



FORSVARET
Forsvarets høgskole

**Luftforsvarets flygerseleksjon: En studie av
prediktiv validitet**

Harald Grindheim

Masteroppgave
Forsvarets høgskole
vår 2023

Forord

Som en avsluttende del av Stabsskolen ved Forsvarets høgskole har jeg det siste året fått anledning til å fordype meg i et tema innenfor seleksjon av flygere i Luftforsvaret. Etter selv å ha gått gjennom datidens seleksjonsprosess for godt og vel to tiår siden, og noe senere gjennom stilling som instruktør ved flygeskolen på Bardufoss, har jeg fått en førstehånds befatning med seleksjonsprosessen som har vekket en interesse og nysgjerrighet for metodene som brukes. Da jeg i forberedelsene til masteroppgaven forsto at det fantes et reelt behov for en ny undersøkelse av opptaksprøvenes prediktive validitet, basert på et eksisterende datasett som ville kunne gjøres tilgjengelig, var det naturlig for meg å gå i den retningen. I en samtale med forsvarssjefen høsten 2022 gav han meg ett råd: Gjør det enkelt for deg selv. Det har jeg forsøkt å gjøre gjennom å holde meg til én, ganske smal, hovedproblemstilling og jobbe godt med den, heller enn å la oppgaven bli for omfattende og komplisert.

Takk til sjefpsykologen i Forsvaret, Ole Christian Lang-Ree, for idéer, tips og inspirasjon, og ikke minst innsatsen for å tilgjengeliggjøre datasettet for min oppgave. Takk til min enestående veileder Monica Martinussen, som har hjulpet meg mye i arbeidet. Monica er virkelig en av verdens fremste eksperter på dette feltet, og har produsert mye av forskningen som denne oppgaven bygger på. Takk også til bidragsytere ved flygeskolen, ved seleksjonscenteret, ved Forsvarets høgskole og til min arbeidsgiver 335 skvadron for god tilrettelegging. Det har vært et privilegium å få avsatt tid og ressurser til å kunne fokusere fullt og helt på arbeidet med masteroppgaven.

Oslo, april 2023

Harald Grindheim

Sammendrag

Denne studien undersøker prediktiv validitet av testene som benyttes for seleksjon av kandidater til militær flygerutdanning i Luftforsvaret, gjennom å sammenligne testresultater med prestasjoner fra Luftforsvarets flygeskole på Bardufoss. Datagrunnlaget innbefatter 505 kandidater som har fullført opptaksprøvene og har startet som flyaspirant ved flygeskolen over en 15-årsperiode. Som prediktorer i undersøkelsen benyttes resultater av papir- og blyanttester i Fase 1, databaserte tester i Fase 2a og intervjuprognoser i Fase 2b, og som kriterier benyttes ulike kontinuerlige prestasjonsmål for teoretiske fag og praktisk flyging, i tillegg til det todelte bestått/ikke-bestått-kriteriet som er mest vanlig å benytte i denne typen studier. Analyser av korrelasjoner og regresjon gir funn som indikerer at testene i rimelig grad er egnet til å predikere flygeskoleprestasjoner, i tråd med forventningene til studien. Sammenhengene er sterkest for prestasjonskriterier innen teoretiske fag, og noe svakere for prestasjonskriterier innen praktisk flyging. For de fem hovedkategoriene av kognitive og psykomotoriske tester finnes det klare sammenhenger med prestasjonskriteriene, med unntak av testkategorien som kalles *Informasjonsprosesseringsevne*. Denne kategorien tester ser ikke ut til å predikere flygeskoleprestasjoner, og det kan hevdes å være oppsiktsvekkende sett i lys av at nyere flytyper stiller økte krav til denne typen evner. Studien kan bidra til en fornyet og oppdatert innsikt i validiteten av seleksjonsprosessen til flygerutdanningen i Luftforsvaret.

Nøkkelord: Flygerseleksjon, flygerutdanning, prediktiv validitet, valideringsstudie, kognitive og psykomotoriske tester.

Summary

This study examines the predictive validity of the tests used for the selection of Norwegian Air Force pilot training candidates, by comparing the test results and results from the initial pilot training in Norway. The data consists of 505 candidates who passed the selection tests and were admitted into initial training over a 15-year period. The predictors in this study are various paper- and pencil tests, computer-based tests and interviews, and the criteria are continuous assessments of practical flying performance and academic performance in addition to the more commonly used dichotomous pass/fail criterion.

Correlation- and regression analysis indicate that the tests used have predictive validity in line with earlier research. The connections between predictors and criteria are clear when it comes to predicting academic performance, and a little less clear when it comes to practical flying performance. All five categories of cognitive and psychomotoric tests have clear connections to the criteria, except from the tests which measure *Information Processing ability*. This category of tests appears to have no ability to predict future flight training performance, which is interesting considering that newer aircraft technology is believed to be more demanding requiring higher levels of this category of abilities. The study can provide an updated insight into the validity of the pilot selection process in the Norwegian Air Force.

Key words: Pilot selection, pilot training, predictive validity, validation study, cognitive and psychomotoric tests.

Innholdsfortegnelse

1 Innledning	1
1.1 Hva handler studien om?	1
1.2 Aktualisering.....	2
1.2.1 Hvorfor er flygerseleksjon viktig for luftfart generelt?	2
1.2.2 Hvorfor er flygerseleksjon viktig for Luftforsvaret?	3
1.2.3 Hvordan har flygerseleksjon utviklet seg historisk?	4
1.2.4 Hvordan utvikler flygerseleksjon i Luftforsvaret seg?	4
2 Teori	6
2.1 Generelt om valideringsstudier	6
2.2 utfordringer med valideringsstudier.....	7
2.2.1 Et lite utvalg.....	7
2.2.2 Et selektert utvalg.....	7
2.2.3 Valg av kriterier	8
2.3 Lignende valideringsstudier	9
2.4 Meta-analyser.....	12
2.5 Jobbanalyser.....	13
2.6 Luftforsvarets valg av prediktorer	14
3 Hvordan seleksjonen gjennomføres	15
3.1 Generelt.....	15
3.2 Søknad og initiell seleksjon	15
3.3 Opptaksprøver Fase 1.....	16
3.4 Opptaksprøver Fase 2.....	16
3.5 Opptaksprøver Fase 3.....	16
3.6 Opptaksråd	17
3.7 LFS.....	17
4 Problemstilling	20
4.1 Forventninger til funn.....	21
4.2 Avgrensninger.....	21
5 Metode	23
5.1 Deltakere	23
5.2 Prediktorer	24
5.2.1 Inngangsverdier for opptaksprøver	24
5.2.2 Opptaksprøver Fase 1, papirbasert	24

5.2.3 Opptaksprøver Fase 2a, databasert	25
5.2.4 Opptaksprøver Fase 2b, intervju	28
5.3 Kriterier.....	28
5.4 Prosedyre.....	30
5.5 Statistiske analyser	30
5.5.1 Bivariate korrelasjoner	30
5.5.2 Hierarkisk multippel regresjon	30
6 Resultater.....	32
6.1 Bivariate korrelasjoner	32
6.2 Hierarkisk multippel regresjon	35
7 Diskusjon.....	37
7.1 Diskusjon delproblemstilling 1	37
7.1.1 Korrelasjoner for hvert av de fem kriteriene	37
7.1.2 Prediktorer som skiller seg ut.....	38
7.1.3 Funn sett i sammenheng	39
7.2 Diskusjon delproblemstilling 2	40
7.2.1 Korrelasjoner for hvert av de fem kriteriene	40
7.2.2 Regresjon	41
7.2.3 Funn sett i sammenheng	41
7.3 Diskusjon delproblemstilling 3	42
7.4 Styrker og svakheter ved studien	44
7.4.1 Reliabilitet generelt	44
7.4.2 Reliabilitet opptaksprøver (prediktorer)	44
7.4.3 Reliabilitet prestasjonsmål ved LFS (kriterier).....	45
7.4.4 Utvalg.....	46
7.4.5 Valg av prediktorer og kriterier	46
7.4.6 Analyser	47
7.4.7 Redusert detaljgrad	47
7.4.8 Egne bias.....	48
8 Konklusjon	49
9 Praktisk nytteverdi og videre studier.....	51
Forkortelser	53
Litteraturliste	54
Vedlegg	58
Godkjenning fra FHS forskningsnemnd.....	58

1 Innledning

1.1 Hva handler studien om?

Denne studien handler om seleksjon av kandidater til flygerutdanningen i det norske Luftforsvaret. Kjernen i seleksjon kan sies å være det å velge ut den eller de, blant flere aktuelle kandidater, som anses å ha de beste forutsetningene for å møte fremtidige jobbkra. Dette foregår i mange ulike sammenhenger, eksempelvis et jobbintervju der en arbeidsgiver forsøker å finne den kandidaten man anser har den høyeste sannsynligheten for å kunne levere det bedriften trenger, opptak til skoler hvor kandidatene med de høyeste karaktersnittene fra tidligere utdanning foretrekkes fremfor de med lavere snitt, eller på et stevнемøte der kandidatene vurderer hverandre opp mot sin forestilling av hva som vil være viktige egenskaper hos en fremtidig partner.

I Forsvaret er bruk av seleksjon meget utbredt. Opptak til ulike skoler og utdanningsløp foregår ved bruk av ulike seleksjonsprosesser (Torjussen & Hansen, 1999). Jobbene som disse skolene og utdanningsløpene leder frem til vil på mange områder være vidt forskjellige, selv om de eksisterer innenfor samme etat og har samme overordnede formål knyttet til de oppgavene Forsvaret skal løse. Eksempelvis er det mulig å se for seg at egenskapene som trengs for å bli en god marinejeger på en del områder vil være forskjellige fra de egenskapene som trengs for å bli en god flyger. Prosessen med å *finne* disse rette egenskapene – seleksjonen – vil derfor også være forskjellig for ulike skoler og utdanningsløp i Forsvaret. Det som er felles er at de har som mål å predikere hvilke kandidater som har de beste forutsetningene for å prestere i den stillingen eller skolen man søker på (Martinussen & Hunter, 2018).

En god seleksjonsprosess innebærer en vitenskapelig tilnærming, anvendelse av de beste tilgjengelige metodene og grundig utførelse i henhold til en høy profesjonell standard (American Educational Research Association, 2014; European Federation of Psychologists' Associations, 2013; International Air Transport Association, 2019). Dette kalles gjerne *evidensbasert seleksjon*, og en sentral del av dette er å undersøke sammenhengene mellom variablene som brukes som *prediktorer* i seleksjonen, og *kriterier* for oppnådde prestasjoner i jobben i den andre enden (Martinussen, 2017). Prediktorer kan typisk være resultater fra opptaksprøver, personlighetstester, intervju eller tidligere skoleprestasjoner, mens kriterier vil være en form for måling av hva kandidaten har oppnådd i jobben eller hvor godt vedkommende presterer. En undersøkelse av sammenhengene mellom disse, som har som mål å belyse prediktorenes individuelle og samlede forklaringskraft, betegnes gjerne som en *undersøkelse av prediktiv validitet*, eller *valideringsstudie*.

Denne studien tar for seg seleksjonsprosessen som har blitt benyttet for opptak til flygerutdanningen i det norske Luftforsvaret i en lengre periode frem til år 2020, ved å undersøke den prediktive validiteten av opptaksprøvene i den samme perioden. Dette gjøres gjennom statistiske analyser av sammenhengene mellom resultater fra opptaksprøvene og resultater fra Luftforsvarets flygeskole (LFS). Opptaksprøveresultatene er altså prediktorer, mens resultatene fra LFS er kriterier i undersøkelsen. På denne måten vil det være mulig å avdekke de ulike testenes grad av egnethet som prediktor for flygeskoleprestasjoner, og studien kan derfor hevdes å være en kvalitetssikring av opptaksprøvene.

Det overordnede formålet med denne studien er å fremskaffe kunnskap som kan bidra til å danne et forbedret grunnlag for kvalitetssikring og videreutvikling av Luftforsvarets flygerseleksjonsprosess gjennom å undersøke den prediktive validiteten til testene som anvendes i dag.

1.2 Aktualisering

1.2.1 Hvorfor er flygerseleksjon viktig for luftfart generelt?

I luftfart generelt gir det å selektere de riktige menneskene til å bemanne flymaskinene en rekke fordeler. Martinussen og Hunter (2018) argumenterer for at en god seleksjon ligger til grunn for at flyselskapet eller virksomheten kan drive trygge og effektive operasjoner og ha fornøyde kunder. En god seleksjon gir et lavere frafall i utdanningen, noe som åpenbart har et økonomisk aspekt. Også for enkeltmennesket gir det mening å ha en jobb man passer til, som er passe utfordrende og hvor man blir verdsatt for å gjøre en god jobb. På den måten kan man se for seg at seleksjonen bidrar til trivsel og lengre ståtid i jobben. Forfatterne hevder at et velfungerende system for seleksjon på sikt bidrar til en mer effektiv og robust organisasjon (Martinussen & Hunter, 2018).

IATA beskriver hvordan evnetester kan benyttes for seleksjon innen kommersiell luftfart, og trekker frem mange av de samme grunnene til at dette er en god investering, oppsummert med økt sikkerhet, reduserte treningsutgifter, økt grad av suksess i trening og operasjoner, et bedre arbeidsmiljø, økt ståtid og en styrking av virksomhetens merkevare (International Air Transport Association, 2019).

1.2.2 Hvorfor er flygeseleksjon viktig for Luftforsvaret?

Det er rimelig å se for seg at mange av fordelene med en god seleksjonsprosess som eksisterer i luftfarten generelt, også gjelder for Luftforsvaret. I tillegg kan det være verdt å nevne noen aspekter knyttet til at Norge er et lite land med et lite luftforsvar, og til det at Luftforsvaret er en del av militærmakten i en demokratisk stat.

Luftforsvaret er ikke gitt ressurser til å drive grunnleggende flygerutdanning selv, i stedet kjøpes det plasser i USA som gir det norske personellet kompetansen som trengs for å operere alle de respektive fly- og helikoptertypene i Luftforsvaret. Med et såpass lite volum, kan man se for seg at det ville blitt uforholdsmessig dyrt å drive en egen nasjonal utdanning til dette formålet, samtidig som man trolig da ville gått glipp av mange av de andre fordelene en utdanning i USA gir, som styrking av relasjoner bilateralt med USA og innad i NATO-alliansen, interoperabilitet og høy kvalitet i utdanningen. Når Norge da velger å sende elever til USA på flygerutdanning, er det en økonomisk kostnad knyttet til dette, ikke bare i form av selve utdanningsplassene, men også det som er relatert til reise og livsopphold. Dette er kostnader som i liten grad lar seg refundere om eleven ikke lykkes i å fullføre utdanningen. Norge har derfor en økonomisk interesse av at sannsynligheten for at elevene fullfører utdanningen er høy. Fra et norsk perspektiv blir utgifter knyttet til seleksjon lave sett i forhold til potensielle tap knyttet til frafall i utdanningen (Martinussen & Hunter, 2018).

Som et lite luftforsvar i et lite land, vil ulykker som fører til tap av materiell og personell få en stor prosentvis innvirkning på Luftforsvarets samlede kampkraft. Det er derfor rimelig å hevde at Norge, i større grad enn andre land med et større luftforsvar, vil være avhengig av en høy grad av sikkerhet og ergo lav sannsynlighet for tap av materiell og personell i ulykker, for å bevare sin kampkraft. En grundig seleksjon kan derfor være en meget god investering (Torjussen & Hansen, 1999).

Som del av en etat (Forsvaret) som på vegne av en vestlig demokratisk stat (Norge) forvalter et mandat for systematisk utøvelse av voldsmakt, er Luftforsvaret avhengig av at befolkningen har tillit til at mandatet forvaltes på en god og forsvarlig måte. En god seleksjon kan hevdes å være et viktig bidrag til dette, både gjennom å sikre at riktig personell bemanner systemene, og gjennom å underbygge befolkningens tillit til at seleksjonen er grundig.

Luftforsvaret bruker relativt mye ressurser på en grundig og god seleksjon av flygere, og det kan synes å være mange grunner til å hevde at det er en lønnsom investering.

1.2.3 Hvordan har flygerseleksjon utviklet seg historisk?

Seleksjon av flygere har foregått i en eller annen form siden menneskeheten for første gang tok steget inn i bemannet luftfart i begynnelsen av forrige århundre (Hunter, 1989). De første psykologiske testene ble tatt i bruk for flygerseleksjon under første verdenskrig, og det kan synes å eksistere en meget sterk kobling mellom fagfeltet psykologi og flygerseleksjon som strekker seg helt tilbake til den perioden. Martinussen (2005) gir i sin oversiktsartikkel et innblikk i hvordan seleksjon av denne kategorien personell i Vesten har utviklet seg gjennom tiden. Hun peker på at ulike prediktorer har vært benyttet, både rene evnetester, men også ulike former for personlighetstester, biografiske data, kunnskap om flyging og tidligere flyerfaring. I takt med at datateknologi i siste del av forrige århundre ble bedre, billigere og mer tilgjengelig, gikk de fleste vestlige land over til å bruke denne teknologien i seleksjon, for helt eller delvis å erstatte tradisjonelle papir- og blyanttester. I nyere tid har mange land gått i retning av å anvende en trinnvis seleksjonsprosess der både enkle og mer komplekse kognitive tester anvendes sammen med mål på motivasjon. Fordelen med en slik trinnvis prosess er at man slipper å teste alle kandidatene med alle testene, og man oppnår dermed reduserte kostnader (Martinussen, 2005). Testutviklingen og de aller fleste valideringsstudiene har blitt utført i militær regi og er basert på søkere til flygerutdanning i forsvaret (Martinussen, 1996).

1.2.4 Hvordan utvikler flygerseleksjon i Luftforsvaret seg?

Koblingen mellom psykologi og flygerseleksjon kan for det norske Luftforsvarets del spores tilbake til 1947 da de første psykologiske testene ble tatt i bruk (Martinussen & Torjussen, 1998). Testene som brukes har siden blitt oppdatert og validert en rekke ganger (Martinussen & Torjussen, 1998; Torjussen & Hansen, 1999). Databaserte tester ble først introdusert i 1996 på eksperimentell basis, og brukt operativt i seleksjonen fra 1998 (Torjussen & Hansen, 1999).

