, 2005, p. 23). It is an internal drive among the individual pilot to seek continuous improvement, not just from a safety standpoint, but beyond to concepts like effectiveness, efficiency, and precision, as depicted in Figure 3.

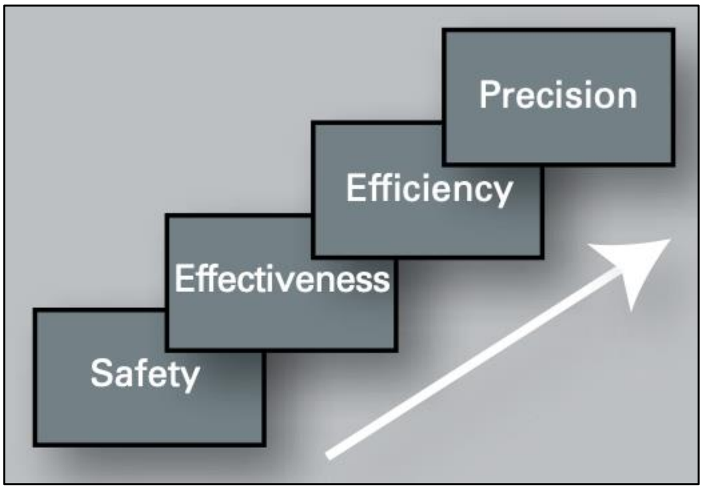

Figure 3: Levels of Airmanship (Kern, 1997)

Both Kern (1997, p. xxx) and Nergård (2014, p. 148) states that professional aviators are obligated to seek the highest standards of airmanship. The concept is indoctrinated from the very beginning of a pilot's career through a fierce selection process with a focus on exceptional performance in a myriad of ability tests, consolidated further by a competitive *ab initio* flight training syllabus with a constant risk of attrition and finally entrenched at the squadrons with constant evaluation of performance. Airmanship is what separates the good from the mediocre, it arranges you in the internal hierarchy of pilots and may materialize in things like selection for aircraft type, job promotions and courses like Top Gun/FWIT[5] and TPS – elements of which have been famously documented in books like Tom Wolfe's "The Right Stuff" and the "Top Gun" movies. As such, the concept is institutionalized in the industry and pilots themselves are highly aware of their place in the hierarchy. Notwithstanding, the concept has evolved from a previous focus on mainly stick-and-rudder skills to now encompassing a complex mix of human, machine, and environmental elements (Kern, 1997, p. xxx). The model from Pierson (2022, p. 242), Figure 4, is an attempt at visualizing some of these elements.

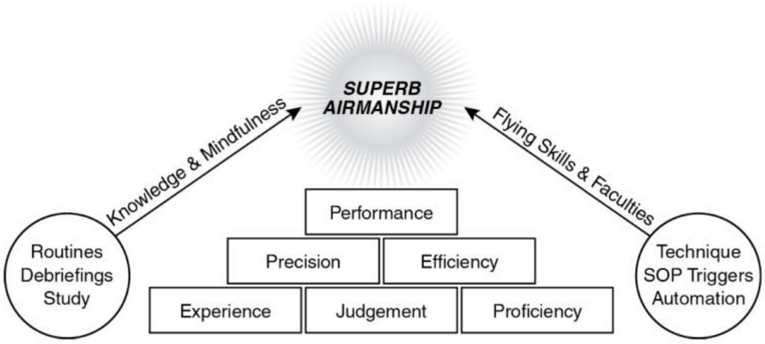

Figure 4: Superb Airmanship Model (Pierson, 2022, p. 243)

Furthermore, the modern approach to safety is now less about building barriers and defenses against errors and more about perceiving safety as an emergent condition created by skilled human practice (Dekker, 2006). This is conceived as a paradigm shift by scholars in safety theory but is construed as semantics by the pilots themselves. Safety is created *on every flight* by each individual pilot through the practice of good airmanship, and this is the way things have always been from a pilot's perspective (Nergård, 2014, pp. 147-148). However, safety is the base level in Figure 3, and ambitious pilots aim for more than just safe operations. They want to employ their vehicle with efficiency and precision, in addition to being safe. Nergård (2014, p. 149) summarizes the concept in this way:

[5] Top Gun is navy slang for USN Strike Fighter Tactics Instructor program where only elite fighter pilots are admitted to become exceptional air combat instructors. FWIT - Fighter Weapons Instructor Training (now called WIC on F-35), is the European equivalent where Norwegian fighter pilots can attend.

"Within the pilot community, airmanship is undisputedly the concept that has been used to define a skilled pilot."

As such, airmanship is an almost inseparable component of a pilot's competency and performance and is why it is embedded in the title of this project. This gives the participants an immediate understanding of the type of competencies the survey is examining.

3.2.2 Experience

"The development of the human mind is a process of learning from our experience" (Bion, 2023, p. iv)

Hence, experience emerges as an important measure of a pilot's competence, manifested through the meticulous logging of flight time, events, training cycles etc. Almost everything a pilot does is documented in some kind of record, providing stakeholders with the ability to track a pilot's overall experience level and currency. There are two main units of measure in this domain, seniority in the form of years of service and flight time. While seniority might influence things like rank, pay, position, social status etc., it is usually flight time that is used as the primary indicator of competence.

Flight time, in the form of *flight hours,* is used to measure a pilot's level of experience and in assessing constructs like when pilots can start their upgrade syllabus[6], as a prerequisite for flight crew licensing[7], as a minimum requirement on pilot job listings, as an icebreaker when sizing up colleagues in the bar and more. These features imply that competency grows with flight time and some studies support this conclusion (Taylor, Kennedy, Noda, & Yesavage, 2007), with G. Li et al. (2003) noting a protective effect of flight experience against risk of crash involvement. However, the evidence is not overwhelming, and some studies reach a less conclusive relationship between flight hours and performance (Todd & Thomas, 2012). As such, there is room for debate regarding how accurately such a unilateral measure reflect pilots' true competence. There is however no doubt that the content of the flight hours flown vary immensely across different aircraft platforms. One flight hour doing air combat maneuvering

---

[6] E.g. minimum flight hours to upgrade from CP to PPC on the P-3 was 1,000 hrs.

[7] Such as the EASA Part-FCL510.A requirement of 1,500 flight hours to get the ATPL(A) rating *(Regulation (EU) No 245/20129)* and the FAA's Extension Act *(Public Law 111-216)* with the same flight hour requirement for holding an ATP certificate in the US.

in an F-35 is certainly a lot different than one hour in transit with the P-8, sowing further doubt around its applicability. However, flight hours are easy to quantify and record and is the most ubiquitous measure of experience in the industry employed today.

Still, more general studies on learning and experience suggest that there is a positive association between the amount of practice, i.e. experience, and perceptions of competency (W. Li et al., 2005). As such, flight hours do to a certain degree reflect amount of practice. Martinussen and Hunter (2018, p. 157) notes in their remark on repetition and performance that *"Although the pace of improvement may slacken, there is never a true plateau in which no improvement occurs."* Suggesting that as more flight experience is accrued performance steadily rises. Thus, experience will be used as a predictor variable in this study and the expectation is that this will positively correlate with perceived competence.

3.2.3 Workload

The concept of workload in aviation has seen historically seen a broad array of definitions. Gartner and Murphy (1976) argues that there is little academic consensus in the conceptualization of the term and several definitions circulate the corpus of articles on the subject. Ratcliffe (1969) reaction to this lack of agreement was that he was ready to accept *"any definition of workload that is not in conflict with common English usage"*.

Workload can be seen either as (1) a set of task demands, (2) the perceived effort experienced by an individual as he works, or (3) as an activity or accomplishment (Gartner & Murphy, 1976, pp. 4-7). It can be further distinguished according to duration, as classified by Howitt (1969) as *immediate, duty-day* or *long-term workload.* "Immediate" would be short-term workload as in the cockpit while conducting a specific flight operation, often referred to as operator/pilot workload. Conversely, "long-term" would be sequences of working days and can be thought of as the integrative overall workload of a job. This thesis aims to measure the latter and should be reflected by the questions stated in the survey.

One of the greatest difficulties regarding studies on workload has been how to measure it (Parks & Boucek Jr, 1989). Objective measurements such as heart rate and other physiological markers has been used but are usually only applicable for short-term workload. One of the most important count variables, particularly suitable for measuring long-term workload, is *time* (ibid). A simple timeline analysis computes the ratio between the *time required (TR)* to perform the demands posed by the job situation, and the *time available (TA)* to actually do it. If the ratio of TR/TA < 1, then spare capacity exists, meaning extra time is available to do other tasks. If TR/TA > 1, then a *workload redline* has been passed where workers now must employ compensations strategies to cope with the insufficient time available. This is also the region where an increase in error-rates typically can be observed (Young, Brookhius, Wickens, & Hancock, 2015, p. 2). This concept is depicted in Figure 5 and is closely interrelated with the Yerkes-Dodson Law (YDL).

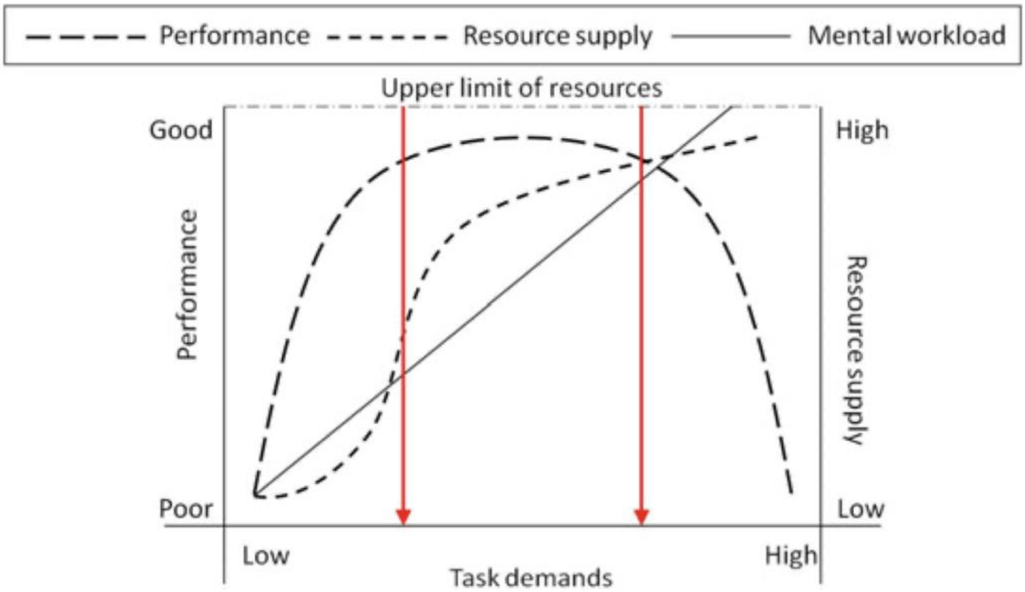

Figure 5: Workload Redline (Young et al., 2015, p. 9).

The YDL from the experiment of Yerkes and Dodson (1908) on the relationship of stimulus to rapidity of habit-formation, can be perceived as one of the closest things to a law of nature in psychology. The analyses of this experiment have historically taken on numerous, and probably too generous, forms of interpretation (Teigen, 1994). However, the general principal seems to have a widely accepted basis, namely that for humans (and other animals) there is a sweet spot when it comes to the stimulus of workload/arousal/stress and subsequent performance. Too little stimulus and the individual gets bored, complacent, and inattentive. Too much stimulus and the individual become stressed, task saturated and prone to error. The concept is illustrated in Figure 6 and is backed up by a wide body of research in the aviation domain (Caldwell, 2005; Salas, Bowers, & Rhodenizer, 1998; Schutte, 2017). One of the better known experiments is Ruffel Smith (1979) simulator study on Boeing 747 crews which indicated that the mean number of errors increased with excessive workload, noting errors like incorrect system operations, navigational mistakes and prolonged response times to aircraft abnormalities. Schutte (2017, p. 234) reaches the same conclusion by his statement *"When humans are in* *periods of high stress and high workload their performance degrades significantly".* Thus, it seems reasonable to assume that workload, as a part of the job situation, also influences the perception of competence, and that aviation organizations should strive to find a good balance between excessive and insufficient workload to optimize performance.

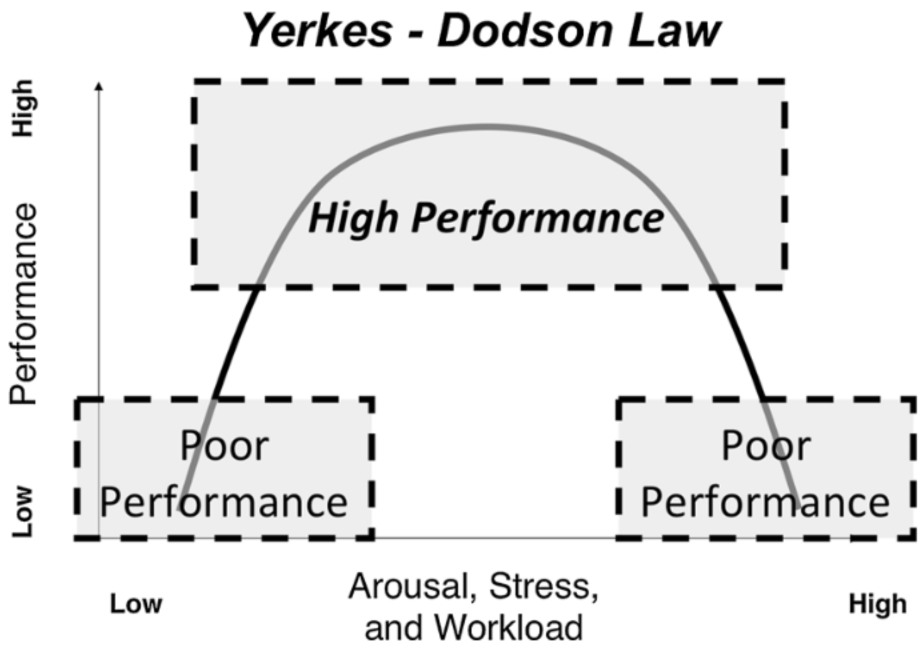

Figure 6: Yerkes-Dodson Law (Schutte, 2017)

Over time, excessive workload is also linked to the phenomena of fatigue and Work-Related Stress (WRS), which unfortunately seems to have become a widespread phenomenon in aviation. An extensive web-based survey of European commercial pilots *(n = 821)* by Cahill, Cullen, Anwer, Wilson, and Gaynor (2021) found that the majority of pilots reported that WRS had a negative impact on their performance. High workload over time was also mentioned as a contributing factor in the near miss of two Norwegian C-130J being seconds away from controlled flight into terrain (CFIT) while conducting tactical low-level formation flying around the Lofoten archipelago, eerily similar to the Kebnekaise-accident with the same squadron eight years prior (Luftforsvaret, 2020; Moldjord, 2023; Rognstrand, 2020). Excessive long-term workload is clearly associated with fatigue, which, according to Helmreich and Foushee (2010, p. 11), can undermine vigilance in otherwise knowledgeable and motivated pilots. The link between excessive workload, fatigue and errors is likely as old as aviation itself with McFarland (1971) summary of the Cambridge Cockpit Studies of WWII Spitfire pilots providing a historical reference:

"Possibly the most significant finding was a general tendency for a sudden increase in errors at the end of the flight. A tired airman, it seems, has an almost irresistible tendency to relax when he nears the airport."