Seleksjonsprosessen som har vært anvendt på utvalget i denne studien presenteres i nærmere detalj i et senere kapittel. Jevnt over har norske elever hatt et lavt frafall og oppnådd gode resultater ved de amerikanske utdanningsinstitusjonene hvor de har fått sin flygerutdanning, og dette kan hevdes å være dokumentasjon for at Luftforsvarets seleksjon treffer godt.

Gjeldende fra 2022 er det etter føringer fra sjef Luftforsvaret foretatt endringer i kravene som stilles til kandidatene ved LFS, som har til hensikt å øke antallet selekterte kandidater. Den påkrevde kurven til progresjon kan sies å ha blitt gjort noe slakere gjennom at krav til tilfredsstillende utførelse av visse momenter under praktisk flyging har blitt skjøvet til punkter lengre ute i programmet¹.

¹ Samtale med hovedinstruktør LFS, 2022

Bakgrunnen for dette er identifiserte personellbehov i Luftforsvaret, vurdert opp mot hvilken grad av kvalitet det bør være nødvendig å kreve av fremtidige militære flygere. Endringene kan komme til å medføre at resultater oppnådd ved LFS før og etter 2022 ikke lar seg sammenligne direkte. Denne studien tar for seg et tallmateriale som strekker seg frem til endringene fant sted og derfor er uberørt av disse, og vil slik sett kunne være en verdifull benchmark for fremtidige studier som forsøker å analysere virkningene av de foretatte endringene.

2 Teori

Det eksisterer store mengder forskning på seleksjon av flygere og dette kapittelet vil ikke kunne romme inngående beskrivelser av alt som kan tenkes å ha relevans. Hensikten med kapittelet er å trekke opp de store linjene og belyse elementer og studier som er sentrale i å danne et teoretisk fundament for denne studien.

2.1 Generelt om valideringsstudier

Som beskrevet i innledningen innebærer en god seleksjonsprosess at den har en vitenskapelig forankring gjennom å basere seg på den beste tilgjengelige dokumentasjonen. Testene som brukes må være egnet til å predikere prestasjoner senere i utdanning eller karriere, og kvalitetssikring av testene innebærer derfor å søke evidens for at testene er egnet til dette formålet. Publikasjonen *Standards for Educational and Psychological Testing* beskriver retningslinjer for hvilke krav som bør stilles til tester som benyttes til seleksjon av personell (American Educational Research Association, 2014). Der understrekes viktigheten av å fremskaffe evidens for systematiske sammenhenger mellom testskåre og ett eller flere utfallskriterier for jobbprestasjon. Dette kalles kriterierelatert evidens, og prediktiv validitet er én type kriterierelatert evidens. Dersom det er ikke foreligger kriterierelatert evidens, kan innholdsrelatert- eller begrepsrelatert evidens også benyttes som et steg på veien (Carretta & Ree, 2003). I praksis betyr det at to av kravene til en god test bør være at den faktisk måler nøyaktig og presist (reliabilitet), og at den eller de egenskapene som måles har betydning for fremtidig jobbprestasjon (prediktiv validitet). Å finne evidens for prediktiv validitet kan gjøres gjennom en lokal valideringsstudie som denne, eller gjennom å sammenligne en større mengde data på tvers av flere studier i en meta-analyse. Forskningen på flygerseleksjon har i stor grad foregått på militære flygere, og fokus har vært på utvikling av tester og søken etter evidens for prediktiv validitet (Carretta & Ree, 2003; Hilton & Dolgin, 1991; Hunter, 1989; Hunter & Burke, 1995).

Evidens for prediktiv validitet påvises gjennom korrelasjoner mellom prediktorer og kriterier. For å avdekke disse korrelasjonene i en lokal valideringsstudie er det vanlig å benytte både en bivariat korrelasjonsanalyse for prediktorene enkeltvis eller gruppevis, og en regresjonsanalyse som forsøker å avdekke i hvor stor grad flere prediktorer til sammen kan bidra til å forklare variasjonen i kriteriene (Martinussen & Hunter, 2018).

Korrelasjoner som påvises i denne typen undersøkelser kan kategoriseres etter størrelse som *små* (*small*, = .10), *middels* (*medium*, = .30) eller *store* (*large*, = .50), i henhold til tommelfingerregler beskrevet av Cohen (2013).

2.2 utfordringer med valideringsstudier

Det finnes noen typiske utfordringer som hefter ved valideringsstudier innenfor flygerseleksjon, der utvalget som undersøkes ofte er lite og strengt selektert, og prestasjonskriteriene er av varierende karakter fra studie til studie.

2.2.1 Et lite utvalg

En lav N gjør det vanskelig å oppnå en statistisk styrke (*statistical power*) som er tilstrekkelig til å kunne si med rimelig grad av sikkerhet at sammenhengene faktisk eksisterer og ikke skyldes tilfeldigheter. Jo høyere N , jo høyere statistisk styrke, og jo høyere lit kan man feste til at de påviste korrelasjonene er reelle og på den bakgrunn kan brukes som evidens for prediktiv validitet.

2.2.2 Et selektert utvalg

At utvalget er (potensielt fremtidige-) flygere som i flere ledd er selektert etter minimumskrav på en rekke variabler og dermed utgjør en på mange områder homogen gruppe, gjør at spredningen i oppnådde skårer blir en del mindre enn den ville ha blitt dersom hele søkergruppen hadde blitt sendt gjennom opptaksprøver og flygerutdanning og blitt målt på de samme variablene. Korrelasjonen mellom prediktor og kriterium i valideringsstudien (observert korrelasjon) blir derfor langt lavere enn det som ville vært tilfellet for en studie av hele gruppen (sann korrelasjon). Denne problematikken kalles *Restriction of Range* (RR) og øker i styrke for hvert ledd av seleksjonen (Martinussen & Hunter, 2018). Det finnes metoder for å korrigere for RR ved å regne om observert korrelasjon til estimerer for sann korrelasjon. En vanlig metode er beskrevet av Dunbar og Linn (Dunbar & Linn, 1991). Det er vanlig å foreta denne korreksjonen i forbindelse med utførelse av meta-analyser for å gjøre korrelasjonsverdiene mest mulig sammenlignbare på tvers av flere studier. RR-korreksjon er en avansert regneoperasjon med en rekke forutsetninger som vil ikke bli foretatt i denne studien. Påviste korrelasjoner må sees i lys av dette og representerer underestimat av de sanne korrelasjonene.

2.2.3 Valg av kriterier

I valideringsstudier benyttes gjerne testresultater fra opptaksprøvene som prediktorer, og her finnes ofte et rikt datamateriale som gir et detaljert bilde av kandidatens ulike evner og trekk. Å finne gode kriterier kan derimot ofte synes å være vanskeligere. I valideringsstudier er det viktig at kriteriene på prestasjoner både er relevante for organisasjonen, har god begrepsvaliditet og reliabilitet (Martinussen, 2005). Kriterier som gjenspeiler prestasjoner i den jobben som skal utføres – som i dette tilfellet vil være mål for hvor godt man utfører jobben som flyger – er kanskje det som vil være mest relevant å bruke i seleksjonsforskning og som forskere bør være mest interessert i. Likevel foregår det meste av forskningen innen seleksjon i luftfart ved bruk av kriterier innhentet under utdanning. Dette kan henge sammen med at data for jobbprestasjoner i mange tilfeller er mindre tilgjengelig eller systematisert, og i større grad må innhentes og systematiseres av forskeren selv, noe som i seg selv kan være ressurskrevende for både forskeren og miljøet det forskes på.

En del av studiene som har blitt gjennomført på flygerseleksjon benytter et dikotomt prestasjonskriterium i form av *bestått/ikke bestått (pass/fail)* utdanning. Carretta (1992) hevder at forholdet mellom variablene som brukes for seleksjon og senere flygerprestasjoner kan forstås mye bedre dersom det i stedet for *pass/fail* benyttes et kontinuerlig mål. Et tall på en skala inneholder mer informasjon og kan på den måten bidra til å avdekke større korrelasjoner enn det et dikotomt mål vil kunne gjøre. Carretta argumenterer for at prestasjonskriterier bør inneholde så mye data om mulig for å oppnå mer stabile prestasjonsindikatorer. Martinussen og Hunter (2018) forklarer at en korreksjon i korrelasjon for dikotomt prestasjonskriterium også kan beregnes matematisk, dersom forholdstallet mellom *bestått* og *ikke bestått* er kjent. Jo lengre unna dette forholdstallet er fra en 50/50-fordeling, jo større blir korreksjonen.

Usikkerhet i målingen av kriterier (kriteriereliabilitet) er også en faktor som trekkes frem av Martinussen og Hunter (2018). Det er en rekke forhold som kan påvirke kriterierelabiliteten, og den kan derfor være vanskelig å fastslå. I noen studier er den oppgitt, og det er mulig å gjøre matematiske korrigeringer for denne usikkerheten, noe som kan være hensiktsmessig særlig i forbindelse med meta-analyser hvor korrelasjoner sammenlignes på tvers av flere studier. I denne studien vil kriteriereliabilitet ikke bli beregnet, men forhold som kan påvirke denne vil bli beskrevet som mulige svakheter ved studien i et senere kapittel.

2.3 Lignende valideringsstudier

I forbindelse med overgangen til databaserte seleksjonstester i Luftforsvaret, gjennomførte Martinussen og Torjussen (2004) en undersøkelse av det testbatteriet som i hovedsak har vært benyttet til seleksjon siden (i opptakets Fase 2a). I tillegg til en faktoranalyse som er grunnlaget for den kategoriseringen av tester som fremdeles er gjeldende, undersøkte de testenes prediktive validitet opp mot kriteriene *bestått/ikke bestått uttaksflyging* og to ulike former for instruktørvurderinger (behaviourally anchored rating scales, BARS). Deltakere i studien var 108 kandidater som alle hadde bestått hele seleksjonen. For kriteriet *bestått/ikke bestått* avdekket undersøkelsen en signifikant korrelasjon med gjennomsnittsskåre for de 13 testene ($r = .26$, korrigert for RR). Flere av enkelttestene hadde signifikante korrelasjoner med ett eller flere kriterier. De sterkeste korrelasjonene i studien ser ut til å være knyttet til testene *Attention*² og *DTG*³, som har r på henholdsvis $.31$, $.42$ og $.37$ (*Attention* opp mot *bestått/ikke bestått* og de to BARS-kriteriene) og $.31$, $.33$ og $.08$ (*DTG* opp mot *bestått/ikke bestått* og de to BARS-kriteriene), korrigert for RR. Testene *Planes*⁴, *Numbers*⁵ og *Patterns* hadde lave r -verdier (mellom $-.04$ og $.11$) og ser ikke ut til å være egnet til å predikere noen av de tre kriteriene. *Patterns* er ikke lenger i bruk og er heller ikke en del av min studie, det er derimot *Planes* og *Numbers*.

Martinussen og Torjussen (1998) gjennomførte tidligere også en studie som blant annet undersøkte prediktiv validitet for testene som da ble brukt til flygerseleksjon i Luftforsvaret. Dette skjedde før overgangen til databasert teknologi, og tok utgangspunkt i testbatteriet som var i bruk den gang. Deltakere i studien var 230 kandidater som i løpet av perioden mellom 1988 og 1992 startet på LFS. Datidens tester er prediktorer i studien, og som kriterier brukes *bestått/ikke bestått* og en skåre beregnet som et gjennomsnitt av fire ulike teorifag. Sammensetningen og utformingen av datidens tester skiller seg noe fra testene som er aktuelle for min studie, så det vil ikke være mulig å gjøre en direkte sammenligning opp mot funn fra min studie. Dog er noen av enkelttestene sammenlignbare, og det er derfor interessant å studere hvilke korrelasjoner studien kunne påvise for disse prediktorene. Opp mot kriteriet *bestått/ikke bestått* finner man signifikante korrelasjoner for *flykunnskap*⁶ ($r = .23$), *engelsk*⁷ (todelt test, $r = .22$ og $.17$) og *instrument comprehension*⁸ ($r = .24$).

² Omtalt som *prediktor 23* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

³ Omtalt som *prediktor 27* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

⁴ Omtalt som *prediktor 19* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

⁵ Omtalt som *prediktor 25* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

⁶ Omtalt som *prediktor 13* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

⁷ Omtalt som *prediktor 07* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

⁸ Omtalt som *prediktor 20* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

Korreksjoner for RR, kriterierelabilitet og dikotomt kriterium, gir estimater for sann korrelasjon på henholdsvis $r = .31, .33, .28$ og $.40$. Opp mot kriteriet *teoriskåre* finner man signifikante korrelasjoner for *tallrekker*⁹ ($r = .23$), *engelsk*¹⁰ (todelt test, $r = .18$ og $.26$), *sortering*¹¹ ($r = .15$), *retningsreversaller* ($r = .16$) og *numbers*¹² ($r = .18$). Korreksjoner for RR og kriterierelabilitet gir estimater for sann korrelasjon på henholdsvis $r = .30, .22, .33, .20, .20$ og $.20$.

En valideringsstudie publisert i 2010 (Lang-Ree et al., 2010) undersøkte korrelasjoner mellom de databaserte seleksjonstestene (i opptakets Fase 2a) og kriteriet *bestått/ikke bestått uttaksflyging* for et utvalg på 208 kandidater som hadde startet på LFS. Utvalget i studien inngår i utvalget i min studie, og variablene som undersøkes finnes blant variablene i min studie. Prediktorene i studien er identiske med prediktorene som benevnes 15 – 28 i min studie¹³, og kriteriet i studien finnes som kriterium 39 i min studie¹⁴. 2010-studien er på den måten sammenlignbar med min studie og derfor særlig interessant. Studien avdekker korrelasjoner som på det meste kan kategoriseres som *små*, hvor korrelasjoner for prediktorene *Sensory Motor Apparatus* ($r = .12$), *Instrument Comprehension* ($r = .17$) og *Hands* ($r = .12$) er signifikante. Korreksjon for RR gir korrelasjon på henholdsvis $r = .16, .22$ og $.18$. Studien ser også på korrelasjon for tre kategori-gjennomsnittsskårer som prediktor, hvor *Spatiale evner* er signifikant ($r = .18, r = .22$ korrigeret for RR), mens *Psykomotorisk koordinasjon* og *Informasjonsprosesseringsevner* ikke er signifikant. Forfatterne peker på at mulige forklaringer for de relativt lave korrelasjonene kan være at det brukes et dikotomt kriterium, og at kriterierelabiliteten er ukjent og lav og kan tenkes å være påvirket av forhold som motivasjon, flysøke med mer.

I en nyere studie (Martinussen et al., 2022) på 188 kandidater til sivil flygerutdanning ved Universitetet i Tromsø (UiT) undersøkes den prediktive validiteten av opptaksprøvene som benyttes for seleksjon til den utdanningen. Av studien går det frem at mange av testene som benyttes i opptaket til UiT-utdanningen er identiske med testene som benyttes i Luftforsvaret. På samme måte som i Luftforsvaret benyttes også personlighetstrekk som et mulig redskap i intervju-delen, uten at det eksplisitt selekteres etter dette. Sted for gjennomføring (Luftforsvarets rekrutterings- og seleksjonssenter, LSES, på Rygge), tidsrammer for, innholdet i og kategorisering av testene som brukes er identisk som for Luftforsvaret. Blant prediktorene som benyttes i studien finnes

⁹ Omtalt som *prediktor 04* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

¹⁰ Omtalt som *prediktor 07* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

¹¹ *Sortering* og *retningsreversaller* ble senere slått sammen til én test, omtalt som *prediktor 23 Attention* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

¹² Omtalt som *prediktor 25* i denne studien, se avsnittet *Prediktorer* i Metode-kapittelet

¹³ Se avsnittet *Prediktorer* i Metode-kapittelet

¹⁴ Se avsnittet *Kriterier* i Metode-kapittelet

gjennomsnittsskårer for kategoriene *General mental ability*¹⁵, *Psychomotor Coordination*¹⁶, *Spatial Ability*¹⁷ og *Information Processing*¹⁸, samt en *skole*¹⁹- og en *flygerprognose*²⁰ som bestemmes på bakgrunn av et intervju med en psykolog og en flyger. Kriteriene består av tre ulike prestasjonsmål innen teoretiske fag, og ett prestasjonsmål innen praktisk flyging (*Extra flying hours needed*). Studien avdekker signifikante, små til middels store ($r = .15 - .38$) korrelasjoner mellom alle de nevnte prediktorene og to av de tre kriteriene innen teoretiske fag, korrigert for RR kan korrelasjonene kategoriseres som middels til store ($r = .28 - .54$). Kriteriet innen praktisk flyging har signifikante korrelasjoner med prediktorene *Psychomotor Coordination* ($r = -.28$), *Spatial Ability* ($r = -.34$), totalskåre for de databaserte testene ($r = -.35$) og for alle testene samlet ($r = -.22$). Korrigert for RR er korrelasjonene henholdsvis $r = -.36, -.42, -.44$ og $-.33$. Regresjonsanalysen indikerer at testene til sammen bidrar til å predikere de ulike kriteriene, og at intervjuprognosene tilfører ytterligere forklaringskraft (utover det som kan forklares gjennom resultatene av testene) for variasjonen i teorieprestasjonskriteriene, men ikke for variasjonen i det praktiske flygerprestasjonskriteriet.

Harlems masteroppgave (Harlem, 2016) undersøker prediktiv validitet for de samme opptaksprøvene som er aktuelle i min studie. Han benytter i hovedsak de samme prediktorene, og benytter prestasjoner fra flere nivåer av videre utdanning i USA som kriterier. Studien avdekker små og middels store korrelasjoner, og indikerer at hvilke tester som predikerer prestasjoner endrer seg gjennom ulike faser i utdanningen (med ulike sett av kriterier). Informasjonsprosesseringsevner ser ut til å få en økende betydning jo lenger ut i utdanningen man kommer. Del to av Harlems oppgave er en jobbanalyse av jagerflyet F-35, som søker å avdekke hvilke egenskaper som er viktige for å gjøre en god jobb som flyger på det systemet. Oppgaven finner at mange av testene viser god prediktiv validitet mot de identifiserte egenskapene, særlig gjelder dette enkelttester i kategoriene informasjonsprosessering og spatiale evner.