Evidently, the short-term and potential long-term effects of excessive workload negatively affect performance and competency, highlighting the importance of a balanced approach to this issue.

3.2.4 Currency

> *"The first keystone in understanding the human condition is the concept of entropy or disorder" (Pinker, 2018, p. 15)*

The Second Law of Thermodynamics states that *"in an isolated system that does not interact with its environment - entropy never decreases (ibid)."* This might seem out of place in a thesis on military pilots, but it is not. Humans are physical systems, and as such governed by the laws of nature. More specifically, the brain, with its capacity to store information, i.e. *knowledge,* is also a physical system made up of a highly improbable collection of atoms (Tooby, Cosmides, & Barrett, 2003, p. 861). Whenever knowledge is acquired, physical changes take place in the brain (Thompson & Krupa, 1994), and the complex web of neurons and synapses are changed into an even higher degree of functional order (Pinker, 2018, p. 21). Functional orders of atoms are an immensely smaller subset of arrangements than non-functional ones and are as such vastly more improbable. Since the second law of thermodynamics states that physical systems tend to move towards more probable states (Tooby et al., 2003, p. 862), i.e. an increase in entropy, the functional network of neurons in the brain storing information tends to move towards disorder with time. This is at its most fundamental level why currency requirements are an important dimension of having competent pilots, because as physical systems we must always battle against entropy's crusade in creating maximum disorder. If effort and energy is not applied in maintaining skills and knowledge, entropy will make sure that these attributes fade over time. This concept is known as skill decay and refers to the loss of trained or acquired skills (or knowledge) after periods of nonuse (Arthur Jr, Bennett Jr, Stanush, & McNelly, 1998, p. 58). If rarely used, skills and knowledge, like muscle, gradually atrophy in the course of time (Schutte, 2017, p. 238). However, they do not necessarily atrophy at the same *rate.*

Arthur Jr et al. (1998, p. 61) distinguishes between *physical task* that require activities such as muscular strength, exertion of force and coordination, and *cognitive tasks* which involve perceptual input, mental operations, problem solving and decision making. They found that *physical tasks* were less susceptible to skill fade than *cognitive tasks* (ibid, p.85). This is supported by Schendel, Shields, and Katz (1978, p. 24) who found that the retention of motor skills were remembered for months or years, as compared to procedural tasks that were forgotten in days, weeks, or months. Moreover, an extensive literature review was performed by Martinussen and Hunter (2018, pp. 156-163) on the subject, essentially reaching the same conclusion. Unfortunately for pilots, this means bad news as their role has shifted drastically over the past fifty years from *physical tasks* to *cognitive tasks* (Schutte, 2017, p. 235).

Long gone are the days where the pilot did most of the flying manually, relied primarily on visual flight rules and had to carefully manage fuel and engine parameters to keep the motor running smoothly. Today, automation does most of the flying part and the fair-weather stick-and-rudder minded pilots of the past are replaced by something closer to a mission manager (ibid). This represents a dramatic shift in the definition of the role as a pilot. With the flying part now taken care of the tasks allocated to the pilots are now more focused on monitoring the automation, operating the mission systems through advanced software and coordinating with other assets. This is particularly true for the new platforms introduced in the RNoAF such as the F-35 (replacing F-16) and the P-8 (replacing P-3). Harlem (2016) concluded from his job analysis of F-35 pilots that cognitive abilities such as information processing and analysis has seen an increase in importance compared to motor skills with the acquisition of the new airframe. Thus, military aviators are also affected by the changeover from physical to cognitive tasks, suggesting that modern day pilots have a higher demand for refresher training because of faster skill decay. Not only has the flying portion of the job become more cognitively challenging. Military air operations are also inherently dependent on procedural knowledge relying on expert operators to employ potentially strategic assets in a high stress environment with little room for error (Haga & Maaø, 2018, p. 141). At a relatively low experience level and rank, military pilots are put at the tip of the spear in air operations with a disproportional amount of responsibility and authority resting on the outermost link in the operation (the pilot), as compared to other military operations where decision making occurs at higher echelons (ibid, p.144). When added up, these characteristics represents conditions that require a particularly high need for refresher training among military pilots, accentuating why a focus on currency is an important dimension of their perceived competence level.

3.2.5 Complexity

"The state or quality of being intricate or complicated"[8]

Complexity is a defining characteristic of today's high-technology, high-consequence systems (Perrow, 1999) and even the use of the term "complexity" has seen a major rise since the 1950's, likely due to the emergence of ever more advanced technology and complex organizational structures. With it, the aviation industry has followed suit producing some of the most technologically advanced pieces of machinery ever produced by man. The industry has also undergone rapid organizational changes to improve safety and efficiency. Complexity can be separated into two concepts; *hardware complexity* and *software complexity*, inspired by the distinction made in the SHEL-model developed by Edwards in 1972 and later adopted by the ICAO as a conceptual tool to capture a systems complexity (Mesarosova, 2020, pp. 239-240). *Hardware* in this context refers to the equipment, i.e. the aircraft, being operated by the *liveware* (the pilots), while *software* refers to the rules, procedures, regulations, and organizational structures that govern the operation.

3.2.5.1 Hardware Complexity

The technological prowess, intellectual capital and financial assets allocated to the F-35 program alone is sufficient to boggle the mind, with an estimated lifetime cost now surpassing $2 trillion (Marrow, 2024). Military pilots are at the receiving end of this breakneck technological arms race and are expected to understand and operate increasingly more complex aircraft and equipment. Comparing a fighter aircraft from the WWII era to that of a fifth-generation fighter like the F-35 today, the difference in complexity is massive. While the Spitfire was equipped with a compass, basic flight instruments and a gun, all components that are relatively easy to understand – pilots flying modern fighter jets like the F-35 needs to understand the same fundamental principles of flight like the Spitfire-pilots, but also the workings of stealth, electro-optical targeting systems, AESA[9] radar, distributed aperture system, countermeasure dispensing systems, fly-by-wire and advanced flight control laws, all-weather operations, flight management systems, data link networks, a range of air-to-air, air-to-surface and anti-ship missiles, precision guided bombs and so the list goes continues. This dramatic increase in hardware complexity is not unique for fighters and follows virtually all

[8] Definition from Oxford Languages

[9] Active Electronically Scanned Array

military aircraft. As another example, the P-8 flight manual is now a 1500-page behemoth filled to the brim with notes, warnings and cautions, operating procedures and systems descriptions expected to be remembered by the pilots. This overwhelming amount of information demands a high level of cognitive effort to comprehend.

Aside for new military capabilities enabled by advances in technology, another driver of increased hardware complexity is the hardware changes enforced through responses to aircraft incidents and accidents. There has historically been a Newtonian approach to accident investigation in which it is presupposed that by collecting enough evidence one can get to the root cause of the accident, and as such create redundancies or barriers that will prevent such incidents from happening in the future (Dekker, Cilliers, & Hofmeyr, 2011). Cases like the crash of Eastern Airlines 401 in the Florida Everglades provides an example of this mindset where the accident led to the requirement of aural alarms when the autopilot goes into control wheel steering mode (FAA, 2022), adding another alerting system in the cockpit aimed to improve safety. However, Perrow (1999) points out that this Newtonian focus may create a false belief that redundancy is the best protection against risk, which in this example can lead to alarm fatigue if too many alerting systems are placed in the cockpit. Another downside is that barriers and redundant systems all add to a systems complexity, which in itself create new categories of accidents. Dekker (2002, p. 8) reaches the same kind of conclusion with his statement:

> *"… aerospace has seen the introduction of more technology as an illusory antidote to the plague of human error. Instead of reducing human error, technology changed it, aggravated the consequences, and delayed opportunities for error detection and recovery."*

The mindset of adding barriers to eliminate risk is likely inspired by accident models such as the Domino model and the infamous Swiss Cheese Model from Reason (1997), Figure 7. This model has probably been a belabored subject in most aviation safety communities in attempts to visualize the trajectory of accidents. The model illustrate how hazards can potentially penetrate through holes in the safety barriers due to active failures and latent conditions. Implicitly the model suggests that adding more defensive layers, i.e. redundancies and safeguards, to the system will increase safety and prevent accidents. However, as Box and Draper (1987, p. 424) succinctly put it: *"All models are wrong, some are useful"*, referring to the trade-off between a models complexity, level of tolerable error and ease of use. Clearly, Reason's model has been useful, all though not a complete representation of reality, but perhaps it has been given a too literal interpretation with a disproportionate focus on adding safety barriers in aviation. As Perrow (1999) argues: *"…this conventional engineering approach to ensuring safety fails because high system complexity makes failures inevitable."* Adding barriers, he claims, means adding to complexity, which in turn only creates new categories of accidents, using the meltdown at Chernobyl as an example where a leading cause of the accident was a trial of a new safety system (ibid).

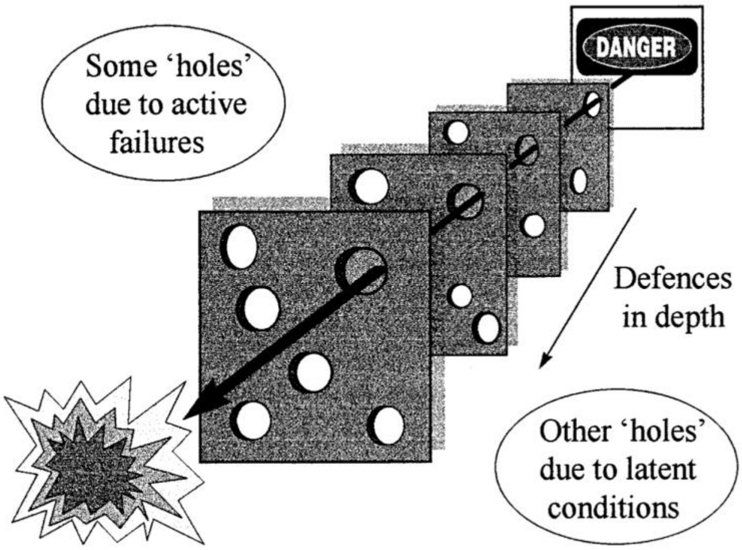

Figure 7: Swiss Cheese Model (Reason, 1997, p. 12)

3.2.5.2 Software Complexity

"*NATOPS is written in blood"*[10]- is common navy slang and a term heard throughout US Navy *ab initio* pilot training. It refers to the iterative changes that occur throughout the lifetime of an aircraft platform, as lessons are drawn from incidents and procedural changes are implemented. It also implies that the volume of information contained in the manual steadily grows resulting in an increase in software complexity. This is not unique to flight manuals but permeate throughout the landscape of aviation and makes its way into standard operating procedures, rules, and regulations. As advances in technology has led to a reduction in "technical failures" the major factor in aviation accidents is now cited as human error (Mesarosova, 2020, p. 239), and must be guarded against by procedural changes, regulations and training programs. The

[10] Naval Air Training and Operating Procedures Standardization refer to the flight manuals used in the USN.

mindset that humans were the main culprit to accidents brought with it the introduction of Crew Resource Management (CRM).

CRM can be traced back to a conference held by NASA and key industry players discussing the problem of how to fully utilize the human resources available on the flight deck (Cooper, White, & Lauber, 1980). This marks a point in time where the aviation industry shifted its attention away from a unilateral focus on technical skills to more broadly emphasize both hard, technical skills and soft, non-technical skills like leadership, communication, and human factors. The aim of the conference was to address what had been identified as the main culprit of aviation accidents lately, namely pilot error in the form of poor interpersonal communications, decision making and leadership (Helmreich, Merritt, & Wilhelm, 1999, p. 19). The introduction of CRM in aviation is usually presented as one of the industry's biggest success stories, linking the advent of CRM with the drastic reduction in the aircraft accident rates occurring after the implementation. As such, education in non-technical "CRM-skills" has been adopted as an integral part of *ab initio* pilot training and constitutes a sizeable component of annual refresher training. However, training in CRM and other soft skills presupposes competency in all the technical KSAO's required to operate an aircraft (Martinussen & Hunter, 2018, p. 171) and brings a non-trivial opportunity cost with it. There are also questions regarding the causality between CRM-training and improved aviation safety. Salas, Burke, Bowers, and Wilson (2001) in their review of 58 published accounts of CRM-training could not establish whether it had an impact on flight safety. Thus, not everyone is fully onboard with the success story of CRM with Moriarty (2015, p. 269) claiming that CRM-training is often hollow, meaningless and a waste of time. If this is indeed the case, training in non-technical skills might lead to a sense of frustration as time is a scarce resource for pilots and this type of training must complement, not displace technical training. Ultimately, this might lead to a reduction in the TR/TA ratio and a subsequent impact on competency and performance.

3.3 Summary

From the introduction and the preceding paragraphs in this chapter it is appropriate to summarize the points made thus far and revisit the situation-specific model for human performance, Figure 8. From this model *feedback* is received by the individual following a *response* from the individual and the subsequent *consequence* to a certain situation. *Experience* is gained when the cycle is repeated, informing the individual whether the response was correct or not. *Currency* also plays a role, due to the ever-looming threat of skill decay, determining if the individual remembers the correct response or not. *Workload* determines how much time and cognitive capacity is available to learn, conduct, and repeat the cycle. Finally, *complexity* regulate the difficulty in achieving the correct response, and at which *rate* the skill decay occurs. Thus, the underlying attributes of the job situation and the individual, i.e. the *input of competence*, regulate the correctness of the response, the *output of competence*, also called performance. As such, these factors all affect the *feedback* coming back to the operator. If the *feedback* is predominantly positive, then it is reasonable to assume that this individual feel more competent than someone receiving mostly negative feedback. The following chapters will examine what role these four predictors play in regulating the *feedback* received by the individual, thus also regulating the perception of competence.

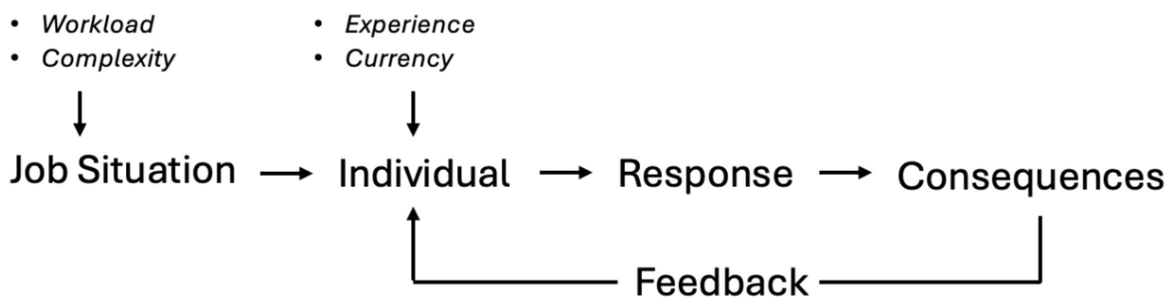

Figure 8: Situation-specific model of human performance. Adapted from (Hoffmann, 1999, p. 277)

4 Methodology

"If the design of an experiment is faulty, any method of interpretation which makes it out to be decisive must be faulty too." (Fisher, 1966, p. 3)

The scientific method involves following a structured process based on rules and principles towards a specified research goal. The path towards this goal is called methodology and is according to Jacobsen (2015, p. 14) what separates science from commonplace observations and claims. The goal of this systematic process of collecting and analyzing data about real world observations is to produce valid and reliable results which accurately describe and explain reality, and possibly predict future outcomes (Tranøy, 2019).