En amerikansk studie fra 2014 på 9641 kandidater til tre ulike flygerutdanningsløp i US Air Force i en 14-årsperiode mellom 1995 og 2008 hadde mål om å undersøke prediktiv validitet for kognitive evner og personlighetstrekk som prediktorer for prestasjoner under flygerutdanningen (Carretta et

¹⁵ Tilsvaret prediktoren som i min studie benevnes *08 GT*, og i begge studier er et gjennomsnitt av de samme testene

¹⁶ Tilsvaret prediktoren som i min studie benevnes *18 PK*, og i begge studier er et gjennomsnitt av de samme testene

¹⁷ Tilsvaret prediktoren som i min studie benevnes *22 SE*, og i begge studier er et gjennomsnitt av de samme testene

¹⁸ Tilsvaret prediktoren som i min studie benevnes *28 IP*, og i begge studier er et gjennomsnitt av de samme testene

¹⁹ Sammenlignbar med prediktoren som i min studie benevnes *31 Skoleprognose*

²⁰ Sammenlignbar med prediktoren som i min studie benevnes *33 Flygerprognose*

al., 2014). Studien fant at kognitive evner har en langt høyere prediktiv validitet enn personlighetstrekk, noe forfatterne hevder å være i tråd med forventningene til studien på bakgrunn av tidligere forskning.

2.4 Meta-analyser

Meta-analyser bruker statistiske analyser for å sammenligne funn fra en mengde studier innenfor samme tema. Dette gir tilgang på et større datamateriale, som gir en høyere statistisk power enn det som normalt vil være mulig å oppnå for enkeltstudier. På den måten vil meta-analyser kunne bidra til å avdekke sammenhenger som ikke vil være mulig å påvise gjennom en enkeltstudie. Meta-analyser er en fellesbetegnelse på en kvantitativ oppsummering av funn fra tidligere undersøkelser (Hunter & Burke, 1995). Sammenligning av data på tvers av studier kan være utfordrende på den måten at et høyt antall studier kan bidra til å gjøre det helhetlige bildet komplisert og uoversiktlig. Det vil ofte eksistere metodiske forskjeller mellom de ulike studiene som undersøkes, eksempelvis når det kommer til hvordan data er samlet inn, sortert og analysert. Variablene i de ulike studiene kan være noe forskjellig, på den måten at de ikke nødvendigvis har nøyaktig samme betydning, og at meta-analysen derfor vil kunne støte på problematikk knyttet til risiko for sammenlikning av «epler og pærer».

Innenfor flygerseleksjon kan meta-analyser bidra til et bedre bilde av den sanne korrelasjonen mellom prediktor og kriterium, og kan på bakgrunn av det vise gjennomsnittlig prediktiv validitet for ulike tester. Meta-analysene kan bidra til å si noe om hvorvidt den prediktive validiteten lar seg generalisere over ulike situasjoner (Martinussen, 1996).

Det er gjennomført flere større meta-analyser som undersøker prediktiv validitet av opptaksprøver på tvers av en mengde enkeltstudier innenfor flygerseleksjon. I en nyere rapport fra European Association for Aviation Psychology beskrives noen av de mest sentrale (Damos et al., 2022). En av disse er en meta-analyse utført av Hunter & Burke (1994), som tar for seg 68 enkeltstudier gjennomført hovedsakelig i USA og Storbritannia, i perioden fra 1940 og frem til 1990. Det samlede utvalget består av til sammen 437.258 sivile og militære flygere eller flygerkandidater. Hovedfunnene viser gjennomsnittlige verdier for prediktiv validitet som varierer mellom $r = .06$ (for utdanning) og $r = .34$ (for *work sample tests*). Prediktorene generelt evnenivå, verbale evner, finmotorikk (*fine dexterity*), alder, utdanning og personlighet viser ingen generaliserbar validitet for flygerprestasjoner. Blant prediktorene som viser bedre validitet finnes kvantitativ evne, spatial evne, teknisk innsikt

(*mechanical*), flykunnskap, generell kunnskap, grovmotorikk, persepsjonstid, reaksjonstid, biografiske data og *job sample*. Martinussens meta-analyse (1996) har et bredere geografisk omfang enn Hunter og Burke, og tar for seg 66 ulike utvalg fra 50 ulike enkeltstudier fra til sammen 11 ulike land. Det samlede utvalget er ikke oppgitt, men estimeres å ligge mellom 17.900 og 26.000, hvor størsteparten er militære flygere eller flygerkandidater. Korrigert for dikotomt kriterium (*pass/fail*), finner Martinussen gjennomsnittlige verdier for prediktiv validitet for ulike kategorier av tester som varierer mellom $r = .14$ (for personlighet) og $r = .30$ (for flygererfaring før testing) mot kriteriet for flygerprestasjon. Prediktorene *kognitive evner* og *psykomotoriske evner* finnes å ha $r = .24$. En samlet skåre for flere tester oppnår $r = .37$. Meta-analysen viser at alle kategorier tester som ble undersøkt i studien, med unntak av personlighet og biografiske data, har generaliserbar prediktiv validitet for flygerprestasjonskriteriet. En meta-analyse utført av AlMamari og Traynor (2019) undersøker prediktiv validitet for en samleskåre fra hele testbatterier opp mot kriterier for flygerprestasjon, gjennom analyser av 118 ulike utvalg fra 89 ulike studier hovedsakelig fra USA og Canada. Det samlede utvalget er på 116.806 flygere eller flygerkandidater, hvor flesteparten har en tilknytning til militær luftfart. Den valgte metoden medfører at korrelasjoner mellom prediktor og kriterium ikke lar seg korrigere for RR eller reliabilitet. Hovedfunn i AlMamari og Traynors studie viser gjennomsnittlige verdier for (ukorrigert) prediktiv validitet som varierer mellom $r = .10$ og $r = .34$. Den beste prediktoren finnes å være flysimulator, som en *work sample test* ($r = .34$), etterfulgt av teoretisk kunnskap (*Acquired Knowledge*, $r = .19$) og generelt evnenivå ($r = .18$).

Martinussen og Torjussen (1998) publiserte mot slutten av forrige århundre også en meta-analyse av et noe mindre omfang, som undersøkte prediktiv validitet for testene som den gang ble benyttet for seleksjon av flygere i Luftforsvaret. Studien finner at *instrument comprehension* ($r = .29$), teknisk innsikt (*mechanical principles*, $r = .23$) og flykunnskap ($r = .22$) er de beste prediktorene for suksess i flygerutdanningen, blant testene i datidens testbatteri.

2.5 Jobbanalyser

I kapittelet *What Do We Want to Select & Assess?* i rapporten med tittelen *Selection in Aviation* (Damos et al., 2022) beskriver Hörmann hva en jobbanalyse er og hvorfor denne typen studier bør ligge til grunn for en god seleksjonsprosess. Han påpeker at en jobbanalyse ikke er en enkeltstående teknikk, men en prosess som innbefatter innsamling og organisering av informasjon om jobben personellet som selekteres til slutt skal settes til å utføre, og hvilke krav den stiller. Denne informasjonen kan være hentet fra ulike kilder som kritiske hendelser, observasjon av jobbutførelse

og intervjuer og undersøkelser av eksperter innen det aktuelle fagfeltet. Informasjonen som samles inn bidrar til en bedre forståelse av hvordan en jobb utføres og hvilke spesifikke krav som bør hensyntas for å oppnå en hensiktsmessig kobling mellom menneskelige egenskaper og jobben som skal gjøres. Målet med jobbanalysen, i seleksjonssammenheng, er altså å definere et sett menneskelige egenskaper som vurderes å ville bidra til å styrke sannsynligheten for at kandidater vil kunne utføre jobben de selekteres til på en vellykket måte.

Det eksisterer en håndfull ulike fremgangsmåter som normalt benyttes for å gjennomføre jobbanalyser, og blant disse er den som kalles *Fleishman's job analysis survey method* (Fleishman, 1975; Fleishman & Reilly, 1992). Metoden baserer seg på spørreskjema hvor ansatte bes indikere i hvor stor grad de vurderer ulike menneskelige egenskaper å være relevante for prestasjoner i jobben. Metoden har blitt benyttet til flere jobbanalyser innen militær luftfart, og blant disse er en analyse gjort på 43 erfarne jagerflygere fra Canada, Norge og USA med mål om å fastslå 27 ulike menneskelige egenskapers relative viktighet for jagerflygerprestasjoner (Carretta et al., 1993). Jagerflygernes responser danner grunnlaget for en rangering av disse egenskapene, og forfatterne finner at situasjonsforståelse (*situational awareness*), hukommelse, motivasjon og resonneringsevne oppnår de høyeste rangeringene. Lederskap og verbale evner (*oral expression, written comprehension* og *written expression*) oppnår de laveste rangeringene i undersøkelsen.

2.6 Luftforsvarets valg av prediktorer

De kognitive og psykomotoriske egenskapene som måles gjennom seleksjonstestene er delt inn i kategorier på bakgrunn av faktoranalyser (Hansen et al., 1989; Martinussen & Torjussen, 1998, 2004). I datagrunnlaget som er gjeldende for denne studien, er disse egenskapene plassert i fem kategorier som benevnes *generell teorikunnskap, teknisk innsikt og romopplevelse, psykomotorisk koordinasjon, spatiale evner og informasjonsprosesseringsevner*. Hver av disse inneholder en rekke enkelttester, som er nærmere beskrevet i metode-kapittelet. Luftforsvarets system for seleksjon av flygere ser ut til å legge stor vekt på kognitive og psykomotoriske evner som prediktorer, og dette kan hevdes å harmonere godt med forskningen på flygerseleksjon som i stort ser ut til å indikere at de beste prediktorene for flygerprestasjoner er nettopp kognitive og psykomotoriske evner, mens prediktorer innenfor kategorier som personlighet, biografiske data og tidligere erfaring er mindre egnet.

3 Hvordan seleksjonen gjennomføres

3.1 Generelt

Dette kapitlet gir en oversikt over innholdet og rekkefølgen i det løpet som fører fra at en søknad til Luftforsvarets flygerutdanning sendes inn, og frem til at kandidaten har bestått LFS og seleksjonsprosessen med det regnes som avsluttet. Dette løpet er felles for alle Luftforsvarets flytyper, og fordelingen av kandidater på de ulike utdanningsløpene som leder frem til de respektive skvadroner og flytyper, skjer først etter at seleksjonen er ferdig. Søknadskrav og løpet for seleksjon og utdanning har i stort vært uforandret i hele perioden som er aktuell i denne studien.

Utdanningsreformen i Forsvaret (Forsvarsdepartementet, 2018) har medført en endring i form av at befalsutdanning ikke lenger inngår som en obligatorisk del av løpet, uten at dette antas å ville påvirke variablene i denne studien. Alle tester gjennomføres og evalueres individuelt for kandidaten.

Kapitlets referanser er Forsvarets egen informasjonsside på internett for flyger- og navigatørutdanningen (Forsvaret, 2022), *LFS Instruktørhåndbok*²¹ samt samtaler med LFS, LSES og Sjefpsykologen i Forsvaret (2022).

3.2 Søknad og initiell seleksjon

Kandidaten sender inn en søknad til flygerutdanning i Luftforsvaret, gjennom post eller digitale kanaler, innen en gitt tidsfrist som normalt er 1. desember for påfølgende opptaksår.

Søknadene mottas og behandles ved Forsvarets personell- og vernepliktssenter (FPVS), som gjør koblinger mot eksisterende data fra sesjon og foretar en initiell seleksjon basert på formelle søknadskrav. Videre vurderinger som leder frem til en liste over kandidater for innkalling til opptaksprøver gjøres i samråd med LSES. Disse vurderingene kan typisk være knyttet til oppfylling av formelle søknadskrav, skolepoeng eller kandidatens historikk med tidligere søknad eller gjennomføring av opptaksprøver. Normalt vil en kandidat ha anledning til å gjennomføre opptaksprøvene kun én gang.

²¹ *LFS Instruktørhåndbok* versjon 2014 og versjon 2020v2, i nyere utgave kalt *LFS Seleksjonshåndbok 129-51A*.

3.3 Opptaksprøver Fase 1

Gjennomføringen av Fase 1 foregår på utvalgte lokasjoner fordelt rundt i Norge, gjerne i de største byene, der Forsvaret enten kan benytte egne fasiliteter eller har anledning til å leie fasiliteter egnet til dette formålet. Testene gjennomføres suksessivt i løpet av en dag og ved bruk av papir og blyant, og måler kandidatenes kunnskap, intelligens og teoriforståelse innenfor ulike kategorier. Testene karaktersettes og vurderes i etterkant. Statistisk vil litt over halvparten av kandidatene som gjennomfører Fase 1 gå videre og bli innkalt til Fase 2.

3.4 Opptaksprøver Fase 2

Gjennomføringen av Fase 2 foregår ved LSES på Rygge flystasjon.

I det som betegnes *Fase 2a* gjennomføres databaserte tester suksessivt over flere dager, og måler ulike kognitive og psykomotoriske evner. Testene karaktersettes og vurderes fortløpende, og frabeordning av kandidater skjer på bakgrunn av resultater fra hver enkelt testgruppe, altså normalt relativt jevnt og hyppig gjennom perioden. Fysiske tester og medisinsk screening er også en del av Fase 2a, uten at disse omfattes av denne studien.

Fase 2b utgjør den avsluttende delen av Fase 2 og omfatter et intervju og en grundig flymedisinsk undersøkelse ved Flymedisinsk institutt (FMI) i Oslo. FMI-undersøkelsen omfattes ikke av denne studien. Intervjuet gjennomføres av en psykolog og en erfaren flyger, i henhold til en egen guide (Forsvarets Sanitet, Seksjon for militærpsykologi, 2022a), og har en varighet på om lag 60 minutter. Intervjuet søker å avdekke forhold som ikke tidligere er avdekket i opptaksprøvene, knyttet til kandidatens motivasjon, interesser, utdanning, personlighet og lederegenskaper. På bakgrunn av intervjuet fastsettes de tre variablene *skole-*, *offisers-* og *flygerprognose*, samt en samlet konklusjon som er enten *Anbefales*, *Anbefales under tvil*, *Anbefales ikke* eller *Anbefales nytt forsøk/modning*.

3.5 Opptaksprøver Fase 3

Når kandidatene som har bestått flygeropptaket Fase 2 igjen møter ved LSES, er det for gjennomføring av Fase 3, som består av *Forsvarets Opptak og Seleksjon* (FOS). Denne delen foregår over flere dager i feltmessige omgivelser og sammen med øvrige kandidater til de ulike krigsskoleutdanningene i Forsvaret. Normalt utgjør denne fasen et svært lite frafall blant

flygerkandidatene. Kandidater som består FOS returnerer til LSES med en FOS-vurdering som tas med inn i det umiddelbart påfølgende opptaksrådet.

3.6 Opptaksråd

LSES avslutter sin del av seleksjonen med et opptaksråd, bestående av en leder utpekt av Luftforsvarets ledelse, sjef LSES, Sjefpsykologen i Forsvaret, sjef LFS, en representant fra FMI og eventuelt andre ressurser. Rådet foretar helhetlige vurderinger av all tilgjengelig informasjon, og avgjør hvilke kandidater som sendes videre til LFS og får status som *flyaspirant*. Kandidater som det hefter noen form for tvil ved vil måtte forvente å bli kalt inn til en samtale for på den måten å kunne frembringe et forbedret beslutningsgrunnlag.

I alle seleksjonens faser frem til oppstart ved LFS, foretas vurderinger av grenseverdier for å være med videre. I disse vurderingene sees antall gjenværende kandidater i sammenheng med antallet kandidater Luftforsvarets ledelse, med utgangspunkt i sitt personellbehov, ønsker å ha gjennom seleksjonen. Antallet kandidater vil på den måten i noen grad kunne styres i retning av et ønsket måltall. Statistisk kunnskap om forventet frafall i seleksjonens ulike faser bidrar i disse vurderingene. Kravene for å bestå opptaksprøvene er derfor ikke absolutte, men er gjenstand for vurderinger og vil kunne variere noe mellom opptaksår. Styring av antall selekterte kandidater gjennom fleksibilitet i krav er ikke funnet å ha vært gjeldende for LFS sin del av seleksjonen i perioden.

3.7 LFS

Denne delen av seleksjonsprosessen foregår i Bardufoss, omtales gjerne som *uttaksflyging* og har en varighet på om lag seks måneder. Normalt gjennomføres to kull per år ved LFS; et *vårkull* i første halvår og et *høstkull* i andre halvår. Kandidater som ikke tidligere har gjennomført førstegangstjeneste, vil normalt gjennomføre første del av denne i høsthalvåret, og deretter starte uttaksflygingen i januar påfølgende år som del av vårkullet. Høstkullet består av kandidater som allerede har gjennomført førstegangstjeneste og dermed kan starte direkte på uttaksflyging.

LFS sitt oppdrag er seleksjon av kandidater til videre flyger- og navigatørutdanning. Navnet *flygeskole* kan slik sett være misvisende, men må forstås som at skoleaktiviteten er et middel for å drive seleksjon, mer enn et mål i seg selv. Det er dog rimelig å hevde at tilegnelsen av kunnskaper, ferdigheter og holdninger under uttaksflygingen bidrar til å gjøre kandidatene bedre rustet i møtet

med utfordringer under utdanningen i USA og i videre karriere, og derfor kan hevdes å være en verdifull og ønskelig bonuseffekt ved å gjennomføre seleksjon på denne måten.

I programmet gjennomføres teoriundervisning med tilhørende skriftlig eksamen i 9 ulike fag²². Et stykke ut i programmet starter kandidatene med praktisk flyging, slik at dagene og ukene består av en kombinasjon av teori og praktisk flyging. Den første og største delen av det praktiske flyprogrammet kalles *Contact*. Flyturene i Contact-fasen gjennomføres under forhold som tillater visuell kontakt med bakken/overflaten, derav benevnelsen. En eksamenstur²³ som gjennomføres mot slutten av Contact-fasen utgjør en stor del av samlet flykarakter²⁴ og hovedkarakter²⁵, og er dermed en viktig del av det samlede grunnlaget for vurderingen av kandidatens prestasjon. Etter Contact-fasen følger flere kortere faser bestående av navigasjon, instrumentflyging, formasjon og nattflyging. Hvor mye av programmet som vil kunne gjennomføres for et bestemt kull styres blant annet av værforhold og andre faktorer som instruktør- og flytilgjengelighet. De delene av flyprogrammet som etterfølger CD-check regnes ikke som kritisk for å gjennomføre en tilfredsstillende seleksjon.

Kandidater som gjennom sine prestasjoner, enten i teori eller praktisk flyging, viser seg å ha problemer med å følge de relativt strenge kravene til progresjon, vil bli gjenstand for et skoleråd. Skolerådet ledes normalt av en flyger med oberst grad og er sammensatt av ulike ressurser som kan bidra til å skape et helhetlig bilde av kandidatens prestasjoner. Skolerådet foretar en vurdering av all tilgjengelig og relevant informasjon og avgjør hvorvidt kandidaten får fortsette i programmet. Mulige utfall av skolerådet kan være å sende kandidaten videre i programmet, terminering og hjemsendelse, mulighet til å fortsette som navigatør-kandidat, eller terminering med mulighet for å komme tilbake på et senere kull dersom modenhet vurderes som en del av problemet. I denne studien regnes hjemsendte kandidater og navigatørkandidater som *ikke bestått*, mens kandidater som kommer tilbake på et senere kull og består uttaksflygingen, regnes som *bestått*. Historisk er det om lag halvparten av kandidatene som starter på uttaksflyging som består. I datasettet for denne studien er tallet noe lavere.

Etter bestått uttaksflyging vil elevene fra høstkullet normalt gjennomføre ventetjeneste ved en av Luftforsvarets flyoperative avdelinger, samtidig med at vårkullet gjennomfører sin uttaksflyging. De selekterte kandidatene fra disse to gruppene, som normalt har gjennomført opptaksprøvene

²² Eksamensresultatene, med noe ulik vektning, utgjør til sammen kriteriet *37 Teorikarakter* i denne studien.

²³ Kriterium *35 CD-check* (Contact Dual checkride) i denne studien.

²⁴ Kriterium *36 Flykarakter* i denne studien.

²⁵ Kriterium *38 Hovedkarakter* i denne studien.

samtidig, vil deretter starte samtidig på Luftkrigsskolen (LKSK) påfølgende høst for å fortsette sitt utdanningsløp, og med det regnes som ferdig selektert.

4 Problemstilling

Studien tar sikte på å undersøke sammenhenger mellom to deler av det som til sammen kan kalles *seleksjonsprosessen* for flygerkandidater i Luftforsvaret. På den ene siden finnes inngangsverdiene til opptaket og resultatene fra opptaksprøvene, og på den andre siden finnes resultatene fra uttaksflygingen ved LFS. Hovedfokus på førstnevnte side vil være resultater fra de kognitive og psykomotoriske testene og intervju (kalt *Fase 2*), da det er disse parameterne som representerer den mest omfattende og ressurskrevende delen av opptaket. Resultatene fra kunnskapsprøvene i *Fase 1* og skolepoeng fra videregående skole, som kan sies å være brukt mer for en grovseleksjon, tas også med i analysene, men tillegges mindre vekt i denne studien. Gjennom kartlegging av nevnte sammenhenger undersøkes opptaksprøvenes prediktive validitet.

Problemstillingen er formulert som følger:

I hvilken grad er det sammenheng mellom resultatene fra opptaket til flygerutdanningen i Luftforsvaret, og resultatene fra Luftforsvarets flygeskole?

Problemstillingen vil bli forsøkt besvart gjennom en tilnærming som innebærer tre ulike delproblemstillinger. Delproblemstilling 1 går på sammenhengene mellom de enkelte testene i opptaket og resultatene ved LFS, med vekt på de testene som har til hensikt å måle kandidatenes kognitive og psykomotoriske evner. Mye av den tidligere forskningen på flygerseleksjon har basert seg på denne typen tester, og funnene som drøftes under denne delproblemstillingen vil kunne sees i sammenheng med funn i andre studier.

Delproblemstilling 1: I hvilken grad er det sammenheng mellom de kognitive og psykomotoriske testene (prediktorer) og resultatene fra LFS (kriterier)?

Videre vil delproblemstilling 2 adressere sammenhenger mellom intervju-delen av opptaket og resultater fra LFS, for å si noe om treffsikkerhet og hvilken verdi denne delen kan tenkes å ha for seleksjonen.

Delproblemstilling 2: I hvilken grad er det sammenheng mellom intervjuprognosene og resultatene fra LFS?

Som en tredje tilnærming for å belyse den overordnede problemstillingen, vil delproblemstilling 3 fra et mer helhetlig perspektiv ta for seg opptaksprøvenes samlede evne til å predikere LFS-prestasjoner.

Delproblemstilling 3: Hvor mye bidrar opptaksprøvene (prediktorer) samlet til å forklare variasjonen i LFS-prestasjoner (kriterier)?

4.1 Forventninger til funn

På bakgrunn av det tidligere studier på flygerseleksjon har vist (Martinussen, 1996), er det rimelig å forvente en positiv sammenheng mellom studiens prediktorer og kriterier. Utvalgets størrelse gir grunnlag for å kunne avdekke små og middels store korrelasjoner. Et strengt selektert utvalg, og det faktum at det ikke foretas korreksjoner for ulike feilkilder²⁶, gjør at de observerte korrelasjonene likevel forventes å være lave. Prediktorene som måler teorikunnskap og akademiske evner forventes å korrelere med kriteriet for akademiske prestasjoner, i tråd med tidligere studier (Martinussen, 1996). De databaserte testene i opptakets fase 2a som måler kognitive og psykomotoriske evner (psykomotorisk koordinasjon, spatiale evner og informasjonsprosessering) forventes å korrelere med kriteriene for praktisk flyging, i tråd med tidligere meta-analyser (Hunter & Burke, 1994; Martinussen, 1996) og valideringsstudier (Carretta et al., 2014; Lang-Ree et al., 2010; Martinussen & Torjussen, 2004).

Når det gjelder intervjuprognosene er det rimelig å forvente at disse tilfører en økt forklaringskraft i variasjonen i prestasjoner utover det som kan forklares gjennom opptaksprøvene, at studieprognosen har en sammenheng med kriteriet for akademiske prestasjoner, og at flygerprognosen har en sammenheng med kriteriene for praktisk flyging.

Det forventes at prediktorene samlet sett bidrar til å forklare en stor del av variasjonen i prestasjonskriteriene.

4.2 Avgrensninger

Studien omhandler det norske Luftforsvarets seleksjon til flygerutdanningen over en lengre periode frem til år 2020. Funn i denne studien har gyldighet innenfor denne rammen, men antas å kunne ha noen grad av overførbarhet til andre vestlige land og institusjoner med sammenlignbare systemer for flygerseleksjon.

Studien er en validering av opptaksprøvene opp mot kriterier på prestasjoner innhentet på LFS, og ikke opp mot prestasjoner videre i utdanningsløpet eller i jobben som flyger i Luftforsvaret. Det vil si at studien ikke tar sikte på å avdekke det gapet som kan antas å eksistere mellom det som måles

²⁶ Nærmere beskrevet i teorikapittelet

gjennom seleksjon, og det som faktisk kreves for å bli en god flyger på de ulike flytypene i Luftforsvaret. Et sånt gap vil kunne avdekkes gjennom validering av den norske seleksjonsprosessen opp mot etablerte prestasjonskriterier senere i utdanningsløpet, eller opp mot kriterier definert ut fra en jobbanalyse av en spesifikk flytype/rolle. Det finnes to eksempler på denne typen studier ved Forsvarets høgskole i nyere tid, som begge har benyttet prestasjoner fra jagerflygerutdanningen i USA som kriterier i sin undersøkelse (Harlem, 2016; Svensson, 2013). Valideringsstudier av seleksjonen opp mot etablerte prestasjonskriterier i andre utdanningsløp og jobbanalyser for andre av Luftforsvarets flytyper, omtales som muligheter for videre studier mot slutten av studien.

Det finnes variabler som har betydning for seleksjonen, men som ikke er en del av denne studien. Disse variablene vurderes som ikke relevant for problemstillingen, og er knyttet til fysisk helse og medisinsk skikkethet, formelle opptakskrav (som statsborgerskap, vandel og alder), prestasjoner under *Forsvarets opptak og seleksjon* (FOS, felles for alle kandidater til utdanning og tjeneste i Forsvaret) og vurdering av *Militært forhold* ved LFS. FOS-skåre kunne muligens vært relevant å inkludere, men har for det aktuelle utvalget i studien en for lav *N* til å være egnet for statistiske analyser.

To av testene som har blitt benyttet til seleksjon på deler av utvalget i undersøkelsen har i løpet av den aktuelle perioden blitt fjernet, og er derfor ikke del av denne studien. Dette gjelder «The Defence Mechanism Test»²⁷ (DMT) og «tappetesten»²⁸.

²⁷ Utviklet i Sverige av Ulf Kragh på 1960-tallet, med formål å velge ut personell til høyrisikoyrker, deriblant flygere. Testen ble brukt som del av flygerseleksjonen i Luftforsvaret frem til 2014, da den på bakgrunn av manglende dokumentasjon for prediktiv validitet ble besluttet tatt bort (ref Forsvarets Doculive-skriv 2014/003046-001).

²⁸ Tappetesten var tidligere en del av intervju-delen av opptaket. Testen ble tatt bort i 2019 da intervjuet gikk over fra fysisk til digital gjennomføring.

5 Metode

5.1 Deltakere

Over en periode på 15 år er det registrert 12650 søkere til Luftforsvarets flygerutdanning. Man kan anta at motivasjonen for å være søker i de fleste tilfeller er knyttet til et ønske om å bli flyger i det norske Luftforsvaret. Av den totale søkermengden har 5216 (41%) gjennomført hele eller deler av settet med opptaksprøver i regi av LSES. Av disse har 505 (like under 10%) startet som flyaspirant ved LFS, normalt i samme eller påfølgende år som opptaksåret. Det er disse 505 flyaspirantene som er deltakere i denne undersøkelsen, avgrenset til kandidater som startet ved LFS innen utgangen av 2021. Kandidater som for eksempel gjennomførte opptaksprøvene i 2020, men av ulike grunner ikke startet som flyaspirant før i 2022, er altså ikke deltakere.

Av de 505 flyaspirantene har 212 (42%) fullført LFS og dermed bestått seleksjonsprosessen. Av de opprinnelige 12650 søkerne utgjør dette om lag 1,68 %.

Samtlige befant seg i aldersspennet fra 18 til 24 år i opptaksåret. Kjønnfordelingen i utvalget er ukjent, men kan antas å ha en sterk overvekt av mannlige kandidater. Alder og kjønn er utelatt som variabler i denne studien. Studien skiller ikke mellom høst- og vårkull ved LFS.

Alle deltakere har på aktuelt tidspunkt tilfredsstilt formelle søknadskrav som for eksempel norsk statsborgerskap og plettfri vandel, samt en rekke medisinske krav (Forsvaret, 2022), som ikke er variabler i denne studien. Disse kravene har vært uendret i perioden.

Ved frabeordning fra seleksjonsflyging ved LFS vil kandidaten i noen tilfeller få tilbud om å fortsette som kandidat til navigatørutdanning. I denne studien som omhandler seleksjon til *flygerutdanningen* regnes disse kandidatene som *ikke bestått*.

Antallet deltakere (N) er for prediktorene om lag 500, med unntak av *Skolepoeng* ($N = 357$) som ligger en del lavere grunnet at disse skårene ikke har blitt registrert systematisk i hele perioden. For kriteriene er N generelt en del lavere (213-226), og gjenspeiler antallet som har kommet langt i seleksjonen og på den bakgrunn oppnådd en skåre på kriteriene. De resterende kandidatene (som mangler kriteriedata), har ikke fått disse innrapportert på bakgrunn av at de har blitt selektert bort i løpet av uttaksflygingen. For kriteriene er unntaket *Flyger (bestått/ikke bestått)*, som eksisterer for hele utvalget og derfor har $N = 505$.

5.2 Prediktorer

Som prediktorer i denne studien benyttes resultater oppnådd på de ulike testene og kategorier av tester i opptaket, i tillegg til *Skolepoeng* og *Alminnelig evnenivå* (AE) som begge er fastsatt en tid før kandidaten gjennomfører opptaksprøvene og derfor kan anses mer som inngangsverdier enn som del av selve opptaksprøvene. Disse prediktorene utgjør til sammen det som i studiens problemstilling omtales som *resultatene fra opptaket til flygerutdanningen i Luftforsvaret*.

Hver prediktor er gitt et unikt tosifret tall som identifikasjon i denne studien for enklere å kunne skille dem fra hverandre.

Skolepoeng er en verdi som fastsettes av instanser utenfor Forsvaret, og benytter en skala som er ulik den som benyttes for de andre prediktorene. Skalaen for skolepoeng er 0-60, hvor 60 representerer høyeste mulige poengsum. Øvrige prediktorer benytter en standard nidelt skala (STANINE), hvor 9 representerer høyeste mulige poengsum (Clark-Carter, 2005).

Beskrivelser av prediktorene er basert på Hansen (1987; Hansen et al., 1989), Martinussen og Torjussen (2004) og Forsvarets Sanitet, Seksjon for militærpsykologi (2022a).

5.2.1 Inngangsverdier for opptaksprøver

01 Skolepoeng: Poengsum oppnådd i videregående skole, i hovedsak basert på karakterer.

02 Alminnelig evnenivå (AE), fra sesjon²⁹.

5.2.2 Opptaksprøver Fase 1, papirbasert

Generell Teorikunnskap (GT)

03 Ravens Advanced Matrices: Testen består av 36 oppgaver og 2 eksempler. Hver oppgave består av en 3x3 matrise med symboler, hvor symbolet nederst til høyre mangler. Kandidaten må velge det riktige manglende symbolet fra 8 alternativer. Testen er ment å måle en generell intelligens.

04 Tallserier: Prøven har 22 oppgaver, og det er antall riktige svar som skåres. Oppgavene består av ufullstendige tallserier, hvor kandidaten skal finne de to neste tallene i rekken. Tallserien er satt opp etter en regel og det gjelder da å finne fram til denne regelen. Testen lader på en numerisk og en generell resonneringsfaktor.

²⁹ Sesjon er en generell, initiell screening av kandidater til tjeneste i Forsvaret

05 Ordforståelse: Testen er ment å måle generell begrepsforståelse, leseinteresse og evne til verbal kommunikasjon.

06 Regning (matematikk): Testen er ment å måle generelle matematikk-kunnskaper på videregående skole-nivå.

07 Engelsk: Tredelt test som er ment å måle generell evne til kommunikasjon på engelsk.

08 GT: Gjennomsnitt av de fem enkelttestene i GT-kategorien (03-07).

Teknisk innsikt og Romopplevelse (TR)

09 Mekaniske oppgaver: Fra figurer skal kandidaten finne frem til hvilket mekanisk prinsipp som gjelder. Testen er ment å måle en generell teknisk innsikt.

10 Figurformer: Kandidaten skal gjenkjenne figurer i ulike posisjoner for å matche en utgangsfigur. Det kreves derfor at kandidatene gjør en mental rotasjon for å finne frem til riktig svar.

11 Plateformer: Dette er en tredimensjonal spatialprøve. Oppgavene inneholder en utgangsfigur og fire tegninger av gjenstander. Utgangsfiguren er en plate som skal enten brettes etter stiplede linjer eller rulles til den danner en av de fire gjenstandene.

12 TR: Gjennomsnitt av de tre enkelttestene i TR-kategorien (09-11).

Flykunnskap

13 Flykunnskap: Generelle spørsmål knyttet til flyging og luftfart. Testen er ment å måle kunnskapsnivå på dette feltet, som kan være relatert til motivasjon.

14 MPAP: Gjennomsnitt av de 9 papirbaserte enkelttestene i Fase 1 (03-07, 09-11, 13)

5.2.3 Opptaksprøver Fase 2a, databasert

Samtlige av de 11 testene i Fase 2a gjennomføres ved bruk av datamaskiner og er programmert av firmaet Psytech LTD London. Av de 11 er to hentet fra PILAPT-testbatteriet utviklet av Burke og Valsler (1999) (17 Trax og 21 Hands), fem av testene er utviklet av Royal Air Force (15 CVT, 16 SMA, 20 Instrument Comprehension, 24 Digit Recall og 26 Vigilance), tre av testene er hentet fra det tidligere norske testbatteriet og gjort om til dataversjon (Torjussen & Hansen, 1999) (19 Planes, 23 Attention og 25 Numbers), og én er hentet fra *The Vienna Test System* utviklet av Schufried (www.schufried.co.at) (27 DTG).

Psykomotorisk koordinasjon (PK)

15 Control of velocity (CVT) 600: Kandidaten kontrollerer en hvit prikk på et horisontalt plan, og avskjærer røde prikker som kommer inn i bildet vertikalt. Kandidatens respons gis med joystick. Testen skåres på bakgrunn av antall treff, og er ment å måle evne til øye-hånd-koordinasjon.

16 Sensory motor apparatus (SMA) 610: Kandidatens oppgave er å styre et punkt horisontalt og vertikalt mot et tråkors på skjermen. Gjennom testen flyttes markøren bort fra tråkorset og kandidatens oppgave er å kompensere for dette og holde markøren så nært tråkorset som mulig. Kandidatens respons gis med joystick og fotpedaler. Testen skåres på bakgrunn av registrert avvik mellom markør og tråkors, og er ment å måle evne til kompleks koordinasjon av hånd-øye-fot.

17 Trax 630: En tredimensjonal målfølgingsoppgave hvor kandidatens oppgave er å styre en markør gjennom en definert bane som kan minne om en innflyging mot rullebanen før landing. Testen krever at kandidaten håndterer de spatiale parameterne som forekommer ved virkelig flyging. Kandidatens respons gis med joystick. Testen skåres på bakgrunn av antall registrerte feilaktige responser, og er ment å måle en kombinasjon av psykomotorisk koordinasjons-, spatiale- og informasjonsprosesseringssevner.