This chapter will present, explain, and justify the methodology chosen for the study. This involves explaining the research design, selection of participants, data collection methods and how the data was analyzed.

4.1 Research Design

The empirical research conducted in this study is based upon the quantitative method with the goal of describing the relationship between the predictor variables *experience*, *complexity*, *workload* and *currency* and the associated outcome variable *perception of competence* to answer the primary research questions presented in paragraph 1.3.1.

Primary data from military pilots at three operational squadrons in the Norwegian Air Force were collected using a questionnaire distributed by email. The survey was open for two months, January and February 2024. Once the data was collected the results were interpreted using statistical techniques, primarily the multivariable technique of regression analysis, to address the primary research question. Likert data and a bivariate correlation matrix was also used to supplement the results and to address the secondary research question.

The quantitative method was chosen with the intent of reaching a larger number of respondents than usually feasible with a qualitative approach within the scope of a master's thesis. An additional reason was the ability to remain anonymous, hopefully to invoke more honest responses from the participants. The aim was to obtain a representative cross-section of a squadron's composition from the *ab initio* co-pilot to the Commanding Officer, and give them the ability to speak freely, without risk of retribution from the organization. Therefore, the survey was made completely anonymous. Furthermore, surveys distributed by email gives the participants the option to complete at their own leisure and is an efficient and flexible way of collecting data from otherwise busy personnel.

A vital step in choosing how to design a research project for testing the theories and hypothesis within is to decide upon (1) what to measure and (2) how to measure it (Field et al., 2012, p. 7).

4.1.1 Variables

> *"Reliable and valid surveys are obtained by making sure the definitions and models you use to select questions are grounded in theory or experience." (Fink, 2009, p. 8)*

The variables and succeeding survey questions were obtained and grounded through both theory and experience. Following a literature review and observations drawn from personal experience a deductive approach was taken in deciding upon what to measure. The resulting hypothesis was that the four predictor variables chosen was some of the main drivers behind variation in *perception of competence* among military pilots. Furthermore, the hypothesis assumes that there is a strong correlation between the *perception of competence* and actual performance.

The intent of measuring the characteristics of primary interest, namely *experience, complexity, workload,* and *currency,* was mainly to describe their relationship with the outcome variable, i.e. a focus on *prediction*. *Inference* as to why these phenomena occur is not the focus of this thesis and are grounds for further research.

4.1.2 Data Collection

To measure the four predictor variables and the outcome variable a self-administered one-time questionnaire was designed. Each variable was attributed questions with the intent of gauging the construct of interest. For some of the variables raw continuous data could be collected (e.g., flight hours), but for most of the variables, data on a lower level of measurement using a Likert-scale had to be utilized. The number of questions attributed to each variable ranged from four (*perception of competence)* to eight (*currency)*.

The population of the study is all active military pilots presently flying for the Norwegian Air Force. Exact numbers of how many pilots are flying at any given time cannot be referenced due to classifications issues, but Grindheim (2023, p. 23) reports that over a 15-year period 212 flight students made it through the selection process at Luftforsvarets Flygeskole (LFS). This gives a rough estimate that the population at hand is somewhere around that number, as 15-20 years is about the average time a military pilot spends actively flying. The RNoAF consists of nine flying squadrons, five of which fly fixed wing aircraft. To reduce complexity data were only collected from fixed wing squadrons based on increased familiarity with their structure from personal experience and due to ease and likelihood of acquiring the necessary permission required to be obtained from the respective unit commanders prior to sending the formal request to Forsvarets Høgskole (FHS).

The surveys were distributed to the 331 (F-35), 333 (P-8) and 335 (C-130) Squadron and all pilots currently employed there were provided with the option to participate. These squadrons employ three very different aircraft, each with their own unique missions sets, providing variation in the predictor variables and possibly the outcome variable. The primary reason to distribute the surveys among several squadrons was to increase the number of respondents (i.e., increase *n*) and to assess data over a broader part of the Air Force to describe tendencies across the whole organization, not just conditions at a single squadron.

4.1.3 Design of Questionnaire

According to Field (2012) a good questionnaire must contain three things – discrimination, reliability and validity. Discrimination was sought to be obtained by a large scale, encouraging variability among the respondents and hopefully resulting in divergent scores in respondents with differing perceptions for the various constructs in question. Secondly, validity, the construct of measuring what you think you are measuring, was sought to be obtained by using a language familiar to the response group. The author has insider knowledge of the military pilot lingo, and as such questions were attempted formulated using familiar and standardized terms that reduces the risk of confusion and misunderstanding. Reverse-phrased questions were included for the variables *workload (WL4)* and *currency (CUR8)* to reduce response bias (Field et al., 2012, p. 799). Finally, a reliability assessment of the questionnaire, i.e. the ability of the survey to produce the same results under the same conditions, through the test-retest method, was not conducted due to the anonymous nature of the survey and time constraints.

The questionnaires were designed by using the University of Oslo's application *nettskjema.no*, a powerful and secure data capturing tool commonly used in academic research in Norway (University of Oslo, 2024). The construct of the questionnaire was designed to measure the four characteristics of interest, in addition to the outcome variable *perception of competence.* A total of 34 questions was added to the questionnaire with the intent of measuring the underlying variables (or factors) of *experience, complexity, workload* and *currency*. The questions consisted of a combination of data fields for continuous variables, interval variables, categorical data and text boxes where respondents could add amplifying information in free text format. A few comments obtained from the text boxes will be cited to augment data in chapter 0, however, no further qualitative analyses were conducted. Numerical values could be inserted for the continuous variables such as *"How many years have you been flying for the RNoAF?"* The primary format for the discrete variables was a 10-point Likert scale ranging from 1 to 10 where (1) usually corresponded to wordings like *"strongly disagree"* or *"least complex"* and (10) corresponded to *"strongly agree"* or *"most complex"*. With a 10-point scale there is no neutral midpoint, which according to Field et al. (2012) forces respondents to either side of the scale and can create noise in the data. However, this effect was attempted mitigated by having a relatively large scale. All the questions were coded for easy reference during the data analysis phase. The codes can be seen in Appendix B, Figure 15, as a code and number in italics at the end of each question, e.g. *(CX3)*, meaning question number three under the factor *complexity*.

The selection of statements followed the *Method Of Constructing An Attitude Scale* by Rensis Likert (1932). Each item was carefully worded as to not form leading or ambiguous questions in a clear and concise language. An iterative process between the researcher and the assigned supervisor was conducted before arriving at the final version of the questionnaire.

4.2 Procedure

The use of primary data obtained from the Norwegian Defense Sector is regulated by a research commission at FHS. Applications are handled on an individual basis and must meet all the requirements set forth by the commission and are described in detail on their website (Forsvarets Høgskole, 2024). Final approval from FHS was granted 21st of November 2023. Since the questionnaires were anonymized, the research project was not required to report a notification form for personal data to Sikt.

The surveys were distributed to each respective unit commander using commercial email applications. Subsequently, the unit commanders promulgated the survey to all pilots within their respective squadron via the internal email system FIS-BASIS, a classified system only accessible on government computers. As such, the unit commanders had to be used as middlemen for distribution. This was a result of the author not physically being in Norway during the data collection phase. The emails sent out to the unit commanders contained a project description, an informational letter, and the online questionnaire.

Due to informational security restrictions regarding F-35 pilots, some questions had to be dropped for the questionnaire sent to 331 Squadron to keep the thesis on an unclassified level. The following questions had to be dropped: EXP2-3, CX2 and CUR3-5. This had a minor, but not insignificant, effect for the data analysis part and compensation strategies to fill in the missing data fields are discussed later.

4.3 Analysis

RStudio, version 4.3.3, were used for all statistical analysis and data visualization used in this thesis. The significance level (α) was set to .05, meaning any p-values below this value were regarded as statistically significant. 5% or .05 was also used as the upper tail for tests using the Fisher Distribution. Linear association is denoted as Pearson's correlation coefficient (r) and the strength of the linear relationship is denoted as R^2.

4.3.1 Data Reduction

> *"The object of the reduction of data is to exclude irrelevant information, and to isolate the whole of the relevant information contained in the data."*
>
> *(Fisher, 1934, p. 7)*

The goal of the data reduction process was to achieve parsimony by explaining the maximum amount of common variance by the smallest number of explanatory constructs. From the initial 34 items, four were omitted as they were deemed outside of the scope of the thesis. Of the remaining 30 items all were sorted into one of four predictor variables or into the factor *perception of competence* for the outcome variable. For the factor *experience* the raw data was inserted directly into the regression analysis without calculating a *score*, explained later in this chapter. For the rest of the factors a *score* was calculated by combining several questions into one single score for each factor, as described by Likert (1932, p. 42). A degree of subjectivity was introduced during this stage of the process as each variable had to be assigned a weight in the calculations. Generally, data of the same measurement level (i.e. continuous or interval) were given the same arithmetic weighting in the formulas. However, in some of the scores a combination of continuous and interval data was combined where equal weighting was not possible, contributing to a moderate amount of subjectivity in the analysis and represents a weakness in the study.

4.3.2 Multiple Regression Analysis

Strategy for selecting the best regression model followed the backward-elimination process described by Kleinbaum, Kupper, Rosenberg, and Nizam (2014, pp. 447-449). This procedure follows a strategy with a focus on prediction. As such, the goal is to find a best-fitting regression model to describe the relationship between the predictor variables and the outcome variable by finding the best subset of *k* predictors. The variables kept in the model are only retained due to their predictive power, no effort is made to massage the model into containing specific variables.

Criterions used for selecting the final model was primarily R^2, F-statistic, p-value and Mallow's C_p. *Ceteris paribus*, lower values of C_p are preferred, and the correct model – or larger models that contain the correct model – should have a value close to $(p + 1)$ or lower.[11] A holistic approach was taken in selecting the final model, thus no single criteria was used as the sole measure for excluding variables (Kleinbaum et al., 2014, pp. 444, 447).

Neither higher order powers of the basic predictors, interactions among predictors nor other transformations of the predictor variables were assessed, meaning only the basic predictors (*EXP, CUR, CX and WL)* were included in the maximum model. The aim was to start with a small maximum model to follow the principle of Occam's Razor to achieve a parsimonious final model (Kleinbaum et al., 2014, p. 441; Martinussen & Hunter, 2018, p. 5)

The test statistic for accepting or rejecting the null hypothesis for the models obtained followed the general form described below:

$$H_0: \beta_1 = \beta_2 = \ldots = \beta_k = 0$$

$$H_A: \text{At least one of the slope parameters are not 0}$$

4.3.2.1 Standardization

For the multiple regression analysis both the predictor variables and outcome variable were standardized. This was performed since some of the factors utilized a combination of both

---

[11] *p* = number of predictor variables in the candidate model

interval and continuous data for calculating the respective *variable score* used in the regression models. All variables were standardized according to the Z-transformation method as follows:

$$Z = \frac{X_i - \bar{X}}{S}$$

This results in the regression coefficient of the intercept (*β_0*) essentially going to zero and the remaining coefficients assuming a value between -1 and +1 (Field et al., 2012, p. 209). The resulting generalized form of the *standardized multiple regression model* is presented below (Kleinbaum et al., 2014, p. 366):

$$\frac{Y_i - \bar{Y}}{S_Y} = \beta_1 \frac{(X_{i1} - \bar{X}_1)}{S_1} + \frac{(X_{i2} - \bar{X}_2)}{S_2} + \cdots + \frac{(X_{ik} - \bar{X}_k)}{S_k}$$

Standardizing the coefficients in the regression model also has the beneficial side effect of easing interpretation of the data as one unit increase in each variable is equal to its standard deviation. Thus, a typical unit increase in the standardized coefficients express how each fraction of the predictor variables influences the outcome variable, without prior knowledge of the scale used (Taboga, 2021). The Z-transformations were performed using Microsoft Excel.

4.3.2.2 Factor Analysis

Factor Analysis (FA) was conducted on the Likert scale response data to exclude irrelevant information from the dataset while at same time retaining as much of the original information as possible. A *maximum-likelihood* FA using four factors was performed to evaluate whether each question pertaining to the respective factor was *actually* measuring what was supposed to be measured. The four factors containing Likert scale data (*workload, complexity, currency and perception of competence)* were already predetermined based on experience and theory (Field et al., 2012, p. 759; Martinussen & Hunter, 2018, p. 35). The FA were performed using a *Varimax* rotation in RStudio. Items that did not load onto their respective factors were omitted from the regression analysis and are described in detail under each variable paragraph.

4.3.2.3 Variables

A visual depiction of the maximum model for the regression analysis is provided in Figure 9. Predictor variables are on the left and the outcome variable is on the right side.

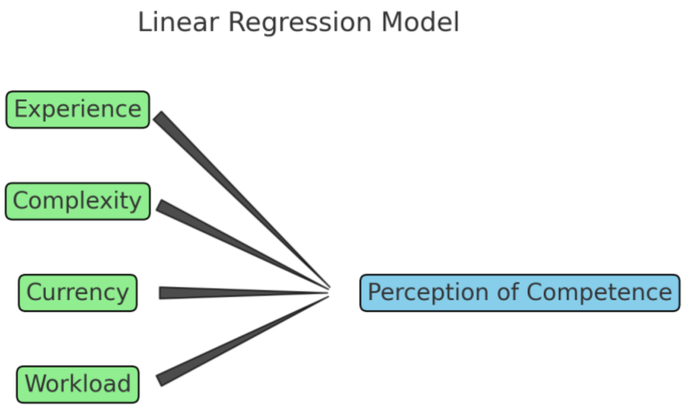

Figure 9: Maximum model

4.3.2.3.1 Experience

Data for the factor of experience was gathered stating three separate questions (EXP1-3). This acquired data about total amount of military flight hours (EXP2), number of years as a pilot (EXP1) and the yearly average flight hours (EXP3). The question of total and yearly flight hours (EXP2 & 3) was omitted for the respondents at 331 squadron due to information assurance (IA). The mean yearly flight hours in the sample from 333 and 335 Squadron were 255 hours. This mean was attributed to each observation from 331 squadron based on years of service to compensate for the missing data fields. E.g. a pilot from 331 with eight years of service was attributed $8 \times 255 = 2040$ flight hours. This gives all the observations from 331 squadron an artificial level of experience but was a necessary work around to be able to run the regression analysis using the complete dataset.

4.3.2.3.2 Complexity

Complexity was measured by collecting information about each pilot's qualification level (e.g. wingman vs formation lead) and through four Likert scale questions (CX3-6). For the Likert scale items the factor analysis indicated that item CX4 ("*how complex is your job as a military aviator?*") did not load onto the complexity factor and was subsequently omitted from the regression analysis. For the rest of the data points, a *complexity score* was calculated and standardized in the following way:

- 1 point attributed to each qualification level above *co-pilot/wingman* (e.g. a qualification level of *aircraft commander* and *instructor pilot* equals 2 points)
    - Qualification data was subsequently standardized using the Z-transformation and given a qualification score.
- A mean score for the variables CX3,5 and 6 were calculated and standardized.
- A total complexity score was calculated by adding both scores and dividing by $n = 2$.

4.3.2.3.3 Currency

Currency was measured by collecting information about last time each pilot was a full-time flight student, with the assumption that more years since being a flight student had a negative impact on currency. Total amount of flight hours in 2023 was also obtained and attributed a positive impact on the *currency score*. Finally, data from three Likert scale questions (CUR6-8) were gathered. All items loaded onto the factor *currency* and was subsequently utilized to calculate a *currency score* in the same fashion as for complexity.

Again, due to IA the participants from 331 Sqn could not report the number of flight hours for 2023, and so an average value based on data from the remaining respondents were calculated and attributed to the respondents from 331 Sqn.

4.3.2.3.4 Workload

Workload was measured solely by Likert scale questions, with a total of seven questions. The factor analysis revealed that only three of the questions (WL3, WL4 and WL7) loaded onto the factor *workload.* The remaining data points were omitted from the regression analysis. An average of the three remaining items was subsequently calculated and standardized before being used in the regression analysis.

4.3.2.3.5 Perception of Competence

Perception of competence was measured by acquiring data from four separate Likert scale questions, aiming to measure both dimensions (input and output) of the concept of competence. Two items (COMP1 & 2) were dedicated to measuring the input of competence, i.e. *competency*, while two items (PERF1 & 2) examined the output of competence, i.e. *performance.* The sum of the data points were divided by $n = 4$ to calculate a mean *competence score*. All questions loaded onto the factor *perception of competence.*

4.3.2.4 Reliability of Model

Scale reliability was calculated for the predictor and outcome variables using Cronbach's α, except for the variable *experience* which did not contain any Likert scale items. Each variable was calculated separately using only the items retained from the factor analysis. Reverse phrased items were attributed a negative score (Field et al., 2012, p. 806). The variables of *competence* ($\alpha = .89$) and *workload* ($\alpha = .88$) indicate high reliability. However, the variables of *currency* ($\alpha = .60$) and *complexity* ($\alpha = .75$) displayed lower reliabilities and may explain their lack of predictive power in the regression model. Notwithstanding, the constructs containing Likert scale items kept in the final model (*competence* and *workload*) display relatively high values of Cronbach’s α - indicating a high overall reliability of the model used as a basis for addressing the primary research question.

Collinearity was assessed by computing the direct correlation between the two remaining predictor variables. Close to zero correlation was observed, indicating no effect from collinearity in the model.

To exhaustively verify the reliability of the model it is necessary to test whether it predicts well for a new random sample drawn from the population under study (Kleinbaum et al., 2014, p. 438). No such efforts have been made.

4.3.3 T-test

A paired two sample t-test were computed between the variables COMP1/2, PERF1/2 and PERF2/3 to assess the existence of a statistically significant difference in the means between the items to address the secondary research questions. The hypothesis for the test statistic was as follows:

$$H_0\colon \mu_1 = \mu_2 \qquad H_A\colon \mu_1 \neq \mu_2$$

5 Results

This chapter will present data obtained from the survey and the resulting products from the ensuing data analysis. First, unaltered observations are presented to visualize the data collected, free from manipulation. Secondly, the regression analysis will be performed in a stepwise fashion, leading to the final regression model – which provides the basis for answering the primary research question. A bivariate correlation matrix is included in Appendix A, Figure 14, to display how the individual items correlate with each other.

5.1 Survey Data

5.1.1 Response Rate and Experience

The following tables contain data about the response rate from the survey and the respondents experience level. No data for flight hours were obtained from 331 Sqn due to IA.

Table 1: Experience in Years - EXP1_yrs (n = 16)

Sqn	N	Mean	SD	Min	Max
331	5	6	4	2	10
333	7	19	12	1	30
335	4	14	11	1	25

Table 2: Experience in Flight Hours - EXP2_hrs (n = 16)

Sqn	N	Mean	SD	Min	Max
331	5	*NA*	*NA*	*NA*	*NA*
333	7	3688	2465	530	7000
335	4	3527	3284	309	7900

5.1.2 Likert Scale Data

The following plots present data from the Likert Scale items contained in the survey. Center of the plots are set to the median value of the scale (5.5). The labels referenced on the plots corresponds to the coding used in the questionnaire contained in Appendix B, Figure 15.

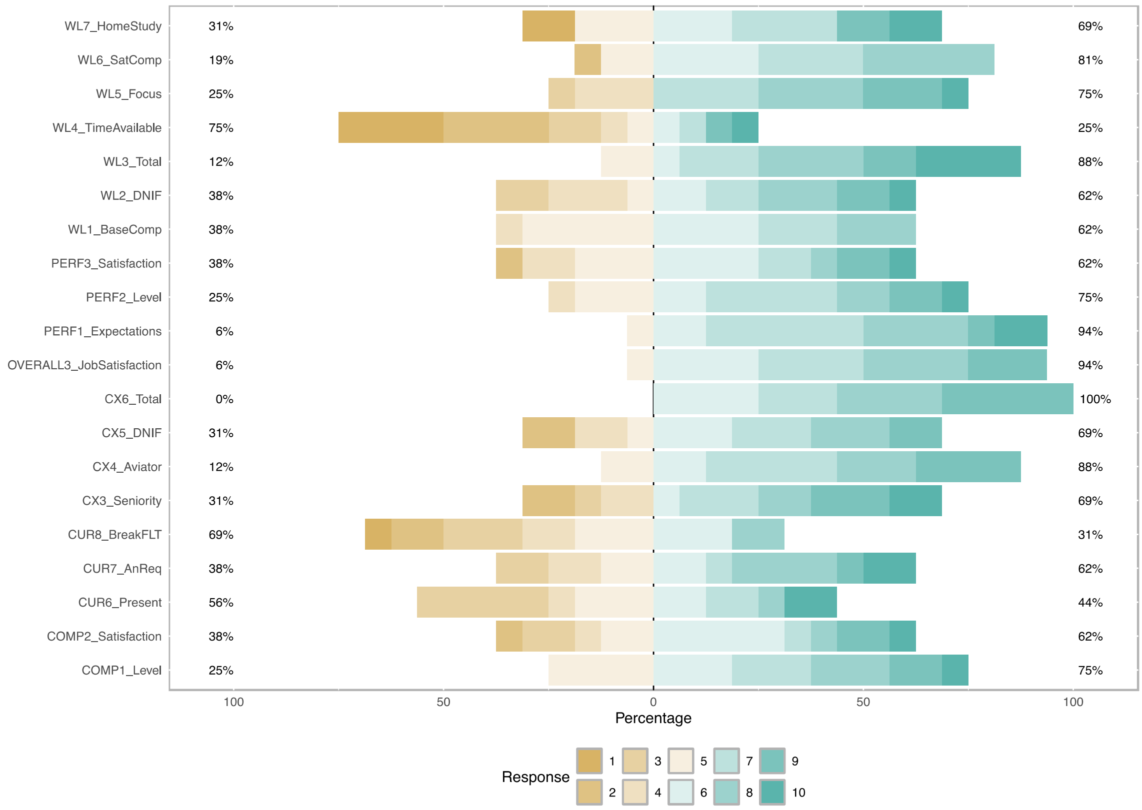

Figure 10: Likert data 1, w/o grouping

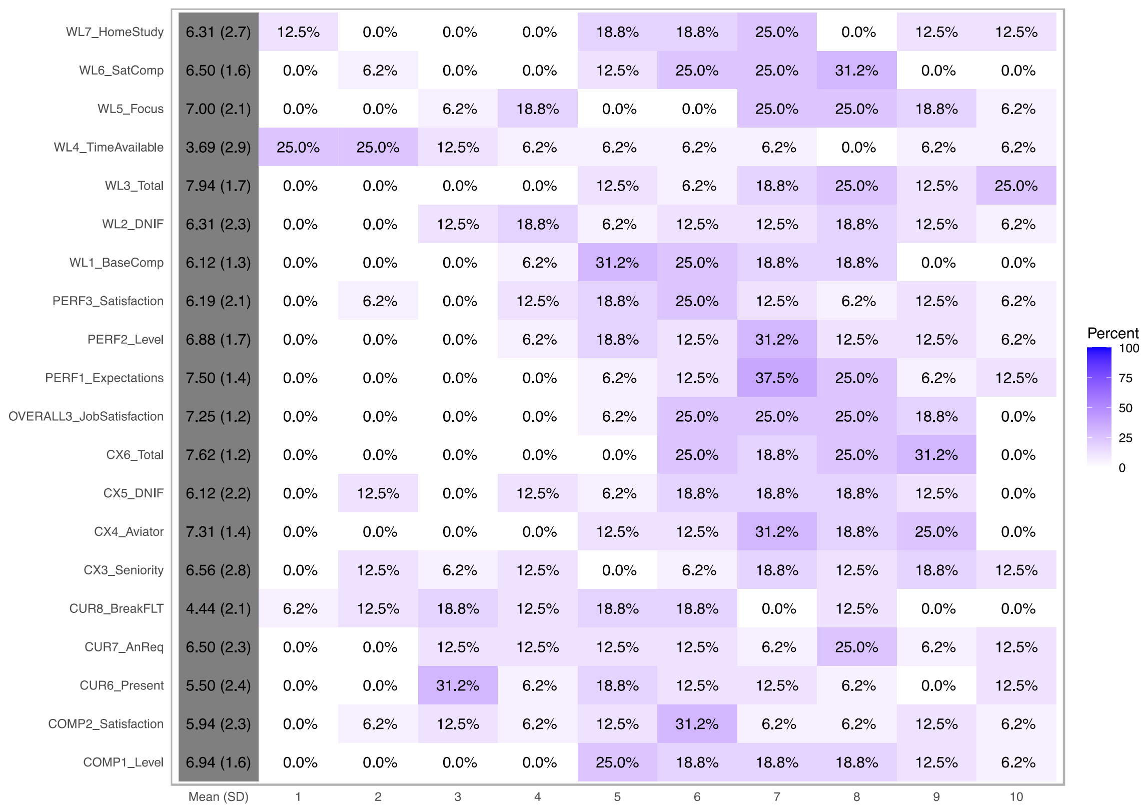

Figure 11: Heat Map

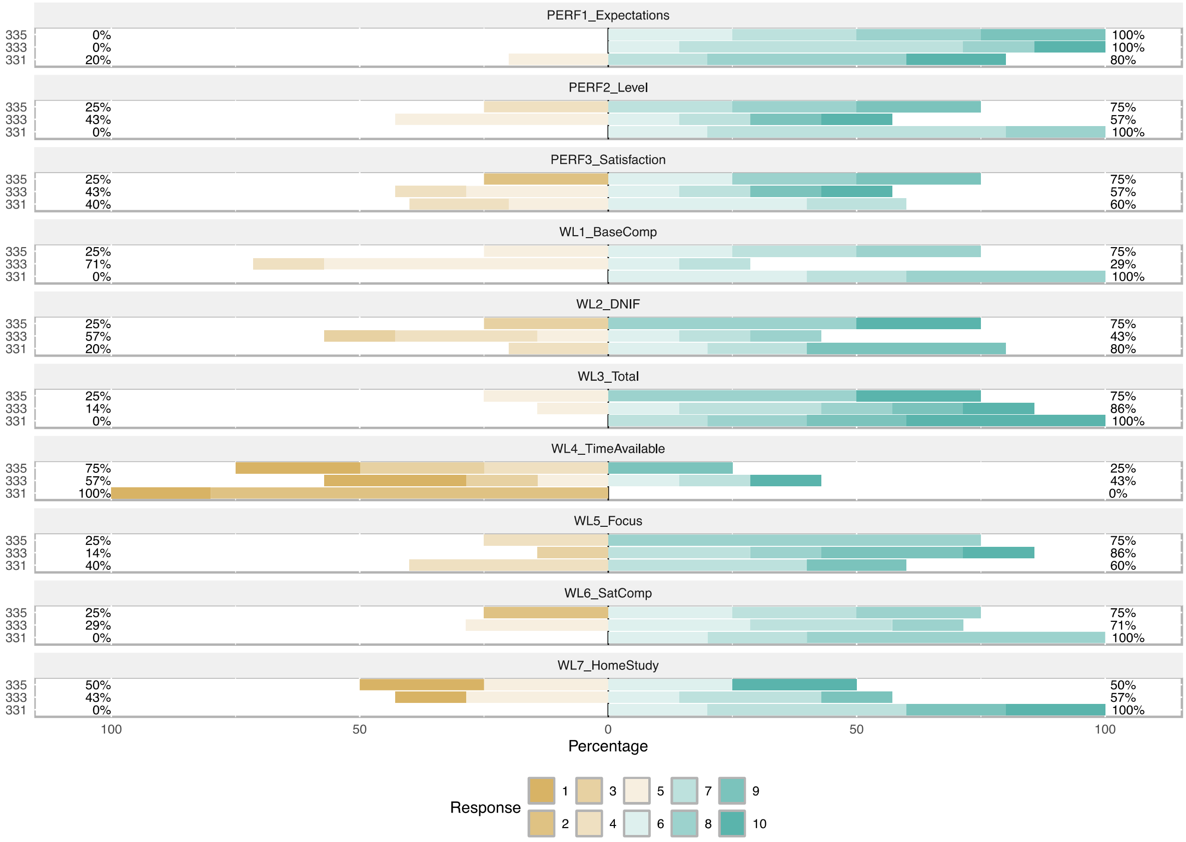

Figure 12: Likert data 2, grouped by SQN

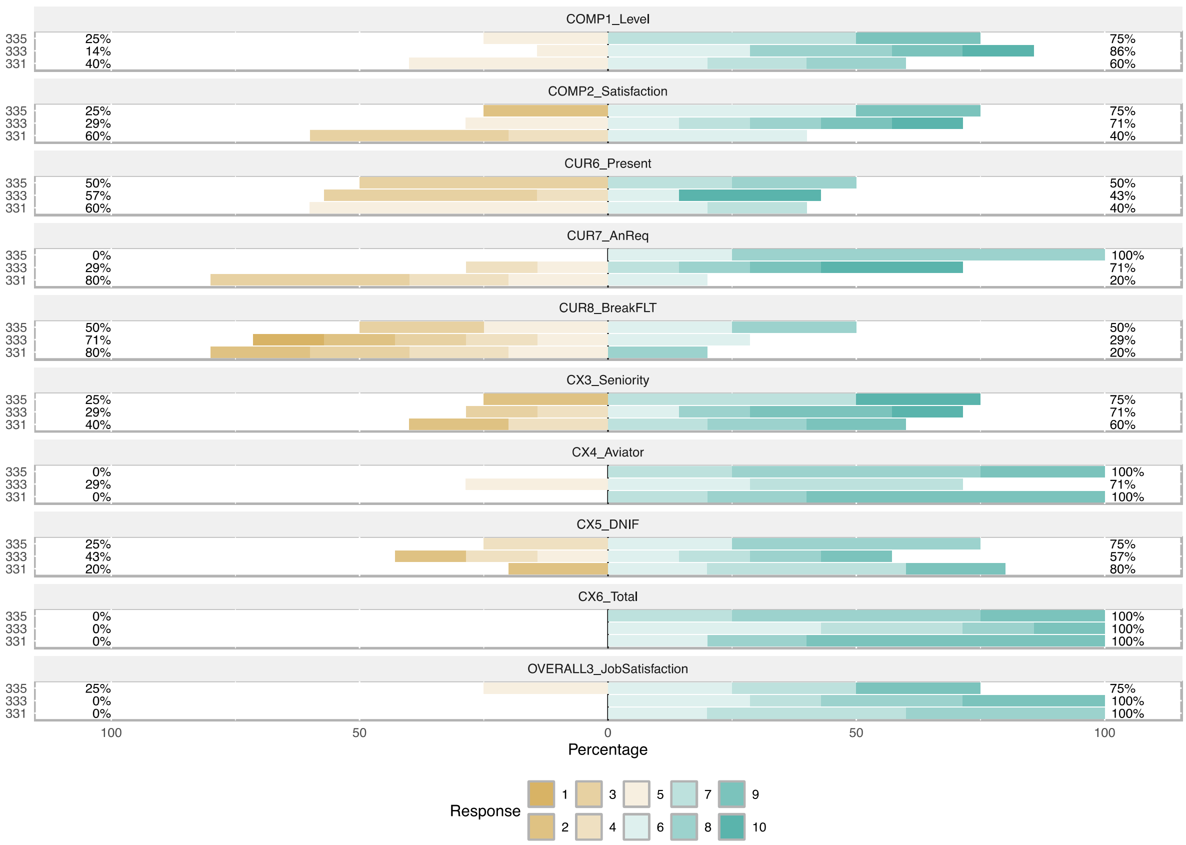

Figure 13: Likert data 3, grouped by SQN

5.2 Multiple Regression Analysis

For the regression analysis a summary of the resulting regression model and an *Analysis of Variance (ANOVA)* table are presented for each model. The criterions discussed in paragraph 4.3.2 provides the means for removing predictor variables, one at a time, until a final model is presented. The resulting model is the principal basis for answering the primary research question. Significance codes of p-values are denoted as: 0 ≤ '***' < 0.001 < '**' < 0.01 < '*' < 0.05.