18 PK: Gjennomsnitt av de tre enkelttestene i PK-kategorien (15-17).

Spatiale evner (SE)

19 Planes 640: En test som involverer romoppfatning og mental rotasjon. Et bilde av et fly presenteres i et av hjørnene på skjermen. Flyet er orientert på en måte som indikerer at det enten beveger seg mot deg eller fra deg, og har i tillegg én av fire ulike konfigurasjoner: Helling («bank») mot høyre, helling mot venstre, opp-ned eller horisontale vinger («straight and level»). En målposisjon indikeres i motsatt hjørne av skjermen, og kandidatens oppgave er å bestemme hvilken stikkebevegelse en tenkt pilot i dette flyet vil måtte utføre for å bevege det i retning målposisjonen. Kandidatens respons gis med joystick og testen skåres på bakgrunn av antall riktige responser i kombinasjon med hurtighet.

20 Instrument comprehension 650: Testen innebærer oppfattelse og forståelse av informasjon som formidles visuelt på skjermen gjennom bilder av seks ulike flyrelaterte instrumenter: Hastighetsmåler, kunstig horisont, høydemåler, svinge-/krenge-indikator, kompass og stige-/synkeindikator. I den første av testens to deler presenteres kandidaten for en visuell fremstilling av alle de seks instrumentene og et sett med verbale beskrivelser. Kandidatens oppgave er å bestemme hvilken av de verbale beskrivelsene som stemmer

overens med instrumentene. I testens andre del presenteres kandidaten for en visuell fremstilling av to av instrumentene (kunstig horisont og kompass) og et sett med bilder som viser et fly orientert på ulike måter. Kandidatens oppgave er å bestemme hvilket av bildene som stemmer overens med instrumentene. Testen skåres på bakgrunn av antall riktige svar og tidsforbruk.

21 Hands 660: Kandidaten gis via hodetelefoner en auditiv kommando som inneholder tre elementer: Positiv/negativ, høyre/venstre og sirkel/firkant. Kandidatene prosesserer informasjonen for så å foreta et visuelt søk etter informasjon som presenteres på skjermen. Kandidaten trenger å lagre og omgjøre verbal informasjon til visuell informasjon, og bruke denne til å velge riktig svaralternativ. Informasjonen som presenteres på skjermen består av bilder av en person som kan være orientert i ulike retninger og holder sirkler eller firkanter i høyre eller venstre hånd. Kandidatens respons gis med tastatur og testen skåres på bakgrunn av antall riktige responser i kombinasjon med hurtighet.

22 SE: Gjennomsnitt av de tre enkelttestene i SE-kategorien (19-21).

Informasjonsprosesseringssevner (IP)

23 Attention 670: Testen er sammensatt av to tidligere anvendte tester kalt «Retningsreversalier» og «Sorteringsprøven». En 4x4-matrise bestående av tall og farger presenteres på skjermen sammen med en auditiv kommando som gis via hodetelefoner. Kommandoen inneholder en angivelse av hvilken del av matrisen kandidaten skal fokusere på, og hvorvidt det er tallet eller fargen som er viktig. Kommandoene gis i et økende tempo, samtidig som eksponeringstiden for den visuelle informasjonen reduseres. Kandidaten avgir sine svar i henhold til hvilken informasjon det spørres etter over hodetelefonene. Testen skåres på bakgrunn av responsenes nøyaktighet.

24 Digit recall 680: Kandidaten presenteres en visuell fremstilling av et sett med tall av varierende verdi. Tallene forsvinner etter en viss tid, og kandidatens oppgave er å taste inn tallene i riktig rekkefølge ved hjelp av tastatur. Testen skåres på bakgrunn av antall riktig inntastede tall og er ment å måle korttidsminne.

25 Numbers 700: Dataversjon av en tidligere anvendt test kalt «Tallsøker». En mengde tall av varierende verdi og fysisk størrelse presenteres visuelt på skjermen i et vilkårlig mønster i fire sekvenser. Kandidatens oppgave er å lokalisere så mange spesifikke tall som mulig i hver av de fire sekvensene. Tilgjengelig tid reduseres gjennom testen. Testen skåres på bakgrunn av summen av antall treff gjennom de fire sekvensene. Testen er ment å måle

persepsjonshastighet og assosieres med hurtig, effektiv og presis bruk av visuell persepsjon samt korttidsminne.

26 Vigilance 720: Kandidaten presenteres for en 9x9-matrise, hvor hver celle kan identifiseres gjennom tallbenevnelser langs matrisens øvre og venstre kanter. Kandidaten pålegges en rutinepreget oppgave og en prioritert oppgave, og må dele sin oppmerksomhet mellom disse.

27 Determinasjonsgerät (DTG) 690: Kandidaten presenteres en auditiv kommando via hodetelefoner. For hver kommando avgis en respons gjennom bruk av fotpedaler og et eget tastatur bestående av fargede knapper. Hastigheten økes gjennom testen. Testen skåres på bakgrunn av antall riktige responser, og er ment å måle selektiv oppmerksomhet og responsorientering.

28 IP: Gjennomsnitt av de fem enkelttestene i IP-kategorien (23-27).

29 MDATA: Gjennomsnitt av de 11 databaserte enkelttestene i fase 2a (15-17, 19-21, 23-27)

30 MTOTAL: Gjennomsnitt av de 20 enkelttestene i fase 1 og fase 2a (03-07, 09-11, 13, 15-17, 19-21, 23-27)

5.2.4 Opptaksprøver Fase 2b, intervju

31 Skoleprognose: Hvordan panelet vurderer at kandidaten vil håndtere det vedvarende kravet til å tilegne seg teoretisk kunnskap og forståelse under flygerutdanningen.

32 Offisersprognose: Hvordan panelet vurderer at kandidaten, gitt utdanning og trening, er egnet som offiser i Forsvaret.

33 Flygerprognose: Hvordan panelet vurderer at kandidaten har de tekniske ferdighetene og den kognitive kapasiteten som er nødvendig for å kunne bli en dyktig operativ flyger.

34 GJ.SNITT INTERVJU: Gjennomsnitt av de tre intervjuprognosene (31-33).

5.3 Kriterier

Som kriterier i denne undersøkelsen benyttes resultater fra uttaksflygingsperioden ved LFS.

Kriteriene kan til sammen hevdes å gi et bilde av prestasjoner ved LFS, og er et uttrykk for det som i studiens problemstilling omtales som *resultatene fra LFS*.

Samtlige skårer fastsettes av skolens instruktører på bakgrunn av metoder beskrevet i skolens egne publikasjoner. Beskrivelser av kriteriene er basert på *LFS Instruktørhåndbok*³⁰ og samtaler med sjef LFS og representanter fra instruktørkorpset (2022).

Skalaen for de fire første kriteriene (35-38) er 1-6, hvor 6 representerer høyeste mulige poengsum. Det gis anledning til å benytte hele og halve tall som karakter. Karakteren 2,0 representerer laveste bestått-karakter. Kandidatene vil ikke kunne få karakter 2,0 på to påfølgende flyturer i programmet, i så tilfelle vil karakteren for den siste av de to turene bli satt til 1,5, selv om prestasjonen isolert sett på den aktuelle turen vurderes å tilsvare karakter 2,0.

Kriteriet *flyger* (39) er dikotomt og indikerer hvorvidt kandidaten har bestått (verdi 1) eller ikke bestått (verdi 0) uttaksflygingen.

Hvert kriterium er, på samme måte som prediktorene, gitt et unikt tosifret tall som identifikasjon i denne studien for enklere å kunne skille dem fra hverandre.

- **35 CD-check:** Resultat av eksamensflyturen som finner sted mot slutten av *Contact*-fasen i flyprogrammet, normalt etter 19 flyturer. Inneholder momenter som måler i hvilken grad kandidaten behersker grunnleggende operasjon av flyet, både på bakken, i treningsområdet og i flyplassens landingsrunde.
- **36 Flykarakter:** Et uttrykk for kandidatens samlede prestasjoner innen praktisk flyging. Karakteren er sammensatt av resultatene av alle tellende flyturer i programmet, med mest vekt på resultatet av eksamensturen i *Contact*-fasen (35 CD-check i denne studien). Frem til 2016 utgjorde CD-check 50%, og gjennomsnittet av de resterende turene 50% av flykarakteren. Gjeldende fra 2016 ble denne vektingen endret til henholdsvis 33,33% og 66,67%.
- **37 Teorikarakter:** Et uttrykk for kandidatens samlede prestasjoner innen akademiske fag. Karakteren er sammensatt av resultatene fra 9 teoriexaminer, hvor disse vektet ulikt. Fagene *engelsk*, *air traffic services* og *aerodynamikk* utgjør 16,67% hver, *flymedisin*, *teknisk kurs* og *meteorologi* 11,11% hver og *airplane general*, *navigasjon* og *instrumentteori* 5,56% hver.
- **38 Hovedkarakter:** Gjennomsnitt av flykarakter og teorikarakter, der begge vektet likt.
- **39 Flyger:** Et dikotomt mål for hvorvidt kandidaten har bestått uttaksflygingen (bestått/ikke bestått).

³⁰ *LFS Instruktørhåndbok* versjon 2014 og versjon 2020v2, i nyere utgave kalt *LFS Seleksjonshåndbok 129-51A*.

5.4 Prosedyre

Data fra søknad og opptaksprøver oppbevares i Forsvarets HR-system (P3), og forvaltes ved LSES med tilgangskontroll gitt fra FPVS. Denne delen av datasettet utgjør prediktorene i undersøkelsen. Til LSES sendes også kullrapporter fra sjef LFS, som er grunnlaget for den delen av datasettet som utgjør kriteriene i undersøkelsen. Sammenstilling av disse delene, anonymisering og utlevering av datasettet for gjennomføring av denne undersøkelsen, er foretatt av Sjefpsykologen i Forsvaret. Variabler som kan bidra til å muliggjøre identifisering av enkeltpersoner er fjernet, og det er ikke utlevert noen form for koblingsnøkkel. Undersøkelsen er derfor å anse som anonym, og etter dialog med NSD er konklusjonen at en formell godkjenning fra NSD for bruk av dette datasettet i prosjektet ikke er nødvendig.

Godkjenning for bruk av datasettet i denne studien er gitt av Forsvarets forskningsnemnd ved Forsvarets høgskole 18.10.2022.

5.5 Statistiske analyser

De statistiske analysene i denne studien er gjennomført ved bruk av dataprogrammet *Statistical Package for the Social Sciences* (SPSS, versjon 27.0.1.0.). Signifikansnivået (α) er satt til .05, som betyr at p -verdier mindre enn .05 vurderes som statistisk signifikante. Korrelasjoner beregnes som Pearsons r . Resultatene korrigeres ikke for statistiske feilkilder, eksempelvis problematikk knyttet til *restriction of range* eller *kriterierelabilitet*.

5.5.1 Bivariate korrelasjoner

Analyser av lineære bivariate korrelasjoner mellom enkeltvis variabler gir grunnlag for å besvare delproblemstilling 1 og 2.

5.5.2 Hierarkisk multippel regresjon

Analyser av hierarkisk multippel regresjon gir grunnlag for å besvare delproblemstilling 2 og 3, gjennom å avdekke i hvor stor grad variasjonen i tre sentrale prestasjonskriterier (36 Flykarakter, 37 Teorikarakter og 39 Flyger) kan forklares gjennom kombinasjoner av prediktorer, analysert stegvis og i den rekkefølgen testene er gjennomført i opptaket.

Det gjennomføres to sett firestegs regresjonsanalyser. I det første settet benyttes én prediktor per steg, hvor alminnelig evnenivå benyttes i steg 1, og de påfølgende stegene benytter gjennomsnittsskårer for henholdsvis papirtester, databaserte tester og intervju. Det andre settet regresjonsanalyser har en tilsvarende oppbygning, men benytter flere prediktorer per steg, hvor steg 1 også innbefatter skolepoeng fra videregående skole, og de påfølgende stegene benytter gjennomsnittsskårer for testkategoriene i opptakets fase 1 (steg 2) og fase 2a (steg 3), før de tre intervjuprognosene benyttes i steg 4. Tilfeller av manglende verdier i datasettet er ekskludert parvis.

Resultatet av regresjonsanalysene er presentert i henholdsvis Tabell 3 og Tabell 4 der i begge tilfeller tre ulike kriterier er forsøkt predikert. Forskjellen mellom resultatene er at analysene i Tabell 3 anvender samleskårer på hvert trinn og dermed er alle prediktorene som inngår i et gitt trinn likt vektet, mens for Tabell 4 inngår flere prediktorer på hvert trinn som gjennom regresjonsanalysen vektet ulikt inn i modellen.

Hvor stor del av variansen i kriteriet (avhengig variabel) som kan forklares av variansen i prediktoren/prediktorene (uavhengig variabel) i det enkelte steg uttrykkes som ΔR^2 . I regresjonsanalysene beregnes også en standardisert betakoeffisient (β) som uttrykker hvor mye en økning på én enhet i den aktuelle prediktoren (uavhengig variabel) tilsvarer i forventet økning på kriteriet (avhengig variabel) gitt at alle øvrige prediktorer holdes konstant. Individuelle β -verdier blir kun fortolket når trinnet som helhet er signifikant. En F -verdi forteller hvorvidt regresjonsmodellen som helhet er statistisk signifikant for de ulike kriteriene.

6 Resultater

6.1 Bivariate korrelasjoner

Tabell 1 viser gjennomsnitt, standardavvik, N og bivariate korrelasjoner for alle variabler i studien.

Tabell 2 er en forenklet utgave av Tabell 1 med høyere lesbarhet, som viser bivariate korrelasjoner for utvalgte variabler. Tabellen viser de fem prestasjonskriteriene, og ni prediktorer som er valgt ut på bakgrunn av at de representerer gjennomsnitt for grupper av tester som er ment å måle tilnærmet samme evne (08, 12, 18, 22 og 28) eller representerer mer overordnede gjennomsnitt av testresultater (14, 29, 30, 34).

Tabell 2. *Bivariate korrelasjoner for utvalgte variabler (N = 213 – 505)*

Variabler	08	12	14	18	22	28	29	30	34	35	36	37	38	39
Prediktorer														
08 GT														
12 TR	.43**													
14 MPAP	.87**	.77**												
18 PK	.06	.11*	.14**											
22 SE	.25**	.39**	.38**	.27**										
28 IP	.23**	.04	.19**	.18**	.32**									
29 MDATA	.26**	.23**	.33**	.65**	.72**	.76**								
30 MTOTAL	.66**	.59**	.78**	.51**	.68**	.61**	.85**							
34 Gj.sn. Int.	.50**	.35**	.51**	.11*	.33**	.33**	.36**	.53**						
Kriterier														
35 CD-check	.07	.12	.13	.06	.13	.08	.13	.16*	.13					
36 Flykarakter	.11	.23**	.23**	.14*	.20**	.03	.16*	.23**	.12	.68**				
37 Teorikarakter	.37**	.27**	.42**	.22**	.24**	.12	.26**	.41**	.24**	.04	.28**			
38 Hovedkarakter	.23**	.30**	.35**	.20**	.24**	.07	.22**	.35**	.17*	.60**	.96**	.53**		
39 Flyger ^a	.00	.08	.07	.13**	.08	.00	.09*	.10*	.10*	.15*	.34**	.06	.22**	

Note: ^aPass/fail

* $p < .05$; ** $p < .01$ (to-halet)

6.2 Hierarkisk multippel regresjon

Tabell 3. Hierarkisk multippel regresjonsanalyse for prediksjon av kriterier for flygerprestasjon (med én prediktorer pr steg)

Variabler	36 Flykarakter ^a		37 Teorikarakter ^b		39 Flyger ^c	
	β	ΔR^2	β	ΔR^2	β	ΔR^2
Steg 1. Inngangsverdier						
01 Alminnelig evnenivå	-.05	.01	.11	.09**	.01	.00
Steg 2. Opptaksprøver Fase 1						
14 MPAP	.23**	.05**	.32**	.10**	.00	.00
Steg 3. Opptaksprøver Fase 2a						
29 MDATA	.09	.01	.14*	.02*	.06	.01
Steg 4. Opptaksprøver Fase 2b						
34 Gj.snitt intervju	-.02	.00	-.01	.00	.08	.00
Total R^2		.06**		.20**		.02
F		3.40*		12.91**		1.81
N		212		213		489

Note. ^aResultatet av tellende flyturer, med mest vekt på CD-checkride ^bKarakter sammensatt av 9 teoriexamener ^cPass/fail. Koeffisienter er hentet fra siste steg i regresjonsanalysen. Verdier i tabellen er avrundet. Summen av avrundede ΔR^2 -verdier kan derfor være ulik summen av eksakte R^2 -verdier.

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

Regresjonsmodellen som helhet er statistisk signifikant for kriteriene 36 Flykarakter og 37 Teorikarakter. Regresjonen er ikke statistisk signifikant for kriteriet 39 Flyger.

Tabell 4. Hierarkisk multippel regresjonsanalyse for prediksjon av kriterier for flygerprestasjon (med flere prediktorer pr steg)

Variabler	36 Flykarakter ^a		37 Teorikarakter ^b		39 Flyger ^c	
	β	ΔR^2	β	ΔR^2	β	ΔR^2
Steg 1 Inngangsverdier		.03		.11**		.01
01 Skolepoeng	.17		.07		.09	
02 Alminnelig evnenivå	.00		.13		.07	
Steg 2 Opptaksprøver Fase 1		.05*		.06**		.01
08 GT	-.04		.16		-.16*	
12 TR	.18*		.09		.05	
Steg 3 Opptaksprøver Fase 2a		.03		.05*		.02*
18 PK	.11		.24**		.14**	
22 SE	.11		.14		.04	
28 IP	-.04		.02		-.06	
Steg 4. Opptaksprøver Fase 2b		.00		.02		.01
31 Skoleprognose	.00		.13		.08	
32 Offisersprognose	-.03		.03		.03	
33 Flygerprognose	.00		-.19*		.04	
Total R^2		.10		.24		.05
F		1.63		4.49**		1.77
N		156		151		351

Note. ^aResultatet av tellende flyturer, med mest vekt på CD-checkride ^bKarakter sammensatt av 9 teorieksamener ^cPass/fail. Koeffisienter er hentet fra siste steg i regresjonsanalysen. Verdier i tabellen er avrundet. Summen av avrundede ΔR^2 -verdier kan derfor være ulik summen av eksakte ΔR^2 -verdier.