5.2.1 Backward-Elimination Procedure

In the following procedure focus will be on the maximum and final model. The steps in between are kept for context but contains less information. The linear regression models were fitted (estimated using Ordinary Least Squares) to predict *Competence* with *years of experience (EXP1_yrs)*, *flight hours (EXP2_hrs)*, *Workload (WL)*, *Currency (CUR)* and *Complexity (CX)* (formula: Comp ~ exp1_yrs + exp2_hrs + wl_score + cur_score + cx_score). Due to standardization the intercept coefficient β_0 is essentially zero and is not presented in the fitted regression equations.

5.2.1.1 Step 0

In *Step 0* the maximum model is presented. The model contains all the empirical data collected in the study except for the questions removed in the factor analysis and the items *PERF1* and *OA1-3*.

Table 3: Summary Maximum Model

Variable	Estimate	Standard Error	t value	Pr (>|t|)	
EXP1_yrs	.094	.294	0.319	.756	
EXP2_hrs	.427	.293	1.458	.175	
WL	-.457	.158	-2.882	.016	*
CUR	.250	.234	1.069	.310	
CX	.103	.236	0.438	.671	

Adjusted R-squared: 0.606

F-statistic: 5.62 on 5 and 10 DF, p-value: 0.01009

Table 4: ANOVA Table Maximum Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr (>F)	
EXP1_yrs	1	3.29	3.29	10.66	.008	**
EXP2_hrs	1	1.32	1.32	4.29	.065	
WL	1	3.67	3.67	11.90	.006	**
CUR	1	0.32	0.32	1.06	.327	
CX	1	0.05	0.05	0.19	.670	
Residuals	10	3.08	0.31			

The model explains a statistically *significant* and *substantial* proportion of variance *(R^2 = .74, F (5, 10) = 5.62*)* and the null hypothesis is rejected. However, the effect of CX is statistically *non-significant (p = .671)*, explains the least amount of variance, has a low estimated regression coefficient and a poor F-statistic. The variable CX is consequently dropped from the model.

5.2.1.2 Step 1

From *step 0* variable CX is removed and a new model with fewer variables is presented.

Table 5: Summary Step 1

Variable	Estimate	Standard Error	t value	Pr (>|t|)	
EXP1_yrs	.120	.276	0.436	.671	
EXP2_hrs	.455	.275	1.655	.126	
WL	-.446	.150	-2.959	.013	*
CUR	.240	.224	1.070	.307	

Table 6: ANOVA Table Step 1

Source	DF	Sum of Squares	Mean Square	F Value	Pr (>F)	
EXP1_yrs	1	3.29	3.29	10.66	.006	**
EXP2_hrs	1	1.32	1.32	4.29	.054	
WL	1	3.67	3.67	11.90	.004	**
CUR	1	0.32	0.32	1.06	.307	
Residuals	11	3.14	0.28			

The new model explains a statistically *significant* and *substantial* proportion of variance *(R^2 = .73, F (4, 11) = 7.53**)* and the null hypothesis is rejected. However, the effect of CUR is statistically *non-significant (p = .307),* explains a low degree of variance and has low F-value. The variable CUR is therefore dropped.

5.2.1.3 Step 2

From *step 1* variable CUR is removed and a new model with fewer variables is presented.

Table 7: Summary Step 2

Variable	Estimate	Standard Error	t value	Pr (>|t|)	
EXP1_yrs	.156	.276	0.565	.582	
EXP2_hrs	.420	.274	1.528	.152	
WL	-.504	.141	-3.563	.003	**

Table 8: ANOVA Table Step 2

Source	DF	Sum of Squares	Mean Square	F Value	Pr (>F)	
EXP1_yrs	1	3.29	3.29	11.36	.005	**
EXP2_hrs	1	1.32	1.32	4.57	.054	
WL	1	3.67	3.67	12.69	.003	**
Residuals	12	3.47	0.29			

The model still explains a statistically *significant* and *substantial* proportion of variance *(R^2 = .70, F (3, 12) = 9.54**)* and the null hypothesis is rejected. From this model it is not obvious which variable to remove. Thus, a partial F-test is conducted to evaluate the remaining variables.

5.2.1.4 Step 3

From the partial F-test the variable EXP1_yrs explains the least amount of variance and is statistically *non-significant.* Consequently, EXP1_yrs is removed and the final model is presented.

Table 9: Summary Final Model

Variable	Estimate	Standard Error	t value	Pr (>|t|)	
EXP2_hrs	.554	.135	4.095	.0013	**
WL	-.489	.135	-3.616	.0031	**

Multiple R-squared: 0.696, Adjusted R-squared: 0.650
F-statistic: 14.94 on 2 and 13 DF, p-value: 0.000427

Table 10: ANOVA Table Final Model

Source	DF	Sum of Squares	Mean Square	F Value	Pr (>F)	
EXP2_hrs	1	4.61	4.61	16.81	.0012	**
WL	1	3.58	3.58	13.07	.0031	**
Residuals	13	3.56	0.27			

The final model explains a *statistically significant* and *substantial* proportion of variance *(R^2 = .70, F (2, 13) = 14.94***)* and the null hypothesis is rejected with only *significant* predictors remaining. Mallow's C_p = 1.55 is low and relatively close to (p + 1 = 3). At a 95% Confidence Interval (CI) none of the predictors contain a regression coefficient of zero. Within this model:

- The effect of EXP2_hrs is *statistically significant* and *positive (beta = 0.55, 95% CI [0.26, 0.85])*
- The effect of WL is *statistically significant* and *negative (beta = -0.49, 95% CI [-0.78, -0.20])*

5.3 T-test

The results from the paired t-test are presented in Table 11.

Table 11: Paired t-test (n = 16)

Factor	Item	Mean	SD	Mean difference	p-value	Significance
Competency	Level (COMP1) Satisfaction (COMP2)	*6.94* *5.94*	*1.6* *2.3*	*1*	*0.015*	*
Performance	Level (PERF2) Satisfaction (PERF3)	*6.88* *6.19*	*1.7* *2.1*	*0.69*	*0.011*	*
Expectations	Level (PERF2) Expectations (PERF1)	*6.88* *7.50*	*1.7* *1.4*	*-0.62*	*0.126*	

6 Discussion

This chapter will first contain a discussion regarding characteristics about the sample and noteworthy observations from the Likert data. Second, follows a discussion on the bivariate correlations among some of the items from the questionnaire. Third, the regression model and findings will be examined. Finally, a holistic overview of the complete dataset and results will be discussed in relation to the primary and secondary research question. Relevant theoretical concepts from chapter 3 will be included where appropriate.

6.1 Survey Data

6.1.1 Sample Characteristics

The final response rate of $n = 16$ was lower than anticipated and may show (1) a sign of *research fatigue* in this community, (2) a lack of time available during working hours and/or (3) simply an unwillingness to participate. Aside from the extra work to complete the survey, there does not seem to be any other downsides to participation, as the survey was anonymous and increased knowledge on this topic should be in the pilots' own interest. From those who did respond a fairly even distribution[12] among the three squadrons were noted, which is favorable and increases the chances of discovering organizational trends prevalent throughout the whole organization.

The experience level across the sample has a high spread as can be seen in Table 1 and Table 2, thus capturing a wide variety of experience, ranging from novice to highly experienced pilots. The least experienced respondent had only 309 flight hours and is probably fresh out of training, while the most senior ones had 7,900 hours or 30 years of service, respectively, which is likely at the top percentile among pilots in the Norwegian Air Force. Furthermore, the spread in the variable CX3[13] is high, with a minimum value of *"2"* and maximum of *"10"*, with observations evenly dispersed among the remaining values, depicted in Figure 11. Consequently, the sample seems to reflect a representative cross section of a squadron's composition of pilots with various experience levels and administrative duties. This suggests that the findings reflect tendencies

[12] Distribution (percentage of sample size): 331 = 31%, 333 = 44%, 335 = 25%

[13] CX3: Seniority level based on rank and position at the squadron.

pertaining to the pilot cadre as a whole, not just a smaller subset of the group. Overall, this increases the face validity of the study.

6.1.2 Likert Scale Data

Reviewing Figure 10 through Figure 13 gives a first impression of the data collected, free from manipulation. Figure 10 and Figure 11 gives an overall impression, while Figure 12 and Figure 13 present differences among the three squadrons. From a first glance - variable WL4[14], WL3[15] and CX6[16] stand out as outliers with particularly high or low reported scores.

6.1.2.1 Workload

WL4 is the variable with the lowest average score *(3.69)* with 50% of the observations reporting either a value of *"1"* or *"2".* The highest reported average score *(7.94)* in the dataset come from WL3 with 62.5% reporting a score of 8-10. The most extreme values were reported by 331 Sqn, with no values above the median line for WL4 and none below the line for WL3, depicted in Figure 12.

Overall, the two variables complement each other in indicating that there seems to be a perception of very high workload at the squadrons, with a particular challenge at the 331 Sqn. As such, there appears to be a disparity between time available (TA) and time required (TR) to perform the sum of all duties demanded by the job, i.e. the ratio of TR/TA > 1. Compensation strategies to mitigate this effect can be to work extra hours, study at home or task shedding. The long-term effects of which may lead to conditions such as fatigue/WRS, dissatisfaction, and reduced competency and performance. An extra cause for concern is that the question explicitly states that this is the sum of work required to maintain a *baseline competency*, meaning airmanship at the lower levels of Figure 3. Thus, it is expected that the work required to maintain a *satisfactory competency* at the higher levels of airmanship will require even more effort and time. These variables seem to be hitting a nerve with such a high percentage of

---

[14] WL4: *"There is enough time available during regular working hours to maintain a baseline competency level as a military aviator and to complete my non-flying related duties"* (Strongly agree/disagree)

[15] WL3: *"Cumulative workload to maintain baseline competency level <u>and</u> complete non-flying related duties"* (Very high/low workload)

[16] CX6: *"How complex is your job overall?"* (Least/most complex)

observations at the extreme end of the spectrum and potentially taps into an underlying frustration regarding excessive workload.

6.1.2.2 Complexity

The variable CX6 stands out as the only variable with all observations on one side of the median line. It has the second highest average score *(7.62)* and the lowest observed variance *(SD = 1.2)* in the dataset. As such, there appears to be a general consensus among the respondents of an overall high complexity involved in the job as a military pilot.

High complexity is likely coupled with the observations on workload in the paragraph above, as there is concordance between task complexity and workload. Increased complexity of cognitive tasks impose higher demands on the working memory load and as such lead to increased time and effort to complete (Young et al., 2015, p. 2).

The three variables WL3, WL4 and CX6 complement and support each other in creating a first impression from the dataset that the job as a military pilot is perceived as complex and difficult which require substantial time and effort to complete the set of tasks demands posed by the job, and that there seems to be a lack of time available to complete these duties. This relationship will be further examined in the succeeding parts of this chapter.

6.1.2.3 Competency and Performance

From the heat map in Figure 11, the variables regarding competency (COMP1[17], COMP2[18]) and performance (PERF2[19], PERF3[20]) have similar numerical values and variation. The mean and standard deviation for the variables regarding perceived level and satisfaction of competence and performance is as follows:

- **Level:** COMP1 = 6.94 (1.6), PERF2 = 6.88 (1.7)
- **Satisfaction:** COMP2 = 5.94 (2.3), PERF3 = 6.19 (2.1)

---

[17] COMP1: *"What is your competency level as a military aviator?"* (Least/most competent)

[18] COMP2: *"How satisfied are you with your competency level as an aviator?"* (Least/most satisfied)

[19] PERF2: *"What is your percieved level of performance as a military aviator?"* (Very low/high performance)

[20] PERF3: *"How satisfied are you with your performance as a military aviator?"* (Very low/high satisfaction)

This harmonizes well with the theoretical framework established in the preceding chapters that there is a strong relationship between competency and performance. On average, the respondents report a higher level of competency and performance than satisfaction.

Another noteworthy observation is that expectations of performance, PERF1[21], is quite a bit higher than the reported level of performance *(7.50 vs 6.88)*, suggesting that on average there is a perceived feeling of not meeting expectations with subpar performance. Out of the 16 observations, 9 (56%) reported performance below expectations, with 5 (31%) reporting on par and only 3 (13%) reported values of performance that exceeds expectations. As noted in paragraph 1.2, perceptions of low performance can have a negative impact on organizational commitment, job satisfaction and turnover rates (Johansen et al., 2014; Robbins, 2013).

6.2 Bivariate Correlations

The bivariate correlations can be seen in Figure 14. Some relationships worth discussing are presented below. As the response scales used between the individual items are phrased differently, no inferences are made from this data set to support conclusions made from the study. As such, the discussion below is informational only.