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

Regresjonsmodellen som helhet er statistisk signifikant for kriteriet 37 Teorikarakter. Regresjonen er ikke statistisk signifikant for de to andre kriteriene i tabellen.

7 Diskusjon

Dette kapittelet består av tre delkapitler som hver vil svare ut hver sin delproblemstilling.

7.1 Diskusjon delproblemstilling 1

Delproblemstilling 1: I hvilken grad er det sammenheng mellom de kognitive og psykomotoriske testene (prediktorer) og resultatene fra LFS (kriterier)?

Dette delkapittelet vil først ta for seg hver av de fem kriteriene (35-39) og belyse korrelasjonene (som er presentert i Tabell 1 og Tabell 2) for de ulike prediktorene som måler kognitive og psykomotoriske evner (prediktorene 01-30). Deretter vil det bli sett nærmere på noen av prediktorene som skiller seg ut, før funnene vil bli sett i sammenheng med tidligere studier.

7.1.1 Korrelasjoner for hvert av de fem kriteriene

Kriterium 35 CD-check er én enkelt flytur, og den viser jevnt over lave korrelasjoner med prediktorene. Kriteriet har signifikant korrelasjon med kun to av enkelttestene (11 Platteformer $r = .016$, 17 TRAX $r = .014$, ref Tabell 1), i tillegg til en signifikant korrelasjon med snittet av enkelttestene (30 MTOTAL $r = .16$).

Kriterium 36 Flykarakter har signifikant korrelasjon med alle nøkkelprediktorene i Tabell 2, foruten 08 GT og 28 IP. Dette kriteriet later jevnt over til å la seg predikere rimelig godt av testene, foruten de to nevnte prediktorene som ser ut til å bidra lite. Av enkelttester finnes det signifikante korrelasjoner for 01 Skolepoeng ($r = .16$), 06 Regning ($r = .16$), 09 Mekanisk ($r = .17$), 10 Figurformer ($r = .20$), 13 Flykunnskap ($r = .23$), for to av de tre testene i PK-kategorien (16 SMA $r = .16$, 17 TRAX $r = .19$) og for to av de tre testene i SE-kategorien (19 Planes $r = .17$, 20 Instrument $r = .18$). Disse funnene kan sies å være i tråd med forventningene om at de databaserte testene som måler kognitive og psykomotoriske evner har sammenhenger med (hoved-) kriteriet for praktisk flyging, med unntak av fraværet av positive, signifikante korrelasjoner mellom dette kriteriet og informasjonsprosesseringssevner (28 IP).

Kriterium 37 Teorikarakter har jevnt over høyere korrelasjoner med prediktorene enn det som er tilfellet for noen av de andre kriteriene. Kriteriet har signifikante korrelasjoner med samtlige utvalgte nøkkelprediktorer i Tabell 2, foruten 28 IP. Tabell 1 gjenspeiler også dette gjennom jevnt over relativt sterke korrelasjoner med testene foruten de som er i IP-kategorien, hvor kun én av de fem testene

har en signifikant korrelasjon (23 Attention $r = .23$) med kriteriet. Som forventet hadde dette kriteriet en rekke små og middels store, positive signifikante korrelasjoner med prediktorene som måler teorikunnskap og akademiske evner (prediktorene 01-14, $r = .10 - .42$) i tråd med intensjonen med disse testene. Generelt har mer generelle intelligens tester vært noe dårligere prediktorer av flygerprestasjoner enn mer spesifikke kognitive eller psykomotoriske tester (Hunter & Burke, 1994; Martinussen, 1996). Her kan nok ulikheter i type kriterium forklare forskjellen mellom denne studien og resultatene fra meta-analysene som i stor grad baserte seg på enten pass/fail i trening eller instruktørvurderinger av selve flygningen.

Kriterium 38 Hovedkarakter er satt sammen av 36 og 37, og har på den bakgrunn også korrelasjoner med prediktorene som gjenspeiler en blanding av det man finner for henholdsvis 36 og 37. Som for 37, har også 38 signifikante korrelasjoner med samtlige utvalgte nøkkelprediktorer i Tabell 2, foruten 28 IP.

Kriterium 39 Flyger er et dikotomt *bestått/ikke bestått*-kriterium. For dette kriteriet finnes det jevnt over lavere korrelasjoner med prediktorene enn det som finnes for de andre (kontinuerlige) kriteriene. Noe av forklaringen her kan ha sammenheng med effekten av dikotomisering, men det er også mulig å se for seg at dette kriteriet har en lavere reliabilitet eller at det måler andre faktorer utover kun å representere et rent prestasjonsmål.

Av nøkkelprediktorene i Tabell 2 finnes signifikante korrelasjoner med 18 PK ($r = .13$), og med de to mer overordnede gjennomsnittsskårene 29 MDATA ($r = .09$) og 30 MTOTAL ($r = .10$). For både 08 GT og 28 IP er observert korrelasjon mot dette kriteriet lik null. Blant enkelttestene finnes det for dette kriteriet signifikante korrelasjoner til 06 Regning ($r = .10$), 09 Mekanisk ($r = .12$), 13 Flykunnskap ($r = .17$) og to av de tre testene i PK-kategorien (16 SMA $r = .12$, 17 TRAX $r = .17$), 20 Instrument ($r = .16$).

7.1.2 Prediktorer som skiller seg ut

Tabell 2 viser korrelasjonene mellom de fem kriteriene og noen utvalgte nøkkelprediktorer. Tabellen viser at disse nøkkelprediktorene jevnt over er godt egnet til å predikere variasjoner i prestasjonskriteriene, med noen unntak som skiller seg ut og kan være verdt å legge merke til.

08 GT har signifikant korrelasjon med 37 Teorikarakter og 38 Hovedkarakter, men ikke for de tre andre kriteriene. Denne gruppen av tester som måler generell teorikunnskap, ser altså ut til å være egnet til å predikere prestasjoner innen teoretiske fag, men ikke prestasjoner som i større grad omhandler praktisk flyging.

Blant prediktorene som gjenspeiler gjennomsnitt av sine respektive grupper av tester som er ment å måle samme eller tilnærmet samme evne (08, 12, 18, 22, 28), er psykomotorisk koordinasjon (18 PK) den eneste som har en signifikant korrelasjon til kriteriet 39 Flyger. Blant testene som har vært brukt til flygerseleksjon i Luftforsvaret i den aktuelle perioden, kan derfor psykomotorisk koordinasjon hevdes å være den kategorien av kognitive tester som i størst grad har evne til å predikere hvorvidt en kandidat kommer til å bestå LFS.

Informasjonsprosesseringssevner (28 IP) skiller seg ut gjennom å være den eneste blant de utvalgte nøkkelprediktorene som ikke har signifikante korrelasjoner med noen av de fem prestasjonskriteriene. Jevnt over kan det se ut til at denne kategorien av tester i svært liten grad har evne til å predikere noe som helst av det som måles ved LFS. Ser man i Tabell 1 på de fem enkelttestene som til sammen utgjør kategori-gjennomsnittet 28 IP (prediktorene 23-27), er det, med unntak av en signifikant korrelasjon mellom 23 Attention og 37 Teorikarakter ($r = .23$), ingen signifikante korrelasjoner mellom disse fem prediktorene og noen av kriteriene.

7.1.3 Funn sett i sammenheng

Kriteriene 35 CD-check og 36 Flykarakter måler noe av det samme, da begge gjenspeiler prestasjoner knyttet til praktisk flyging³¹. Førstnevnte er resultatet av én enkelt flytur, mens sistnevnte er resultatet av en mengde flyturer over en lengre periode. Disse to ulike metodene for å måle det som kan antas å være samme verdi (noe i retning av *praktiske flygerferdigheter*), resulterer i ulik grad av reliabilitet ved at kriteriet som består av flere turer har flere målepunkter og dermed bedre reliabilitet. Denne forskjellen medvirker trolig til at de påviste korrelasjonene med prediktorene generelt er høyere for 36 Flykarakter enn for 35 CD-check.

Kriterium 37 Teorikarakter har generelt noe høyere korrelasjoner med prediktorene enn det som er tilfellet for kriteriene 35 CD-check og 36 Flykarakter. Førstnevnte gjenspeiler teori, mens de to sistnevnte gjenspeiler praktisk flyging. Det kan derfor se ut til at de aktuelle seleksjonstestene fungerer godt som prediktorer for prestasjoner innen teoretiske fag, og i noe mindre grad fungerer som prediktorer for prestasjoner innen praktisk flyging.

At testene som måler informasjonsprosesseringssevner har så liten evne til å predikere prestasjoner ved LFS, er oppsiktsvekkende. Dette funnet er det som i størst grad bryter med forventningene til studien og tidligere funn. Dette funnet i seg selv gir dog ikke grunnlag for å hevde at IP-kategorien av tester ikke har noe for seg. Det kan godt tenkes at evner knyttet til informasjonsprosesseringssevner er viktig

³¹ Se avsnittet *Kriterier* i Metode-kapittelet

i jobben som flyger i Luftforsvaret, slik Harlem fant i sin studie på F-35 (Harlem, 2016), og at spørsmålet man heller bør stille er om Luftforsvaret hadde vært tjent med en uttaksflyging som i større grad inkluderer informasjonsprosesseringsevner som del av det som måles og vurderes (kriteriene). At psykomotorisk koordinasjon er den blant testgruppegjennomsnittsskårene som i størst grad ser ut til å predikere hvilke kandidater som består LFS ($r = .13$), mens informasjonsprosesseringsevner *ikke* ser ut til å være egnet til å predikere dette ($r = .00$), er interessant. Harlem peker nettopp på funn som indikerer at med nyere flymaskiner blir psykomotorisk kapasitet mindre viktig som en konsekvens av at jagerflyene blir mer avanserte og lettere å fly, mens informasjonsprosesseringsevne blir mer viktig. Hvilke evner som kreves for å være en god operativ flyger på Luftforsvarets øvrige flytyper eksisterer det lite forskning på, og det er mulig å se for seg at disse evnene kan avvike noe fra F-35, men samtidig virker det sannsynlig at overgang fra eldre til nyere teknologi vil medføre endrede evnekrav i retning økt viktighet av informasjonsprosesseringsevne.

7.2 Diskusjon delproblemstilling 2

Delproblemstilling 2: I hvilken grad er det sammenheng mellom intervjuprognosene og resultatene fra LFS?

Dette delkapittelet vil først belyse funn fra korrelasjons- og regresjonsanalysene som er relatert til de ulike intervjuprognosene (prediktorene 31-34), før disse sees i sammenheng.

7.2.1 Korrelasjoner for hvert av de fem kriteriene

For kriterium 35 CD-check finnes det kun små, positive korrelasjoner til de ulike intervjuprognosene, hvorav ingen er statistisk signifikante. Gjennomsnittet av de tre prognosene har en korrelasjon på $r = .13$ med CD-check.

Kriterium 36 Flykarakter har en signifikant korrelasjon ($r = .16$) med skoleprognosen, men ikke med noen av de andre intervjuprognosene. Gjennomsnittet av de tre prognosene har en korrelasjon på $r = .12$ med dette kriteriet, omtrent som for CD-check.

Kriterium 37 Teorikarakter er der man finner de største korrelasjonene relatert til intervjuprognoser. Kriteriet har en middels stor, signifikant korrelasjon ($r = .34$) til skoleprognosen, og en liten, signifikant korrelasjon ($r = .14$) til flygerprognosen. For prognosegjennomsnittet finnes en signifikant korrelasjon ($r = .24$) til kriteriet.

Kriterium 38 Hovedkarakter er satt sammen av 36 og 37, og har på den bakgrunn også korrelasjoner med intervjuprognosene som gjenspeiler en blanding av det man finner for henholdsvis 36 og 37. Som for 37, har også 38 signifikante korrelasjoner med skoleprognosen ($r = .23$), flygerprognosen ($r = .18$) og med gjennomsnittet av de tre prognosene ($r = .17$), men ikke med offisersprognosen ($r = -.03$).

For *bestått/ikke bestått*-kriteriet 39 Flyger, så er korrelasjonene noe lavere, men ser ut til å følge samme mønster som for de øvrige kriteriene. Det finnes også her små, signifikante korrelasjoner til skoleprognosen ($r = .10$), flygerprognosen ($.09$) og gjennomsnittsskåren ($r = .10$), men ikke til offisersprognosen ($r = .04$).

7.2.2 Regresjon

Steg 4 i regresjonsanalysene som er presentert i Tabell 3 og Tabell 4 viser at ingen av de tre intervjuprognosene eller gjennomsnittet av disse ser ut til å tilføre ytterligere forklaringskraft i variasjonen i de tre valgte kriteriene enn det som kan forklares gjennom opptaksprøvene.

7.2.3 Funn sett i sammenheng

I hvor stor grad er så de ulike prognosene egnet til å predikere prestasjoner ved LFS? Det kan se ut til at offisersprognosen jevnt over er lite egnet til å predikere noen av kriteriene i undersøkelsen. Det kan tenkes at prognosen er et uttrykk for evner eller et potensiale som ikke måles gjennom de kriteriene som er valgt ut i denne undersøkelsen, eksempelvis kan man se for seg en mulighet for at det eksisterer en sammenheng mellom offisersprognosen og henholdsvis prestasjoner under FOS og vurdering av militært forhold ved LFS.

Praktiske flygerprestasjoner ser ut til vanskelig å la seg predikere gjennom intervjuprognoser, med jevnt over lave korrelasjoner. For de to kriteriene som i sin reneste form er knyttet til praktiske flygerprestasjoner – 35 CD-check og 36 Flykarakter – finnes det altså kun én signifikant korrelasjon mot intervjuprognosene, og den er interessant nok knyttet til *skoleprognose*, og ikke til *flygerprognose*. Blant kriteriene ser akademiske prestasjoner – målt gjennom kriteriet 37 Teorikarakter – ut til å være det som i størst grad lar seg predikere gjennom intervju.

Det kan heller ikke utelukkes at skole- og flygerprognosene er egnet til å predikere det dikotome *besått/ikke bestått*-kriteret (39). Korrelasjonene er lave men signifikante. Hvor mye av grunnen for de lave korrelasjonene som skyldes feilkilder knyttet til dikotomt kriterium er vanskelig å si uten å gjøre denne korreksjonen matematisk, men det vil nok være rimelig å se for seg at de korrigerede korrelasjonstallene ville vært en del høyere. Funnene kan i noen grad hevdes å være i tråd med

forventningene til undersøkelsen. Studieprognosen later til å ha en klar sammenheng med akademiske prestasjoner, mens flygerprognosen bare i noen grad ser ut til å ha sammenheng med kriteriene for praktiske flygerprestasjoner.

Både korrelasjons- og regresjonsanalysene bidrar til å skape et inntrykk av at intervjudelen av opptaket tilfører svært liten forklaringskraft for variasjonen i prestasjoner ved LFS, utover det som kan forklares gjennom de foregående testene i seleksjonsprosessen. På den måten kan det etterlatte inntrykket bli at intervjuet ikke har noe for seg. Det som dog er viktig å ta høyde for, er at analysene i denne studien er avgrenset til kandidater som har kommet gjennom intervjuet og fått anledning til å starte uttaksflygingen ved LFS. Analysene ser ikke på kandidater som har blitt selektert bort i intervjuet, og er derfor ikke egnet til å si noe om hvorvidt intervjuet fungerer for å luke ut disse (uegnede) kandidatene. Det kan tenkes at intervjudelens hovedfunksjon ikke nødvendigvis er å predikere hvor sterkt kandidatene kan komme til å prestere ved LFS, men heller er å selektere bort kandidater som det er grunn til å stille spørsmål ved hvorvidt er egnet for en karriere som flyger i Luftforsvaret. Intervjuet kan på den måten tenkes å ha mer en «select-out»- enn en «select-in»-funksjon, slik Raymond E. King beskriver det i en artikkel om personlighet- og psykopatologi-vurderinger i flygerseleksjon (King, 2014).

I en studie fra 2016 som på et overordnet nivå forsøker å oppsummere forskning på bruk av psykologiske metoder innenfor seleksjon av personell, peker forfatterne på at kombinasjonen av evnetester og intervju viser god prediktiv validitet for fremtidige jobbprestasjoner (Schmidt et al., 2016). Intervjuet kan synes å bidra til å øke validiteten av seleksjonsprosessen som helhet, heller enn at det nødvendigvis tilfører mye isolert sett. Intervjuet kan på en slik bakgrunn anses som del av en helhet heller enn et isolert ledd i seleksjonen.

Det er også mulig å se for seg at intervjuet kan ha andre funksjoner enn kun som redskap for seleksjon. Intervjuet kan gi kandidatene anledning til å stille spørsmål og få svar som bidrar til en bedre forståelse av egne prestasjoner, seleksjonsprosessen, utdanningsløpet og videre karriere, og kan på den måten påvirke kandidatens selvinnsikt, motivasjon og opplevelse av en rettferdig seleksjon.

7.3 Diskusjon delproblemstilling 3

Delproblemstilling 3: Hvor mye bidrar opptaksprøvene (prediktorer) samlet til å forklare variasjonen i LFS-prestasjoner (kriterier)?