The strongest negative relationship observed in the matrix is between the variables WL7[22] and WL4 *(r = -.79)*. This is unsurprising as they measure the same concept of time in relation to workload. If time available during working hours to maintain competency is insufficient, then a compensation strategy is to study at home outside of working hours. This is a known concept to many aviators in the military, especially during periods of high stress like initial qualification or upgrade programs. It is unknown from the dataset if respondents were undergoing such programs in the timeframe of participation, but it would likely only be a small subset of the group. Studying during spare time is never explicitly demanded by the employer but goes back to paragraph 3.2.1.1 and the strong sense of obligation to seek the highest standards of airmanship (Kern, 1997; Nergård, 2014). Although unsurprising, it is still probably grounds for concern that such a large portion of the respondents report that they must study outside of working hours to achieve a satisfactory competency level. This is work that is not being

[21] PERF1: *"What is the expected level of performance directed towards you as a military aviator?"* (Very low/high expectations)

[22] WL7: "*I have to study outside of working hours to acquire/maintain a satisfactory competence level"* (Strongly disagree/agree)

compensated for in a traditional monetary way and indicates the level of self-discipline and extra work imposed by the pilots themselves to meet a personal performance standard. Sacrificing spare time for performance is likely not a sustainable model in the long-run and achieving the necessary competency should ideally be conducted inside of working hours to prevent unnecessary fatigue and frustration. As such, it would seem logical that this phenomenon has a negative impact on job satisfaction, and indeed this is supported in the data with a negative correlation between the variables WL7 and OA3[23] *($r = -.29$).*

The strongest positive relationship (aside from intrafactorial relationships) is between the variable CUR6[24] and PERF2 *($r = .85$).* This is another unsurprising find and is in line with the preceding theory regarding *currency*, paragraph 3.2.4. As the data indicates, perceived performance increases with currency. This supports the conclusion that currency training is an important aspect of performance and competency and may indicate that the introduction of new modern aircraft in the Air Force has imposed higher demands for refresher training with the transition from physical to cognitive tasks. The three squadrons participating in the study all operate modern aircraft (C-130J, P-8, F-35) and is such exposed for high levels of automation and complex mission software which may contribute to the strong correlation between these variables.

The second strongest positive correlation was found between WL1[25] and CX4[26] *($r = .84$).* Still, another expected find that underlines the theory in paragraph 3.2.5. The variables seem to be coupled in that an increase in complexity corresponds to an increase in workload, leading to further strain on the TR/TA ratio.

---

[23] OA3: *"How satisfied are you with your job overall?"* (Very low/high satisfaction)

[24] CUR6: *"How do your rate your currency at the present moment?" (Least/most current)*

[25] WL1: Workload to maintain baseline competency level as a military aviator (Very low/high workload)

[26] CX4: «*How complex is your job as a military aviator?*» *(Least/most complex)*

6.3 Regression Analysis

The linear multiple regression analysis from paragraph 5.2.1.4 yielded the following model:

$$Competence_{score} = 0.55EXP2_{hrs} - 0.49WL_{score}$$

This model is *statistically significant* with none of the regression coefficients containing zero in their 95% confidence interval. As such, the data indicates that both flight hours and workload likely contribute to predict the perception of competence with a multiple R-squared value of $R^2 = .70$, indicating that 70% of the variation in competence is explained by the constructs of experience and workload. This is a substantial proportion of variation and is larger than expected. The juxtaposition of the predictors, with their opposing slopes, likely tap into differing constructs related to competence. By removing the variables *currency*, *complexity* and *years of experience* the *adjusted R-squared* value increased from *.61* to *.65*, instilling more confidence in the model. Based on the data from this sample none of these variables seems to have predictive power over the outcome variable. However, the small sample size negates discovery of relationships with low correlation.

Still, a large part (30%) of the variation in the model is not being explained by the regressors indicating that there are additional constructs factoring into the perception of competence that is not being captured by the model.

6.3.1 Experience

The regressor *EXP2_hrs (flight hours)* is the most significant variable in the model and explains ~39% of the variation in the outcome variable with $p = .0013$**. It has a positive association and a beta of *.55* with *perception of competence*, meaning the feeling of competence will increase by *.55* for every unit increase in flight hours.

This is in line with expectations and theory; however, such a strong relationship was perhaps a bit surprising as it challenges some of the theory presented in paragraph 3.2.2. It was further assumed that there would be some negative effects from experience as the administrative duties increase with seniority. This does not seem to play a major role in the data. The findings underline that flight hours can be used as a measure of competency and that the two constructs are closely interrelated. As Martinussen and Hunter (2018) stated, learning never really plateaus which seems to be supported by this model. It appears thus to legitimize the frequent use of flight hours in aviation as a prerequisite for employment, upgrade programs and license ratings.

6.3.2 Workload

The regressor WL explains less variation (~30%) in the outcome variable but is still *statistically significant* with $p = .0031^{**}$. Further, the magnitude of the coefficient is large with a negative association of *-.49.* As such, the model indicates that workload has a strong negative impact on the perception of competence. This appears to be in concordance with the theories on performance and workload with an emphasis on the concept from the performance curves presented in Figure 5 and Figure 6. Complemented by the findings in paragraph 6.1.2.1 it appears as though the respondents in the survey are located on the upper right half of these curves, close to the workload redline, with an increase in demands contributing to a sharp decline in performance. Adding more work is thence clearly unfavorable and stresses the need for a holistic approach considering the total burden of workload in attempts to maximize competence. Any push in the direction of increased demands seems to result in overload by crossing the workload redline.

6.4 Interpretation of Results

6.4.1 Primary Research Question

> *How do characteristics of the profession as a military aviator cause variation in the perception of competence?*

The characteristics that emerge as significant predictors of the perception of competence in this study are the constructs of *experience* and *workload.* The influence of experience on competence seems to be driven mainly by the amount of flight hours accrued, not years spent flying. The recency of the flight hours are probably also an important characteristic of feeling competent, but *Currency* did not seem to explain a significant proportion of variation in a model already containing *experience* and *workload*. However, a strong bivariate correlation *($r = .80$)* between currency (CUR6)[27] and the *competence score*, Figure 14, can be observed, indicating that there might be constructs in the data that is not reflected in the regression model. *Complexity* was also dropped from the regression model for its lack of predictive power although high values of complexity were reported in the Likert scale data, particularly for item CX6. However, this item was dropped from the FA and as such was not included in the regression analysis. Thus, the overall factor *complexity* may be measuring something other than

[27] CUR6: *"How do you rate your currency at the present moment?"* (least/most current)

actual complexity. The construct of actual complexity may be explained partly by *workload* in the model as these concepts are closely interrelated as noted by the strong bivariate correlation *($r = .61$)* between total perceived workload (WL3) and overall job complexity (CX6).

The result that more flight hours positively contribute to the feeling of competence is hardly surprising but reinforces the importances of letting pilots practice their skillset on a frequent basis and serves as a reminder of the alternative cost associated with demands posed by tasks and duties not related to flying. As such, an alternative interpretation of this data can be that any tasks not related to flying will have a detrimental impact on flight hours thus leading to a reduced feeling of competence. Tasks not related to flying will also have a negative impact on *workload.*

From the final model *workload* contribute negatively to the perception of competence. A similar impression is emerging from the unaltered Likert scale data, paragraph 6.1.2.1, and from the bivariate correlation matrix, Figure 14. Combined, they imply that as workload increase there is a subsequent drop in the perception of competence, indicating that many of the observational units in the study are hovering close to the workload redline. The conditions that seem to influence the perception of high workload are high complexity (CX6) and a disparity between TR and TA (WL4), forcing many to study outside working hours to maintain proficiency. The disparity can be further exacerbated if non-flying related workload (WL2[28]) is high, and several observations are noted at the highest end of the scale, with 50% of the observations reporting *"7"* or above. This is illustrated by one of the few qualitative remarks obtained from the survey:

> *"Full time job is not enough to fill all annual training requirements for military aviators in Norway. Even less so for staff/part time flyers. Average 150-180% of a man-year required."*

As such, the characteristics that cause variation in the perception of competence among military aviators in Norway seems to be predominantly *experience* through flight hours and *workload.* The characteristics oppose each other, and stakeholders should attempt to maximize flight hours while alleviating workload to foster pilots that are feeling competent.

[28] WL2: Workload to complete non-flying related duties. (Very low/high workload)

6.4.2 Secondary Research Question

Are pilots in the RNoAF satisfied with their own level of competence?

This is a somewhat vague and almost dichotomous question and was therefore not stated directly in the survey, to avoid forcing respondents to either a *"yes"* or *"no"* answer. Yet, indications can be drawn from the variables on competency and performance. Overall, the observations on levels of competency (COMP1) and performance (PERF2) are relatively high implying that most of the participants feel *"good enough for government work"*. However, the reported values for satisfaction on the same constructs, Table 11, are on average lower, indicating that there is some dissatisfaction among the respondents. Both t-tests on these factors showed a statistically significant difference between the means of reported *levels* and the corresponding *satisfaction.* This is complemented by the findings from paragraph 6.1.2.3 that the average reported levels on performance are a bit lower than expectations *(6.88 vs 7.50).* However, this difference was not statistically significant. Furthermore, the quality and amount of data on this topic is limited due to the small number of items and low sample size. As such, it is not appropriate to give a conclusive "yes" or "no" answer to this research question.

6.5 Comparison with Existing Literature

None of the findings in this study seems to be in opposition to the literature reviewed in the thesis, rather the findings support and accentuate certain aspects important for competency and performance among military pilots. The aviation industry's move towards Evidence Based Training (EBT) seems like a step in the right direction with its stronger focus on pilot competencies regardless of accumulated flight hours (Mendonca et al., 2019, p. 3). However, this study indicates that flight hours, although not all-encompassing, can still be used as an easy and convenient measure of competency. Consequently, it further seems to support the studies of Taylor et al. (2007) and G. Li et al. (2003) with a positive association between experience and performance.

Furthermore, the study seems to strongly support Young et al. (2015) in the concept of a *workload redline* with a sharp decline in performance when entering the overload region - considering the strong negative association of workload from the regression model. Complimented by the very low values of the Likert scale data observed in item *WL4 (Time available)*, the study further seems to indicate the applicability of using time as a count variable for measuring workload, by using a simple timeline analysis like the TR/TA ratio as suggested by Parks and Boucek Jr (1989). Finally, the results are in line with the thesis from Wibe and Hanssen (2023) and their findings of high workload among Norwegian F-35 pilots.

6.6 Implications

Findings from the study imply the need for a careful and balanced approach in managing workload. The emerging empirical evidence suggest that most of the observational units in the sample are seriously close to a *workload redline* depicted in Figure 5. Thus, simply adding another strap-on CRM course to the annual currency requirements is unlikely to contribute positively to the perception of competence if other aspect of the burden of workload is not reduced. If increased demands are posed from the recurrent training program or operational missions, other tasks need to be removed to avoid incursions into the overload region. The lowest hanging fruits for stakeholders will likely be to alleviate task demands that can be automated out or performed by personnel not qualified as pilots. These colleagues cannot step into the cockpit when workload is too high, but they can help with administrative duties thus lowering the overall burden faced by pilots, and as such contribute to raising performance.

A focus on job satisfaction also emerges as an important aspect for stakeholders seeking to maximize performance, considering the strong positive relationship between *flight hours* and *competence.* Building experience through flight hours is expensive and time-consuming, but there seems to be no cheap substitute. Alas, experienced pilots are difficult to replace, which demands low turnover rates to avoid losing critical personnel due to dissatisfaction. This requires organizations to facilitate and schedule for training programs that foster a high level of proficiency among pilots so they can keep flying with a strong sensation of feeling competent. This should also have a synergistic effect on safety as a competent pilot will be more likely to manage situations in flight that have been unforeseen in the training program and for which he has not been specifically trained (ICAO, 2013, pp. I-1-1).

6.7 Limitations

"Torture the data long enough and it will confess to anything" – Ronald H. Coase

An inherent limitation of a web-based survey is that it is difficulty of knowing whether all participants actually received the survey (Martinussen & Hunter, 2018, p. 27). This was certainly the case for this project as the surveys were distributed across the Atlantic Ocean via fiber optic cables, reverted over to a classified email system, and subsequently disseminated by middlemen. This created a frustrating distance between the researcher and the observational units which likely lowered the response rate, as adequate follow-up throughout the data collection phase was deficient. A reminder was sent out halfway through the collection period but did not have a meaningful impact on the response rate. As the survey was voluntary, a restrained approach in encouraging participation was essential to avoid nudging participants into consent. These conditions are likely contributors to the most significant weakness of the study, namely the low sample size.

The study's sample size of $n = 16$ represents a response rate of less than 10% of the population, making it questionable whether any reliable conclusions can be drawn from the regression analysis. Military pilots have historically been subject to much research, suggesting a possibility that there is an element of "research fatigue" in these communities, which might partly explain the low response rate. On a more general note, web-based surveys have seen a decline in engagement, with the problem of non-respondents rising (Jacobsen, 2015, p. 311). This is likely due to its dramatic increase in usage due to the low cost and convenience of use alongside the explosion of internet use. As such, a low response rate is not a unique problem for this study. However, there are some very real challenges in working with small sample sizes. Correlation coefficients fluctuate from sample to sample due to natural variation and the fluctuation is much more pronounced in smaller samples (Field et al., 2012, p. 769). This lowers the confidence in the results significantly. A small sample size also precludes discovery of relationships with smaller correlations. The effect is exacerbated by the use of constructed data for some of the data fields for the respondents from 331 Sqn, reducing both the reliability and validity of the study.

Furthermore, most of the data obtained in this study are based on indirect and subjective measures such as self-reported Likert scale items. This likely represents another weakness as the scale used in the questionnaire is not exactly equally interpreted by all respondents. Meaning, a reported value of *"8"* by one respondent on the Likert scale may not exactly represent the same for someone else also reporting *"8"* on the same question, and it certainly does not mean that those who reported *"8"* feel that they are exactly twice as good/bad as someone reporting a value of *"4"*. This is why Field et al. (2012, p. 9) argue that self-report data using Likert scale items should not be treated as continuous, interval data – which would preclude using it in most statistical analysis. However, Likert (1932, p. 42) claims that if one collects responses about a series of statements regarding the same factor/variable an average numerical value can be computed and the distribution of this data will approach a normal distribution as n increases - which has been the strategy for this thesis. Thus, the empirical data obtained in this thesis are treated either as *interval* (data from Likert scale items) or *continuous* (e.g. flight hours) which opens up for a broader range of available statistical analysis (Martinussen & Hunter, 2018, p. 23).

Due to time constraints no pilot test was performed prior to distributing the questionnaires, which is according to Fink (2009, p. 6) a crucial step in conducting surveys that can improve the clarity of the language and increase response rates. This could have also helped eliminate superfluous questions and reduced the overall length of the survey, possibly increasing the response rates (Martinussen & Hunter, 2018, p. 27). The survey had 34 questions, with an average response time of 15 minutes, which might have been too lengthy for some of the potential respondents and possibly deterred them from completing the survey.

Finally, the author has insider knowledge and is part of the population under study. As such, there is a risk of underlying predetermined and prejudicial opinions that might skew the results as statistics are inherently subjective (Field, 2012, p. 808). Furthermore, the condition as an *insider* makes the study susceptible to *blind spots* where the researcher does not have the critical distance needed to approach the subject in an unbiased way (Jacobsen, 2015).

6.8 Ethical Considerations

The study followed the principles of the Helsinki Declaration (WMA, 2013). All participants were explicitly informed that the survey was voluntary, and participation could be withdrawn at any time, without question. An assessment of privacy and personal data was conducted, and the research was defined as anonymous as judged by the standards of the Norwegian Social Science Data Services (Sikt) and accomplished by using functions for anonymity on *nettskjema.no.* As such, the research was not required to report a notification form to Sikt. An informational letter was attached to the email containing the survey which explained respondent's rights and how their data would be handled. Written consent to participate was not collected, but the informational letter contained information stating that completion of the survey would be regarded as consent.

7 Conclusion

The primary purpose of this study has been to examine the relationship between characteristics of the job as a military pilot and the individual's perception of competence to further expand the knowledge base on the topic. This will in turn enable stakeholders to make more informed decisions to better facilitate for pilot competency and performance, ultimately leading to increased operational efficiency and safety. Secondly, the study has surveyed pilot's satisfaction with competence, i.e. the feeling of being good enough, with the assumption that feelings of incompetence can lead to lower job satisfaction.

The main findings are primarily based upon a statistically significant multiple regression model consisting of the predictor variables *experience* and *workload* which explained a substantial amount of variance *(70%)* in the outcome variable *perceived competence.* The variables *currency* and *complexity* did not display predictive power in the model.

Experience, in the form of *flight hours*, indicates a strong positive association with the feeling of competence, and is in line with expectations and prior studies on human performance. The findings instill confidence in a continued use of *flight hours* as a measure of competency, already prevalent in the aviation industry. Further, it accentuates the susceptibility of aviation organizations to turnover rates which calls for high levels of job satisfaction to prevent pilots from quitting early.

The second predictor of competence, *workload*, indicate a strong negative association and is perhaps the most interesting result from the study. The empirical data clearly show that a majority of the sample is working close to a *workload redline*, which results in a sharp decline in performance if crossed. Running this close to maximum capacity leaves little room for adding task demands and as such it becomes a zero-sum game where increased effort in one facet of the job will result in a detrimental effect on other components, which can lead to frustration among ambitious workers striving for perfection. Particularly, it appears to be a disparity between the time required (TR) to maintain high levels of competence, and the time available (TA). As with aircraft performance which sharply degrades when Power Required (P_R) exceeds Power Available (P_A) (Anderson, 1999, p. 239), the same principle seems to apply to the humans in the system if TR exceeds TA. In military aviation, nobody wants a low P_S[29]. The same indignation should probably also apply to a low T_S[30].

The data is less conclusive on whether the pilots feel good enough for the job or not. On the one hand, the reported scores on level of competency, performance and overall job satisfaction are high. However, on the other hand the self-report data on satisfaction with competency and performance is quite a bit lower, with a statistically significant difference in the means from the paired t-tests, suggesting at least some facets of dissatisfaction. This is complimented by reported performance levels below expectations. On this topic the jury is still out which lay grounds for further research.

Finally, lapses in safety are still being labeled as operator error with accident reports filled with connotations of poor airmanship. Historically, this has been counteracted by numerous approaches like technological improvements, added safeguards and new training programs aimed to enhance crew coordination, reduce errors and fully utilize the human resources on the flight deck. While these initiatives have achieved various levels of success, *the individual flyer,* according to Kern (1997, p. xxx), remains the key to meeting the last great challenge in aviation - human error. That may be, and if such is the case then *the individual flyer,* i.e. *the pilot*, need to be afforded optimized working conditions. That means finding a well-balanced workload, facilitating for performance at the top of the Yerkes-Dodson curve with a buffer to the workload redline, while continuously building flight hours and managing task demands faced by the pilots. Only by optimizing these components can the system be made to work as safely and efficiently as possible.

7.1 Recommendations for Further Research

Initially, further studies on the same topic with a larger sample size should be conducted to validate the findings of this research. Cross-validation to assess the ability to generalize the findings should also be performed. Secondly, a logical extension of this study would be to investigate further the relationship between *perceived* competence and *actual* competence, as the positive relationship between the two is only an assumption made in this study. Finally, a

[29] P_S (Specific excess power) = P_A - P_R

[30] T_S (Specific excess time) = T_A - T_R (made up term by the author)

closer investigation of pilot's satisfaction with competence seems appropriate as no conclusive results regarding this topic were provided in this thesis.

References

Adler, A. B., Thomas, J. L., & Castro, C. A. (2005). Measuring up: Comparing self-reports with unit records for assessing soldier performance. *Military psychology, 17*(1), 3-24.

Anderson, J. D. (1999). *Aircraft performance and design*. Washington D.C.: McGraw Hill.

Arifin, M. A. (2021). Competence, competency, and competencies: A misunderstanding in theory and practice for future reference. *International Journal of Academic Research in Business and Social Sciences, 11*(9), 755-764.

Arthur Jr, W., Bennett Jr, W., Stanush, P. L., & McNelly, T. L. (1998). Factors That Influence Skill Decay and Retention: A Quantitative Review and Analysis. *Human performance, 11*(1), 57-101. doi:10.1207

Baker, D. P., & Dismukes, R. K. (2002). A Framework for Understanding Crew Performance Assessment Issues. *The International Journal of Aviation Psychology, 12*(3), 205-222. doi:10.1207/s15327108ijap1203_2

Bion, W. (2023). *Learning from experience*: Taylor & Francis.

Box, G. E., & Draper, N. R. (1987). *Empirical model-building and response surfaces*: John Wiley & Sons.

Cahill, J., Cullen, P., Anwer, S., Wilson, S., & Gaynor, K. (2021). Pilot work related stress (WRS), effects on wellbeing and mental health, and coping methods. *The International Journal of Aerospace Psychology, 31*(2), 87-109.

Caldwell, J. A. (2005). Fatigue in aviation. *Travel medicine and infectious disease, 3*(2), 85-96. doi:10.1016/j.tmaid.2004.07.008

Cooper, G. E., White, M. D., & Lauber, J. K. (1980). *Resource management on the flight deck.* Paper presented at the NASA/Industry Workshop.

Damos, D. L. (2011). KSAOs for military pilot selection: A review of the literature. *Air Force Personnel Center, Randolph AFB, Texas*.

Dekker, S. (2002). The re-invention of human error. *Human Factors and Aerospace Safety, 1*(3), 247-265.

Dekker, S. (2006). *The field guide to understanding human error*. Aldershot: Ashgate.

Dekker, S., Cilliers, P., & Hofmeyr, J.-H. (2011). The complexity of failure: Implications of complexity theory for safety investigations. *Safety Science, 49*(6), 939-945. doi:https://doi.org/10.1016/j.ssci.2011.01.008

Field, A. (2012). Discovering statistics using R, companion website. Retrieved 3rd of May from https://studysites.uk.sagepub.com/dsur/study/DSUR%20Smart%20Alex-Labcoat%20Leni-Self%20Test%20Answers/DSUR%20Chapter%2017%20Web%20Material.pdf

Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. London: Sage Publications.

Fink, A. (2009). *How to conduct surveys: A step-by-step guide* (4th ed.): SAGE publications.

Fisher, R. A. (1934). *Statistical methods for research workers* (5th ed.): Oliver and Boyd.

Fisher, R. A. (1966). *The design of experiments* (Vol. 21): Oliver and Boyd Edinburgh.

Flight Safety Australia. (2005). The Right Stuff. *Flight Safety Australia*. Retrieved from https://skybrary.aero/sites/default/files/bookshelf/1132.pdf

Forsvarets Høgskole. (2024, 3rd April). *Bruk av data fra Forsvaret i forskning.* Retrieved 9th April 2024 from https://www.forsvaret.no/forskning/forsvarets-forskningsnemd

FAA. (2021). *Airplane flying handbook (FAA-H-8083-3C)*. Oklahoma City: United States Department of Transportation, Federal Aviation Administration, Airman Testing Standards Branch.

FAA. (2022). Eastern Airlines Flight 401, N310EA. Retrieved 26th of April from https://www.faa.gov/lessons_learned/transport_airplane/accidents/N310EA

Gartner, W., & Murphy, M. (1976). *Pilot workload and fatigue: A critical survey of concepts and assessment techniques*. Retrieved from https://ntrs.nasa.gov/api/citations/19770004731/downloads/19770004731.pdf
Grindheim, H. (2023). *Luftforsvarets flygerseleksjon: En studie av prediktiv validitet.* FHS, Retrieved from https://fhs.brage.unit.no/fhs-xmlui/handle/11250/3087354
Haga, L. P., & Maaø, O. J. (2018). *Forsvarets doktrine for luftoperasjoner* (8269151408). Retrieved from https://fhs.brage.unit.no/fhs-xmlui/handle/11250/2634745
Harlem, T. O. (2016). *Seleksjon flygere F-35: Neste generasjon jagerfly, en ny generasjon flygere?* FHS, Retrieved from https://fhs.brage.unit.no/fhs-xmlui/handle/11250/2436208
Helmreich, R. L., & Foushee, H. C. (2010). Why CRM? Empirical and theoretical bases of human factors training. In *Crew resource management* (pp. 3-57): Elsevier.
Helmreich, R. L., Merritt, A. C., & Wilhelm, J. A. (1999). The Evolution of Crew Resource Management Training in Commercial Aviation. *The International Journal of Aviation Psychology, 9*(1), 19-32. doi:10.1207/s15327108ijap0901_2
Hoffmann, T. (1999). The meanings of competency. *Journal of European industrial training, 23*(6), 275-286. doi:10.1108/03090599910284650
Howitt, J. (1969). *Flight-deck workload studies in civil transport aircraft.* Paper presented at the AGARD Conference Proceedings.
IATA. (2023). *Competency Assessment and Evaluation for Pilots, Instructors, and Evaluators Guide Manual*. Retrieved from https://www.iata.org/contentassets/c0f61fc821dc4f62bb6441d7abedb076/competency-assessment-and-evaluation-for-pilots-instructors-and-evaluators-gm.pdf
ICAO. (2013). *Manual of Evidence-based Training*. Retrieved from https://skybrary.aero/bookshelf/books/3177.pdf
Jacobsen, D. I. (2015). *Hvordan gjennomføre undersøkelser? Innføring i samfunnsvitenskapelig metode* (Vol. 3): Cappelen Damm Akademisk.
Johansen, R. B., Laberg, J. C., & Martinussen, M. (2014). Military Identity as Predictor of Perceived Military Competence and Skills. *Armed Forces & Society, 40*(3), 521-543. doi:10.1177/0095327x13478405
Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits self-esteem, generalized self-efficacy, locus of control, and emotional stability with job satisfaction and job performance: a meta-analysis. *The Journal of applied psychology, 86 1*, 80-92.
Kern, T. (1997). *Redefining airmanship*. New York: McGraw-Hill.
Kleinbaum, D. G., Kupper, L. L., Rosenberg, E., & Nizam, A. (2014). *Applied regression analysis and other multivariable methods* (5 ed. Vol. 601). Boston, MA: Cencage Learning.
Knutsen, Ø. (2023). *Safety i det norske jagerflyvåpenet: En analyse basert på teorier rundt Resilience Engineering og Safety-II.* FHS, Retrieved from https://fhs.brage.unit.no/fhs-xmlui/handle/11250/3113118
Li, G., Baker, S. P., Grabowski, J. G., Qiang, Y., McCarthy, M. L., & Rebok, G. W. (2003). Age, Flight Experience, and Risk of Crash Involvement in a Cohort of Professional Pilots. *American Journal of Epidemiology, 157*(10), 874-880. doi:10.1093/aje/kwg071
Li, W., Lee, A. M., & Solmon, M. A. (2005). Relationships Among Dispositional Ability Conceptions, Intrinsic Motivation, Perceived Competence, Experience, Persistence, and Performance. *Journal of Teaching in Physical Education, 24*, 51-65.
Likert, R. (1932). *A Technique For the Measurement of Attitudes* (Vol. 22). New York.
Luftforsvaret. (2017). *Bestemmelser for Militær Luftfart*. Retrieved from https://www.forsvaret.no/soldater-og-ansatte/regelverk/bestemmelser-millitaer-luftfart.pdf

Luftforsvaret. (2020). Rapport om alvorlig Luftfartshendelse, Mosken, 11.mars 2020, med to C-130J, Operert av Luftforsvaret. Forsvarsintern undersøkelsesrapport.
Marrow, M. (2024, 15th of April). F-35 program's life time price tag. Retrieved 5th of May from https://breakingdefense.com/2024/04/f-35-programs-lifetime-price-tag-tops-2-trillion-pentagon-wants-jets-to-fly-longer/
Martinussen, M., & Hunter, D. R. (2018). *Aviation Psychology and Human Factors* (2nd ed.). Boca Raton, Florida: CRC Press.
McFarland, R. A. (1971). Understanding Fatigue in Modern Life. *Ergonomics, 14*(1), 1-10. doi:10.1080/00140137108931216
Mendonca, F. A., Keller, J., & Dillman, B. G. (2019). Competency Based Education: A Framework for a More Efficient and Safer Aviation Industry. Retrieved from https://commons.erau.edu/publication/1706
Mesarosova, K. (2020). Personality in Pilot Selection and Training Is There a Right Stuff? In R. Bor, C. Eriksen, T. P. Hubbard, & R. King (Eds.), *Pilot Selection* (1 ed., pp. 235-254): CRC Press.
Moldjord, C. (2023). Læringskultur i operative miljøer – Luftforsvaret som eksempel. In *Militær leder- og ledelsesutvikling i teori og praksis* (pp. 431-453): Universitetsforlaget.
Moore, D. R., Cheng, M.-I., & Dainty, A. R. J. (2002). Competence, competency and competencies: performance assessment in organisations. *Work study, 51*(6), 314-319. doi:10.1108/00438020210441876
Moriarty, D. (2015). *Practical human factors for pilots*. London, England: Academic Press.
Nergård, V. (2014). Airmanship - a qualitative approach. *Aviation, 18*(3), 147-156. doi:10.3846/16487788.2014.969882
Nergård, V., Hatlevik, O. E., Martinussen, M., & Lervåg, A. (2012). An airman's personal attitude: pilots' point of view. *Aviation (Vilnius, Lithuania), 15*(4). doi:10.3846/16487788.2011.651789
Parks, D. L., & Boucek Jr, G. P. (1989). Workload prediction, diagnosis, and continuing challenges. In *Applications of human performance models to system design* (pp. 47-63): Springer.
Perrow, C. (1999). *Normal Accidents: Living with High Risk Technologies - Updated Edition*: Princeton: Princeton University Press.
Pierson, K. (2022). *Pilots in Command : Your Best Trip, Every Trip*. Newcastle, WA: Aviation Supplies & Academics, Incorporated.
Pinker, S. (2018). *Enlightenment now: The case for reason, science, humanism, and progress*: Penguin UK.
Prop 87 S. (2024). *Forsvarsløftet - for Norges Trygghet*. Retrieved from https://www.regjeringen.no/contentassets/27e00e5acc014c5ba741aacfff235d99/no/pdfs/prp202320240087000dddpdfs.pdf
Ratcliffe, S. (1969). Mathematical models for the prediction of air traffic controller workload. *The Controller, 9*, 18.
Reason, J. (1997). *Managing the risks of organizational accidents* (1 ed.). Aldershot: Ashgate.
Rjaanes, M., Kalveland, M., Olsen, K. E., Haugen, R., Beadle, A. W., & Aarønæs, L. (2020). *Teknologiske trender–mulige konsekvenser for Luftforsvaret* (8246432877). Retrieved from https://ffi-publikasjoner.archive.knowledgearc.net/handle/20.500.12242/2760
Robbins, S. P. (2013). *Organizational behavior* (15th, global ed. ed.). London: Pearson.
Rognstrand, A. (2020, 20th of July 2020). Hercules i nestenulykke i mars: Marginer skilte fra katastrofe. Retrieved 4th of May from https://www.forsvaretsforum.no/flyulykke-