Dette delkapittelet vil ta for seg funnene i regresjonsanalysene som er presentert i Tabell 3 og Tabell 4, og se disse i sammenheng. De to analysene foregår begge i fire steg som inneholder noe av det samme. I Tabell 3 ser vi at steg 1 er gjort med AE som prediktor. Analysen gir en positiv β -verdi og signifikant ΔR^2 for teori-kriteriet, men lite for de to andre kriteriene. Bildet ser ut til å være det samme for steg 1 i Tabell 4, hvor bruk av skolepoeng i tillegg til AE som prediktor gir en signifikant ΔR^2 for teori-kriteriet, men lite for de to andre kriteriene. I steg 2 brukes gjennomsnittsskåren for papirtestene i Fase 1 som eneste prediktor i Tabell 3, mens analysen presentert i Tabell 4 benytter de to prediktorene GT og TR, som begge er en del av nevnte gjennomsnittsskåre. I begge analyser ser vi i steg 2 signifikante verdier opp mot kriteriene for samlet flyprestasjon og samlet teoriprestasjon, men lite opp mot det dikotome *flyger*-kriteriet. Prediktoren 12 TR har en signifikant β -verdi på .18 mot kriteriet 36 Flykarakter. Når vi i steg 3 legger til resultater fra opptakets Fase 2 som prediktorer – henholdsvis med gjennomsnittsskåre for datatestene i Tabell 3 og de tre gruppegjennomsnittene PK, SE og IP i Tabell 4 – viser analysene at dette har størst betydning opp mot kriteriet for samlet teoriprestasjon. Gjennomsnittsskåren for datatestene gir som vist i Tabell 3 signifikante verdier for β og ΔR^2 mot teori-kriteriet, på henholdsvis .14 og .02. I Tabell 4 gir steg 3 også en signifikant ΔR^2 mot teori-kriteriet, og en signifikant β -verdi på .24 for prediktoren 18 PK. Steg 3 i Tabell 4 er også signifikant for det dikotome kriteriet 39 Flyger. Prediktoren 18 PK har også her en signifikant β -verdi, på .14. Prediktorene i dette steget av regresjonsanalysen kan også synes å tilføre noe forklaringskraft opp mot variasjonen i kriteriet for samlet flyprestasjon (36), men ikke nok til at det er statistisk signifikant. Når intervjuprognosene benyttes i steg 4, ser vi av begge tabeller (3 og 4) at ingen verdier er statistisk signifikante. Intervjuprognosene kan derfor se ut til å tilføre liten ytterligere forklaringskraft for variansen i de tre valgte kriteriene, utover det som kan forklares gjennom foregående tester i opptaket.

Verdier for total R^2 i både Tabell 3 og Tabell 4 indikerer at en relativt stor andel av variansen i prestasjonskriteriet for teoretiske fag kan tilskrives variansen i skårene på opptaksprøvene som helhet. Dette er i noe mindre grad tilfelle for variansen i kriteriet for samlet flyprestasjon, og i enda mindre grad for variansen i det dikotome *bestått/ikke bestått*-kriteriet *flyger*. Det kan altså i noen grad se ut til at opptaksprøvene samlet sett bidrar til å predikere variasjon i prestasjonskriteriene ved LFS, og da særlig innenfor teoretiske fag og i noe mindre grad innenfor praktisk flyging.

7.4 Styrker og svakheter ved studien

7.4.1 Reliabilitet generelt

Studien benytter en mengde data generert gjennom opptaksprøver i ulike former, teorieksamener og praktisk flyging. Alle disse ulike formene for testing har mulige svakheter ved seg som kan bidra til å skape avvik mellom *sann* og *observert* skåre. Studien har ikke gått inn på å undersøke datamaterialets reliabilitet, og kan derfor si lite sikkert om denne. Det kan dog være nyttig å peke på noen forhold som kan tenkes å ha innvirkning på reliabilitet. Påfølgende avsnitt vil gå nærmere inn på forhold som kan tenkes å kunne påvirke reliabilitet spesifikt for delene av datasettet som er knyttet til henholdsvis opptaksprøvene og karakterene fra uttaksflygingen, men først noen forhold som kan ha innvirkning generelt, knyttet til menneskelige faktorer. Kandidatene i utvalget er mennesker, som påvirkes ulikt i ulike testsituasjoner. Nervøsitet, stress, forventningspress, søvn og restitusjon, ernæring, dagsform, personlige forberedelser og grad av motivasjon, er noen eksempler på en mengde menneskelige faktorer som kan tenkes å innvirke på hvor tett opptil sitt fulle potensial en kandidat presterer i en testsituasjon (American Educational Research Association, 2014).

Registrering og sammenkobling av data er delvis manuelt utførte prosesser og kan derfor inneholde en risiko for menneskelige feil.

7.4.2 Reliabilitet opptaksprøver (prediktorer)

Det kan virke sannsynlig at opptaksprøvene generelt har en høy grad av reliabilitet. De gjennomføres ved et eget senter på Rygge tiltenkt formålet, på en standardisert måte, i kontrollerte omgivelser, ved bruk av utstyr som jevnlig funksjonstestes og kvalitetssikres, og under oppsyn av kvalifisert personell som har som hovedoppgave å drive seleksjon. De samme testene har blitt benyttet gjentatte ganger over lang tid. Skåring foregår etter etablerte standarder. Testene som brukes som prediktorer i undersøkelsen har vært tilnærmet like for hele utvalget. Databasert testing har vært fullt implementert i den aktuelle perioden.

Kandidatenes grad av kjennskap til innholdet i testene før gjennomføring er en usikkerhetsfaktor. En slik kjennskap vil kunne gi en kandidat et mulighetsrom for å gjøre spesifikke forberedelser som kan bidra til en kunstig høy skåre, og på den måten svekke en tests reliabilitet. LSES tilstreber en størst mulig grad av hemmelighet rundt testene for å minimere dette mulighetsrommet, gjennom ikke

åpent å publisere detaljert informasjon om testene³², og gjennom å oppfordre alle kandidater til ikke å spre informasjon om testene til andre. Det er ukjent hvor effektive disse tiltakene er.

Lang-Ree gjennomførte en liten studie for å undersøke test-retest-reliabilitet på fire av testgruppene i opptaksprøvene, med om lag 18 måneder mellom gjennomføringene (Forsvarets Sanitet, Seksjon for militærpsykologi, 2022b). Utvalget var på $N = 40 - 50$. Funnene i studien er test-retest-korrelasjonsverdier (rtt) på mellom .43 og .88 for enkelttestene, og mellom .60 og .85 for gruppegjennomsnittene (GT $rtt = .85$, PK $rtt = .67$, SE $rtt = .80$, IP $rtt = .60$, MDATA $rtt = .69$, MTOTAL $rtt = .77$). Jevnt over viser gruppegjennomsnittene en høyere rtt enn enkelttestene, og befinner seg i kategorien «adequate» ($0.60 \leq rtt < 0.70$) eller høyere, i henhold til etablerte standarder for kategorisering av test-retest-reliabilitet (European Federation of Psychologists' Associations, 2013). Endringen i gjennomsnittet av MTOTAL-skårer er fra 5.46 til 6.30, og økningen er noe sterkere for datatestene enn for papirtestene. Studien har åpenbare svakheter i form av et lite utvalgt, det faktum at utvalget ved andre gangs gjennomføring har hatt kjennskap til innholdet i testene, at tiden mellom gjennomføringene har gitt kandidatene god anledning til å gjøre forberedelser, og at kandidatene i tiden mellom gjennomføringene kan antas å ha hatt en stigende interesse for fly og flyging som kan ha medført at de har tilegnet seg kunnskap og ferdigheter relatert til dette feltet, eksempelvis gjennom bruk av simulator. Studien kan likevel synes å danne grunnlag for en styrket antakelse om at testene har en rimelig høy grad av test-retest-reliabilitet.

7.4.3 Reliabilitet prestasjonsmål ved LFS (kriterier)

LFS er en profesjonell, militær avdeling som kan virke å være kjennetegnet av høy grad av standardisering og dedikasjon til oppdraget om å selektene kandidater til flygerutdanningen. På den bakgrunn kan det virke sannsynlig at reliabiliteten for kriteriene i studien generelt er høy. Det er likevel noen aspekter ved denne typen prestasjonsmålinger som vil kunne bidra til å svekke reliabilitet. LFS-instruktører er, i likhet med kandidatene, mennesker. Instruktørene vil alltid ha noen grad av subjektivitet i sine vurderinger av prestasjoner. Ulike instruktører vil kunne bedømme en identisk prestasjon ulikt, uansett hvor god standardiseringen er. Instruktører med ulik personlighet, alder, erfaring og utstrålt autoritet vil også kunne påvirke en kandidat ulikt, og på den måten påvirke hvor tett opptil sitt fulle potensial en kandidat greier å prestere på en gitt manøver eller flytur. Kandidater med tidligere flyerfaring vil ha et annet utgangspunkt ved LFS enn kandidater som ikke har denne typen erfaring. Det er rimelig å tenke seg at tidligere flyerfaring vil kunne bidra til bedre prestasjoner både innen teoretiske fag og innen praktisk flyging. Hvor mye påvirkning denne faktoren

³² Se også avsnittet *Redusert detaljgrad* for vurderinger rundt dette aspektet i forbindelse med denne studien.

har på prestasjoner, om den hensyntas i instruktørens vurderinger og eventuelt blir forsøkt kompensert for, er ikke undersøkt. De ulike flyskrogene som brukes til uttaksflyging kan tenkes å ha visse særegenheter og individuelle forskjeller, utviklet gjennom ulike påkjenninger fra tusenvis av flytimer over mer enn 40 år. Gruppedynamikken i et LFS-kull vil kunne spille inn på prestasjoner. En gitt kandidat med et gitt utgangspunkt kan eksempelvis tenkes å ville prestere bedre i et velfungerende, samarbeidende og støttende kull enn den samme kandidaten ville gjort i et kull som ikke kan beskrives på samme måte. Grad av kontinuitet i flygingen kan påvirke prestasjoner, da avbrudd eksempelvis grunnet vær, instruktør- eller flytilgjengelighet kan tenkes å ville svekke muligheten for en jevn og kontinuerlig akkumulering av ferdigheter. Ulike betingelser i form av årtidsvariasjoner i temperatur, vær og lysforhold kan tenkes å ville medføre forskjeller i prestasjoner mellom vår- og høstkull ved LFS. Frafall (*ikke bestått*) og variasjon i prestasjoner kan også tenkes å skyldes andre grunner enn manglende evne til å prestere. Andre grunner kan være utfordringer knyttet til eksempelvis fysiologi og flysyke, motivasjon, støtte hos familie og nærstående, personlig overbevisning eller livssyn, disiplinærsaker, sikkerhetsklarering, med mer.

7.4.4 Utvalg

Utvalget i undersøkelsen er relativt stort sammenlignet med andre valideringsstudier innenfor flygerseleksjon, og gir tilstrekkelig statistisk power til å kunne avdekke statistiske sammenhenger av en viss størrelse. Det eksisterer manglende data for enkelte variabler. Dette skyldes mangelfulle rutiner for registrering bakover i tid, og det foreligger ingen grunn til å tro at det eksisterer systematisk fravær av data som vil kunne skape skjevheter i de statistiske analysene.

Utvalget er strengt selektert i flere faser, og undersøkelsen treffes derfor av problematikk knyttet til *restriction of range*³³, noe som fører til en underestimering av de faktiske eller sanne korrelasjonene.

7.4.5 Valg av prediktorer og kriterier

Det benyttes et bredt sett av prediktorer i undersøkelsen, som i stor grad gjenspeiler variablene som brukes for seleksjon frem til kandidatene starter ved LFS. Disse representerer både enkelttester, gjennomsnittsskåre for grupper av tester og overordnede gjennomsnittsskåre. Variablene som er valgt bort har enten for lav N til å være egnet for statistiske analyser (eksempelvis FOS-skåre), eller regnes som ikke relevante for å besvare problemstillingen (eksempelvis medisinske undersøkelser og testing av fysisk kapasitet).

³³ Se Teori-kapittelet

Kriteriene gjenspeiler i stor grad den informasjonen som fremkommer i kullrapportene som sendes fra sjef LFS til LSES, og gir et bilde på prestasjoner med en rimelig god oppløsningsgrad. I tillegg til variablene som brukes som kriterier i denne undersøkelsen, settes det en skåre for *Militært forhold* ved LFS. Denne har ikke blitt valgt som kriterium da den antas i liten grad å være egnet som et prestasjonsmål i konteksten av denne undersøkelsen. For å øke oppløsningsgraden på kriteriene kunne samtlige karakterer fra flyturer og teorieksamener vært brukt, i tillegg til å inkludere andre grunner til frafall (enn manglende evne)³⁴ som kriterier i undersøkelsen. En økning av oppløsningsgrad av denne typen ville medført mer kompliserte analyser i tillegg til en mer ressurskrevende datainnsamling da disse dataene i mindre grad er tilgjengelig. De valgte kriteriene i undersøkelsen gir flere kontinuerlige mål på prestasjoner, i tillegg til det dikotome *bestått/ikke bestått*-kriteriet som er mest vanlig å bruke i denne typen undersøkelser.

7.4.6 Analyser

Det kan tenkes å eksistere ikke-lineære sammenhenger mellom prediktorer og kriterier. Det er kun styrken på lineære sammenhenger som undersøkes i denne studien.

Resultatene er ikke korrigert for statistiske feilkilder.

Logistisk regresjon regnes som en mer optimal analysemåte enn hierarkisk multipl regressjon når kriteriet er dikotomt, som pass/fail-kriteriet (39 Flyger) i denne studien. Sistnevnte analysemåte er dog mest vanlig å benytte i valideringsstudier innen flygerseleksjon, og måten er også valgt i denne studien for lettere å kunne sammenligne funn med tidligere studier.

7.4.7 Redusert detaljgrad

Studien publiseres med gradering *ugradert*, som vil si at den vil være tilgjengelig for å kunne leses av alle. Informasjon som kan være egnet til å beskrive forhold omkring Luftforsvarets kampkraft på ulike flysystemer, eller som på andre måter kan tenkes å ville resultere i en høyere gradering enn *ugradert*, er utelatt fra studien. Eksempelvis beskrives ikke nøyaktige årstall eller hvor stor andel av utvalget i studien som har fullført den videre flygerutdanningen og hvilke systemer de er fordelt på.

Å publisere en for detaljert omtale av testene som benyttes, vil kunne undergrave seleksjonsprosessen gjennom å bidra til å redusere testenenes måleegenskaper³⁵, og på den måten svekke prediktiv validitet. Hensynet til denne uønskede effekten veier tyngre enn den ytterligere

³⁴ Se avsnittet *Reliabilitet prestasjonsmål ved LFS (kriterier)*.

³⁵ Nærmere beskrevet i avsnittet *Reliabilitet opptaksprøver (prediktorer)*.

verdien mer detaljerte beskrivelser kan tenkes å ville tilføre studien, og detaljgraden i beskrivelser av testmetoder har derfor blitt holdt på et overflatisk nivå.

7.4.8 Egne bias

Forfatteren har selv bakgrunn som flyger i Luftforsvaret. Seleksjonen fant sted på et tidspunkt som ligger utenfor denne studien, så egne prestasjoner er ikke en del av datamaterialet.

Seleksjonsprosessen er dog i så stor grad lik at jeg kan sies å ha en førstehånds opplevelse av gjennomføringen. I årene 2009 og 2010 hadde jeg stilling som instruktør ved LFS, og kan derfor tenkes å ha hatt en indirekte påvirkning på deler av datamaterialet. Disse erfaringene har gitt grunnlag for en forståelse og en nysgjerrighet for seleksjonsprosessen. Erfaringene kan tenkes å kunne bidra både til potensielt feilaktige antakelser om at prosessen foregår på samme måte nå som da, og til en bias i retning av en oppfatning av stort samsvar mellom prediktorene i opptaksprøvene og prestasjonskriteriene ved LFS og at disse variablene til sammen danner et fullkomment bilde av hva som definerer en god fremtidig flyger. Denne typen tankemønstre er forsøkt lagt til side for å unngå at disse kommer i veien for en åpen, nøytral og vitenskapelig tilnærming til studien, samtidig som at erfaringene brukes som en hjelp til å forstå feltet som er gjenstand for studien.

8 Konklusjon

Studien har på bakgrunn av et datamateriale som innbefatter 505 kandidater til flygerutdanning i Luftforsvaret over en lengre periode, påvist sammenhenger mellom resultater fra opptaksprøvene og resultater fra uttaksflygingen ved LFS i Bardufoss gjennom korrelasjons- og regresjonsanalyser. I opptaksprøvene er det de kognitive- og psykomotoriske testene, i tillegg til intervjuprognoser, som har blitt benyttet som prediktorer i denne studien. Som kriterier har studien benyttet ulike mål for prestasjoner innen praktisk flyging og teoretiske fag. De påviste sammenhengene mellom prediktorer og kriterier indikerer at testene i rimelig grad kan hevdes å fungere som prediktorer for fremtidige flygeskoleprestasjoner målt gjennom kriteriene ved LFS. På den måten kan opptaksprøvene sies å ha en prediktiv validitet for flygeskoleprestasjoner. Funnene i studien anses jevnt over å være i tråd med det som vil kunne være rimelig å forvente på bakgrunn av tidligere studier innenfor feltet flygerseleksjon. Sammenhengene ser ut til å være tydelige for prestasjonskriteriet innenfor teoretiske fag, og noe mindre tydelige for prestasjonskriteriene innenfor praktisk flyging. For kriteriet *bestått/ikke bestått uttaksflyging* er sammenhengene jevnt over lavere, i tråd med det som er rimelig å forvente for et dikotomt kriterium. Gjennomsnittsskårer for grupper av tester i opptaket som er ment å måle samme eller tilnærmet samme evne hos kandidaten, ser ut til å være jevnt over gode prediktorer for LFS-prestasjoner, med ett unntak som kan virke noe oppsiktsvekkende. Resultater av testene som er ment å måle evne til informasjonsprosessering ser ikke ut til å ha sammenheng med variasjonen i oppnådde resultater ved LFS. Dette er et funn som kan gi grunnlag for videre forskning knyttet til målemetodene og kriteriene som brukes ved LFS og hvorvidt disse er egnet til å selektere kandidater med de evnene som gir best forutsetninger for å prestere godt som flyger innenfor et bredt og stadig nyere, mer teknologisk avansert og trolig mer informasjonsprosesseringskrevende spekter av flytyper i Luftforsvarets inventar.