hercules-luft/hercules-i-nestenulykke-i-mars-marginer-skilte-fra-katastrofe/130550?noLog=1
Ruffel Smith, H. (1979). *A simulator study of the interaction of pilot workload with errors, vigilance, and decisions*. Retrieved from https://ntrs.nasa.gov/api/citations/19790006598
Rønningstad, S. H. (2024). Alternativkostnad. Retrieved 29th of April from https://snl.no/alternativkostnad
Salas, E., Bowers, C. A., & Rhodenizer, L. (1998). It Is Not How Much You Have but How You Use It: Toward a Rational Use of Simulation to Support Aviation Training. *The International Journal of Aviation Psychology, 8*(3), 197-208. doi:10.1207/s15327108ijap0803_2
Salas, E., Burke, C. S., Bowers, C. A., & Wilson, K. A. (2001). Team Training in the Skies: Does Crew Resource Management (CRM) Training Work? *Human factors, 43*(4), 641-674. doi:10.1518/001872001775870386
Schendel, J. D., Shields, J. L., & Katz, M. S. (1978). *Retention of motor skills, review*(Vol. 313).
Schutte, P. (2017). How to make the most of your human: design considerations for human–machine interactions. *Cognition, technology & work, 19*(2-3), 233-249. doi:10.1007/s10111-017-0418-2
SHK. (2013). *Sluttrapport RM 2013:02*. Retrieved from https://www.havkom.se/assets/reports/Swedish/RM-2013_02_no.pdf
Strebler, M. (1997). *Getting the Best Out of Your Competencies*: ERIC.
Stueland, E. (2023). *En analyse av risiko og complacency hos jagerflygere.* FHS, Retrieved from https://fhs.brage.unit.no/fhs-xmlui/handle/11250/3087352
Svensson, K. (2013). *Flygarseleksjon i Forsvaret: ein studie av seleksjonssystemet sin prediktive validitet.* FHS, Retrieved from https://fhs.brage.unit.no/fhs-xmlui/handle/11250/100112
Taboga, M. (2021). Linear Regression with standardized variables. *Lectures on probability theory and mathematical statistics.* Retrieved 2nd Mar 2024 from https://www.statlect.com/fundamentals-of-statistics/linear-regression-with-standardized-variables
Taylor, J., Kennedy, Q., Noda, A. M., & Yesavage, J. A. (2007). Pilot age and expertise predict flight simulator performance. *Neurology, 68*, 648 - 654.
Teigen, K. H. (1994). Yerkes-Dodson: A law for all seasons. *Theory & Psychology, 4*(4), 525-547.
Thompson, R. F., & Krupa, D. J. (1994). Organization of memory traces in the mammalian brain. *Annual review of neuroscience, 17*(1), 519-549.
Todd, M., & Thomas, M. (2012). Flight Hours and Flight Crew Performance in Commercial Aviation. *Aviation, space, and environmental medicine, 83*, 776-782. doi:10.3357/ASEM.3271.2012
Tooby, J., Cosmides, L., & Barrett, H. C. (2003). The Second Law of Thermodynamics Is the First Law of Psychology: Evolutionary Developmental Psychology and the Theory Of Tandem, Coordinated Inheritances: Comment on Lickliter and Honeycutt (2003). *Psychological Bulletin, 129*(6), 858-865. doi:https://doi.org/10.1037/0033-2909.129.6.858
Tranøy, K. E. (2019, 18th Feb). *Metode.* Retrieved 8th April 2024 from https://snl.no/metode
University of Oslo. (2024). *Nettskjema.no.* Retrieved 9th April 2024 from https://nettskjema.no/?lang=en

Wibe, M., & Hanssen, S. (2023). *Sikkerhet og sikkerhetskultur i kampflymiljøet i Luftforsvaret.* Nord universitet, Retrieved from https://nordopen.nord.no/nord-xmlui/bitstream/handle/11250/3100279/WibeHanssen.pdf?sequence=1&isAllowed=y
Wickens, C. D., & Dehais, F. (2019). Expertise in Aviation. In *The Oxford Handbook of Expertise*. Toulouse: Oxford University Press.
WMA. (2013). *Declaration of Helsinki*. Retrieved from https://www.wma.net/policies-post/wma-declaration-of-helsinki-ethical-principles-for-medical-research-involving-human-subjects/
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology and Psychology, 18*(5), 459-482. doi:https://doi.org/10.1002/cne.920180503
Young, M. S., Brookhius, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: mental workload in ergonomics. *Ergonomics, 58,* 1-17. Retrieved from https://doi.org/10.1080/00140139.2014.956151

Abbreviations

AESA......... Active Electronically Scanned Array
ANOVA .... Analysis of Variance
ATP.................................Airline Transport Pilot
ATPL ....Airline Transport Pilot License
CFIT ....Controlled Flight Into Terrain
CP .... Co-Pilot
CRM .... Crew Resource Management
CV ....Curriculum Vitae
DF...................................... Degrees of Freedom
DNIF......................... Duty Not Involving Flying
EASA............. European Aviation Safety Agency
EBT..............................Evidence Based Training
FA.. Factor Analysis
FAA ....Federal Aviation Administration
FHS....................................Forsvarets Høgskole
FOC .... Full Operational Capability
FWIT .... Fighter Weapons Instructor Training
IA................................... Information Assurance
IATA...... International Air Transport Association
ICAO .. International Civil Aviation Organization
ILS ....Instrument Landing System
KSAO .... Knowledge, Skills, Abilities, and Other characteristics
LFS .... Luftforsvarets Flygeskole
NASA .... National Aeronautics and Space Administration
NATOPS........ Naval Air Training and Operating Procedures Standardization
PPC .... Patrol Plane Commander
RNoAF.....................Royal Norwegian Air Force
SD..Standard Deviation
SHEL ....Software, Hardware, Environment, Liveware
SQN ....Squadron
TEM..................... Threat and Error Management
TPS .... Test Pilot School
TA.. Time Available
TR .... Time Required
USN .... United States Navy
WIC.......................... Weapons Instructor Course
WMA ....World Medical Association
WRS.................................. Work-Related Stress
YDL .... Yerkes-Dodson Law

Codes:

COMP ....Competency
CUR ....Currency
CX... Complexity
EXP..Experience
OA ....Overall
PERF..Performance
WL ....Workload

Appendix A – Bivariate Correlation Matrix

The following plot present the bivariate correlation matrix for most of the items contained in the survey. The complete phrasing for each item can be found in the questionnaire. Variables of low correlation are intentionally unreadable, as they are of little interest. Stronger correlations have a higher contrast to stand out in the matrix.

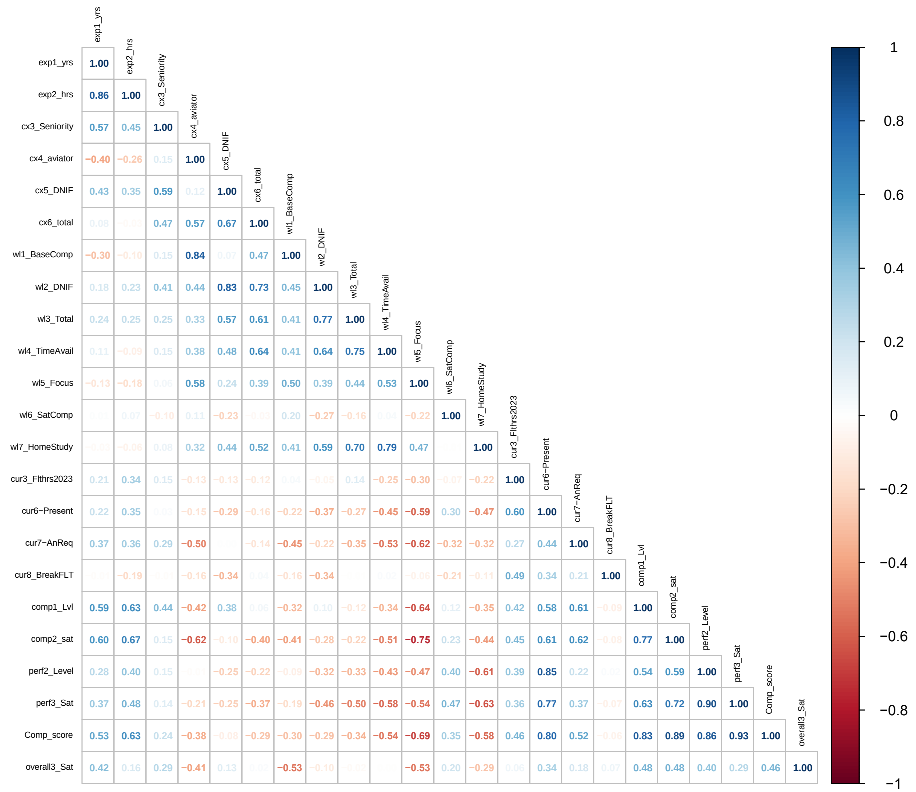

Figure 14: Correlation Matrix

Appendix B – Questionnaire

Airmanship and perception of competence - Final version

Formalities

Are you a military pilot actively flying today?
If you somehow ended up with this survey and is not a military pilot, please refrain from completing this survey.

Yes
No

Which squadron do you currently work at?

331
333
335

Experience (EXP)

How many years have you been flying for the RNoAF? ***(EXP1)***
Count years since your initial qualification as an operational co-pilot/wingman in the RNoAF at your respective squadron. Do not include years not spent flying (such as non-flying staff tours, fulltime student at the Air Force Academy etc)

How many military flight hours do you have in total? ***(EXP2)***
Include all military flight time, including flight training. A rough estimate is acceptable (e.g. ~2100hrs)

How many military flight hours do you fly on average each year? ***(EXP3)***
A rough estimate is acceptable (e.g. ~360hrs)

Complexity (CX)

What are your qualifications as a military aviator in the RNoAF? ***(CX1)***
Pick all that applies

Co-pilot
Wingman
Aircraft Commander (Besetningssjef)
Formation lead (Fighter)
Aircraft Instructor Pilot
Simulator Instructor Pilot
Functional Check Flight Pilot (Prøveflyger 2)
Instrument Card Check Evaluator
Other - please elaborate in textbox below

Other qualifications ***(CX2)***

What is your seniority level at the squadron? ***(CX3)***
As measured by rank and position at the squadron
1 = Least senior
10 = Most senior

Figure 15: Survey

Nettskjema

In your opinion - how complex is your job as a military aviator? ***(CX4)***
Consider only the flying-related portion of your job.
1 = Least complex
10 = Most complex

In your opinion - how complex is the non-flying related part of your job? ***(CX5)***
Consider only the non-flying related portion of your job.
1 = Least complex
10 = Most complex

In your opinion - how complex is your job overall? ***(CX6)***
Consider the cumulative sum of both the flying-related and all non-flying related duties in your job.
1 = Least complex
10 = Most complex

Workload (WL)

Baseline competency level for the following questions is defined as a competency level meeting the *minimum* requirements for competency as set forth by annual currency requirements for flying and squadron/RNoAF regulations and flight manuals (e.g. SOP, NATOPS, BML etc). I.e., good enough to get by, but not necessarily a satisfactory competency level as judged by your own standards.
Satisfactory competency level meets the baseline requirements *and* is a competency level satisfactorily as judged by your own standards.

On average - How would you rate your workload to maintain a baseline competency level as a military aviator? ***(WL1)***
Consider only the workload and competency level for the flying-related portion of your job.
1 = Very low workload
10 = Very high workload

On average - How would you rate the workload to complete your non-flying related duties? ***(WL2)***
Consider only the workload for the non-flying related portion of your job.
1 = Very low workload
10 = Very high workload

On average - How would you rate the cumulative overall workload based on the two previous questions? ***(WL3)***
Baseline competency level and completing your non-flying related duties.
1 = Very low workload
10 = Very high workload

Consider this statement: "There is enough time available during regular working hours to maintain a baseline competency level as a military aviator *and* to complete my non-flying related duties" ***(WL4)***
1 = Strongly disagree
10 = Strongly agree

Consider this statement: "I am able to fully focus on flying activities when such activities are being conducted" ***(WL5)***
I.e. non-flying related duties does not distract/demand your attention when preparing/conducting/de-briefing flights or conducting other flying related training/studying

1 = Strongly disagree
10 = Strongly agree

On average - How would you rate your workload to acquire/maintain a satisfactory competency level as a military aviator? *(WL6)*
Consider only the workload and competency level for the flying-related portion of your job.
1 = Very low workload
10 = Very high workload

Consider this statement: "I have to study outside of working hours to acquire/maintain a satisfactory competency level as a military aviator" *(WL7)*
1 = Strongly disagree
10 = Strongly agree

Currency (CUR)

How many years since last time you were a fulltime flight student? *(CUR1)*
Initial flight training or conversion course (e.g. P-8/F-35 conversion)

How many years since last time you did a pilot upgrade? (CUR2)
E.g. upgrading from CP to PPC (Besetningssjef), wingman to formation lead etc

How many military flight hours did you acquire in 2023? *(CUR3)*
A rough estimate is acceptable (e.g. ~360hrs)

How many days since your last flight? *(CUR4)*

How many days are in average between your flights? *(CUR5)*
Only count working days

How do you rate your currency at the present moment? *(CUR6)*
Consider only flying related currency.
1 = Least current
10 = Most current

Consider this statement: "I am able to fullfill the annual currency requirements for flying" *(CUR7)*
Consider only flying-related currency requirements set forth by RNoAF/Squadron regulations and flight manuals.
1 = Strongly disagree
10 = Strongly agree

Consider the following statement: "I feel uncomfortable in the aircraft after having a longer break from flying" *(CUR8)*
1 = Strongly disagree
10 = Strongly agree

Competency (COMP)

Competency in this regard refers to the cumulative sum of your skills, attitude and knowledge to successfully operate as a military aviator.

In your opinion - what is your competency level as a military aviator? *(COMP1)*
Consider only the flying-related part of you job.

1 = Least competent
10 = Most competent

How *satisfied* are you with your competency level as an aviator? *(COMP2)*
1 = Least satisfied
10 = Most satisfied

Are there any other factors you would like to mention that strongly influence your competency level? *(COMP3)*
If possible, give a rating based on a 1-10 scale

Performance (PERF)

In your opinion - what is the *expected* level of performance directed towards you as a military aviator? *(PERF1)*
Expectations as projected by squadron leadership and colleagues.
1 = Very low expectations
10 = Very high expectations

In your opinion - what is your *percieved* level of performance as a military aviator? *(PERF2)*
1 = Very low performance
10 = Very high performance

How satisfied are you with your performance as a military aviator? *(PERF3)*
1 = Very low satisfaction
10 = Very high satisfaction

Overall (OA)

How do you rate your overall competency level? *(OA1)*
Overall competency, including both flying and non-flying related duties.
1 = Very low competency
10 = Very high competency

How do you rate your overall job performance? *(OA2)*
Overall performance, including both flying and non-flying related duties.
1 = Very low performance
10 = Very high performance

How satisfied are you with your job overall? *(OA3)*
Overall impression, including both flying and non-flying related duties.
1 = Very low / I can't get no (satisfaction)
10 = Very high satisfaction

Finishing remarks

If you have any other remarks that you feel are worth mentioning, please feel free to elaborate in the textbow below.

Generated: 2024-05-28 21:12:50.

UiT The Arctic University of Norway