Studien finner at intervjudelen av opptaket isolert sett ser ut til å tilføre svært liten forklaringskraft for variasjonen i prestasjoner ved LFS, utover det som kan forklares gjennom de foregående testene i seleksjonsprosessen. På den bakgrunn kan det være fristende å trekke en slutning om at intervjuet har lite for seg. Dog kan intervjuet tenkes å være et viktig bidrag til seleksjonsprosessen samlede prediktive validitet når det anses som en del av en helhet heller enn et isolert ledd i seleksjonen. Det er også viktig å huske at utvalget i studien er begrenset til kandidater som har kommet videre fra intervjuet og fått anledning til å starte uttaksflygingen ved LFS, og at studien kun sier noe om variasjonen i prestasjoner blant disse kandidatene. Studien ser altså ikke på kandidater som har blitt selektert bort i intervjuet, og det kan tenkes at intervjuets hovedfunksjon er å luke ut disse uegnede kandidatene, heller enn å produsere nøyaktige prediksjoner av fremtidige prestasjoner. Intervjuet

kan også bidra til å påvirke kandidatens selvinnsikt, motivasjon og opplevelse av en rettferdig seleksjon.

Studiens regresjonsanalyser kan synes å gi grunnlag for å hevde at opptaksprøvene som har blitt benyttet for seleksjon av flygere i Luftforsvaret i den aktuelle perioden, samlet sett bidrar til å predikere variasjon i kriteriene for prestasjoner ved LFS, og da særlig innenfor teoretiske fag og i noe mindre grad innenfor praktisk flyging.

Studien har sine styrker i et stort datagrunnlag analysert etter metoder som representerer en norm innen forskning på feltet flygerseleksjon, og bruk av flere kontinuerlige prestasjonskriterier enn det dikotome *bestått/ikke bestått* som er mest vanlig å bruke i denne typen studier. Studien har også mulige svakheter som gjør at innholdet ikke nødvendigvis kan ansees som fulle og hele sannheter om flygerseleksjonsprosessen i Luftforsvaret. Analyseresultatene er ikke korrigeret for statistiske feilkilder.

9 Praktisk nytteverdi og videre studier

Studien kan bidra til en fornyet og oppdatert innsikt i seleksjonsprosessen til flygerutdanningen i Luftforsvaret, både for ledelse og beslutningstakere, for instansene som er direkte involvert i prosessen, for andre miljøer som har sitt virke innenfor seleksjon og for forsknings- og kunnskapsmiljøer med interesse for feltet. Studien kan sees som en bekreftelse på at opptaksprøvene ser ut til å være rimelig gode prediktorer for prestasjoner under uttaksflygingen. Samtidig belyser studien sammenhenger, eller kanskje snarere *manglende* sammenhenger, mellom informasjonsprosesseringsevne og kriteriene som benyttes under uttaksflygingen. Dette bør gi grunn til å stille spørsmålstegn ved om den siste delen av seleksjonsprosessen – uttaksflygingen ved LFS på Bardufoss – bruker målemetoder som er egnet til å fange opp de egenskapene som er viktige for å lykkes som flyger på et bredt spekter av flytyper som er i stadig fornyelse. Harlem gjorde sin jobbanalyse av F-35 (2016), som gir verdifull innsikt i hvilke egenskaper som i størst grad kan ha betydning for å lykkes som flyger på Norges nye jagerfly. Tilsvarende studier for andre flytyper kan tenkes å ville gi et vitenskapelig grunnlag for å mene noe om hvilke egenskaper som bør være viktige å bruke som kriterier dersom opptaksprøvene eller målemetodene ved LFS skulle vurderes endret. Det har foregått overganger fra eldre til nyere teknologi ved flere av Luftforsvarets flyoperative avdelinger de siste tiårene, uten at det nødvendigvis eksisterer god kunnskap om hvilke endrede krav til egenskaper som følger med den nye teknologien. Eksempelvis har C-130E blitt byttet ut med C-130J (taktiske transportfly), P-3 er i ferd med å byttes ut med P-8 (maritime patruljefly), og nye helikopter vil sannsynligvis erstatte det nylig skrinlagte NH-90 for operasjoner på kystvakt og fregatt. Jobbanalyser av disse nyere systemene vil kunne tilføre kunnskap som bidrar til et grunnlag for å kunne videreutvikle seleksjonsprosessen. Kunnskap om disse kravene som ligger i «den spisse enden» vil kunne være verdifulle for videreutvikling av hele seleksjonsprosessen, både opptaksprøvene og metodene som brukes ved LFS. Damos (1996) hevder at testbatterier brukt til seleksjon av militære flygere bør valideres opp mot prestasjonskriterier i jobben som flyger, men finner at majoriteten av denne typen testbatteri er lagt opp for å predikere prestasjoner under utdanning. Noe av forklaringen hevdes å være at det er problematisk å etablere reliable prestasjonskriterier for militære flygere, hvor jobben spenner fra hjemlig trening og øving i fredstid til skarpe operasjoner i krig og konflikt. Mange ulike faktorer vil spille inn på graden av suksess oppnådd i krig og konflikt, og det kan være metodisk krevende å frembringe et godt vitenskapelig grunnlag for etableringen av gode prestasjonskriterier som dekker hele spennet. Det er likevel mulig å tenke seg

at jobbanalyser som sier *noe*, om ikke alt, om krav i jobben som flyger på Luftforsvarets ulike flytyper, vil kunne gi verdifull innsikt.

Forskning på sammenhenger mellom personlighetstrekk og flygerprestasjoner gir få tydelige konklusjoner (Martinussen et al., 2022). Personlighetstrekk vektlegges ulikt i flygerseleksjon i ulike land, og brukes ikke i Luftforsvarets flygerseleksjon utover at resultatene av en personlighetstest inngår som ett av flere redskaper som kan anvendes i forbindelse med gjennomføring av intervjuet. Det betyr at det finnes noe data om personlighetstrekk for kandidater til flygerutdanningen i nyere tid, som det ville vært mulig å gjøre vitenskapelige undersøkelser av. Undersøkelser av sammenhenger mellom personlighetstrekk og prestasjonskriterier ved LFS, ved institusjoner senere i utdanningsløpet eller ved operative avdelinger kan tenkes å ha potensiale til å gi interessante funn.

Kravene til progresjon under uttaksflygingen er endret med virkning fra 2022³⁶. Studier som søker å avdekke i hvor stor grad disse endringene påvirker nivået på prestasjoner senere i utdanningsløpet eller videre i karrieren som flyger, vil kunne bidra med verdifull kunnskap. Eksempelvis kan man se for seg at en undersøkelse som sammenligner resultater oppnådd av norske elever ved utdanningsinstitusjoner i USA før og etter at endringene fikk effekt, vil kunne peke på endrede uttaksflygingskrav som en mulig forklaring. Det tar riktig nok noe tid fra endringene trådte i kraft til det vil være mulig å se konsekvenser, da kandidatene har et relativt langt utdanningsløp foran seg og det tar tid å få gjennom nok kandidater til at datamaterialet vil være egnet for statistiske analyser.

³⁶ Se avsnittet *Hvordan utvikler flygerseleksjon i Luftforsvaret seg?* i innledningen.

Forkortelser

AE	Alminnelig evnenivå
BARS	Behaviourally anchored rating scales
CD	Contact Dual (visuell flyging, med instruktør)
CVT	Control of velocity
DTG	Determinasjonsgerät
FMI	Flymedisinsk institutt (Oslo)
FOS	Forsvarets opptak og seleksjon (felles for alle kandidater til utdanning og tjeneste i Forsvaret)
FPVS	Forsvarets personell- og vernepliktssenter (Hamar)
GT	Generell teorikunnskap
IATA	The International Air Transport Association (den internasjonale organisasjonen for lufttransport)
IP	Informasjonsprosessering
LFS	Luftforsvarets flygeskole (Bardufoss)
LKSK	Luftkrigsskolen (Trondheim)
LSES	Luftforsvarets rekrutterings- og seleksjonssenter (Rygge)
NSD	Norsk senter for forskningsdata
PK	Psykomotorisk koordinasjon
RR	Restriction of range
TR	Teknisk innsikt og romopplevelse
SE	Spatiale evner
SMA	Sensory motor apparatus
UiT	Universitetet i Tromsø

Litteraturliste

- ALMamari, K., & Traynor, A. (2019). Multiple test batteries as predictors for pilot performance: A meta-analytic investigation. *International Journal of Selection and Assessment*, 27(4), 337–356. <https://doi.org/10.1111/ijsa.12258>
- American Educational Research Association (Red.). (2014). *Standards for Educational & Psychological Testing*. American Educational Research Association.
- Burke, E., & Valsler, C. (1999). The Pilot Aptitude Tester (PILAPT) Handbook. *People Technologies*.
- Carretta, T. R. (1992). Understanding the Relations Between Selection Factors and Pilot Training Performance: Does the Criterion Make a Difference? *The International Journal of Aviation Psychology*, 2(2), 95–105. https://doi.org/10.1207/s15327108ijap0202_2
- Carretta, T. R., & Ree, M. J. (2003). *Pilot selection methods*.
- Carretta, T. R., Rodgers, M. N., & Hansen, I. (1993). *The identification of ability requirements and selection instruments for fighter pilot training*. ARMSTRONG LAB BROOKS AFB TX.
- Carretta, T. R., Teachout, M. S., Ree, M. J., Barto, E. L., King, R. E., & Michaels, C. F. (2014). Consistency of the Relations of Cognitive Ability and Personality Traits to Pilot Training Performance. *The International Journal of Aviation Psychology*, 24(4), 247–264. <https://doi.org/10.1080/10508414.2014.949200>
- Clark-Carter, D. (2005). Stanine Scores. I B. S. Everitt & D. C. Howell (Red.), *Encyclopedia of Statistics in Behavioral Science* (s. bsa640). John Wiley & Sons, Ltd. <https://doi.org/10.1002/0470013192.bsa640>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences* (0 utg.). Routledge. <https://doi.org/10.4324/9780203771587>
- Damos, D. L. (1996). Pilot Selection Batteries: Shortcomings and Perspectives. *The International Journal of Aviation Psychology*, 6(2), 199–209. https://doi.org/10.1207/s15327108ijap0602_6

Damos, D. L., Eaglestone, J., Hörmann, H.-J., Stadler, K., & Wium, J. (2022). Selection in Aviation.

European Association for Aviation Psychology.

Dunbar, S. B., & Linn, R. L. (1991). Range restriction adjustments in the prediction of military job performance. I *Performance assessment for the workplace, Vol. 1; Vol. 2: Technical issues.* (s. 127–157). National Academy Press.

European Federation of Psychologists' Associations. (2013). *EFPA review model for the description and evaluation of psychological tests.* European Federation of Psychologists' Associations. <http://assessment.efpa.eu/documents/>

Fleishman, E. A. (1975). Toward a taxonomy of human performance. *American Psychologist, 30,* 1127–1149. <https://doi.org/10.1037/0003-066X.30.12.1127>

Fleishman, E. A., & Reilly, M. E. (1992). *Handbook of human abilities: Definitions, measurements, and job task requirements.* (s. ix, 132). Consulting Psychologists Press.

Forsvaret. (2022). *Flyger- og navigatør--utdanning.*

<https://www.forsvaret.no/utdanning/utdanninger/flyger>

Forsvarets Sanitet, Seksjon for militærpsykologi. (2022a). *Intervjuguide LFS.*

Forsvarets Sanitet, Seksjon for militærpsykologi. (2022b). *Testbatteriet for seleksjon av flygere ved LSES på Rygge: Beskrivelse av metoder og prosedyrer for gjennomføring av psykologiske tester ved opptak til LFS.*

Forsvarsdepartementet. (2018). *Prop. 1 S.* <https://www.regjeringen.no/no/dokumenter/prop.-1-s-20182019/id2613164/>

Hansen, I. (1987). Psykologiske metoder ved flygeruttak. Del A; Beskrivelse av testene.

Militærpsykologiske Meddelelser, 15. Forsvarets psykologiske og pedagogiske senter.

Hansen, I., Heim, J., Nordby, K., Storsve, O., & Værnes, R. (1989). Psykologiske metoder ved flygeruttak. Del B: Utprøving av nytt testbatteri. *Militærpsykologiske Meddelelser, 16.*

Forsvarets psykologiske og pedagogiske senter.

-
- Harlem, T. O. (2016). *Seleksjon flygere F-35: Neste generasjon jagerfly, en ny generasjon flygere?* [Masteroppgave, Forsvarets høgskole]. <https://fhs.brage.unit.no/fhs-xmlui/handle/11250/2436208>
- Hilton, T. F., & Dolgin, D. L. (1991). *Pilot selection in the military of the Free World*. John Wiley & Sons.
- Hunter, D. R. (1989). Aviator selection. I *Military personnel measurement: Testing, assignment, evaluation*. (s. 129–167). Praeger Publishers.
- Hunter, D. R., & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *The International Journal of Aviation Psychology*, 4(4), 297–313.
- Hunter, D. R., & Burke, E. F. (1995). *Handbook of pilot selection*. Avebury Aviation Brookfield, VT.
- International Air Transport Association. (2019). *Pilot Aptitude Testing, Guidance Material and Best Practices* (3rd Edition).
- King, R. E. (2014). Personality (and Psychopathology) Assessment in the Selection of Pilots. *The International Journal of Aviation Psychology*, 24(1), 61–73.
<https://doi.org/10.1080/10508414.2014.860844>
- Lang-Ree, O. C., Martinussen, M., & Ødegaard, P. E. (2010). *Pilot selection in the Norwegian Air Force: A validation study*. Poster presented at the European Association for Aviation Psychology 29th Conference, Budapest, Hungary.
- Martinussen, M. (1996). Psychological Measures As Predictors of Pilot Performance: A Meta-Analysis. *The International Journal of Aviation Psychology*, 6(1), 1–20.
https://doi.org/10.1207/s15327108ijap0601_1
- Martinussen, M. (2005). Seleksjon av flygere og flygeledere. *Tidsskrift for Norsk psykologforening*.
<https://psykologtidsskriftet.no/fagartikkel/2005/04/seleksjon-av-flygere-og-flygeledere>
- Martinussen, M. (2017). Pilot selection: An overview of aptitude and ability assessment. I *Pilot mental health assessment and support: A practitioner's guide*. (s. 23–39). Routledge/Taylor & Francis Group.

-
- Martinussen, M., & Hunter, D. R. (2018). *Aviation psychology and human factors* (Second edition). CRC Press, Taylor & Francis Group, CRC Press is an imprint of the Taylor & Francis Group, an informa business.
- Martinussen, M., Lang-Ree, O. C., Mjøen, H., Svendsen, B., & Barone, A. (2022). Predicting Commercial Pilot Training Performance. *Aviation Psychology and Applied Human Factors*. <https://doi.org/10.1027/2192-0923/a000232>
- Martinussen, M., & Torjussen, T. (1998). Pilot Selection in the Norwegian Air Force: A Validation and Meta-Analysis of the Test Battery. *The International Journal of Aviation Psychology*, 8(1), 33–45. https://doi.org/10.1207/s15327108ijap0801_2
- Martinussen, M., & Torjussen, T. M. (2004). Initial validation of a computer-based assessment battery for pilot selection in the Norwegian Air Force. *Human Factors and Aerospace Safety*, 4, 233–243.
- Schmidt, F. L., Oh, I.-S., & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years. *Fox School of Business Research Paper*, 1–74.
- Svensson, K. (2013). *Flygarseleksjon i Forsvaret: Ein studie av seleksjonssystemet sin prediktive validitet* [Masteroppgave, Forsvarets høgskole]. <https://fhs.brage.unit.no/fhs-xmlui/handle/11250/100112>
- Torjussen, T. M., & Hansen, I. S. (1999). Forsvaret, best i test? Bruk av psykologiske tester i Forsvaret, med spesiell vekt på flygarseleksjon. *Tidsskrift for norsk psykologiforening*, 36, 772–779.

Vedlegg

Godkjenning fra FHS forskningsnemnd



Vår saksbehandler

Audun Benjamin Bengtson, aubengtson@mil.no
+47
FHS/FAGSTAB/SEK FOU ADM

Vår dato

2022-10-18

Vår referanse

2022/039358-002/FORSVARET/ 910

Tidligere dato

Tidligere referanse

Til

Harald Grindheim
.
..

Kopi til

FSAN/FSANFAG/IMA/SEKSJ MILPSY/Ole Christian Lang-Ree

Tillatelse til å innhente opplysninger i og om Forsvaret til forskningsformål

1 Bakgrunn

Forsvarets høyskole (FHS) har mottatt din søknad av 30. september 2022 om tillatelse til å innhente opplysninger i og om Forsvaret til forskningsformål. Prosjektet det skal innhentes data til er en masteroppgave, og følgende problemstillinger er oppgitt: «I hvilken grad er det sammenheng mellom de kognitive og psykomotoriske testene og resultatene oppnådd ved Luftforsvarets flygeskole (LFS)? I hvilken grad er det sammenheng mellom intervjuprognosene og resultatene fra LFS, og hvor mye bidrar disse til å forklare variasjonen i prestasjoner utover det som forklares ved hjelp av testene og skolepoeng?». Det søkes om tilgang til Forsvarets registerdata og tillatelse er innhentet ved Ole Christian Lang-Ree (FSAN).

2 Drøfting

Vurdering av søknader om tillatelse til å innhente opplysninger i og om Forsvaret til forskningsformål er regulert av *Bestemmelse om utlevering av personopplysninger til forskning og gjennomføring av spørreundersøkelser*, fastsatt av sjef HR-avdelingen i Forsvarsstaben 1. mai 2018.

I henhold til punkt 2.3 og 2.4 i denne bestemmelsen er det en forskningsnemnd oppnevnt av sjef FHS som har myndighet til å behandle søknader om tillatelse til datainnsamling i Forsvaret. Kriterier og rettsgrunnlag som skal legges til grunn for vurderingen er omtalt i punkt 4.1 og 4.2.

Forskningsnemnda har vurdert din søknad som tilfredsstillende i henhold til gjeldende krav.

3 Vedtak

Søknad om tillatelse til å innhente opplysninger i og om Forsvaret til forskningsformål innvilges. Tillatelsen gjelder til prosjektslutt 1. juni 2023.

4 Vilkår for tillatelsen

Det er kun gitt tillatelse til innhenting av det datamaterialet som fremgår av søknaden. Data hentet fra Forsvaret skal ikke benyttes til andre formål enn den aktuelle masteroppgaven. Ved prosjektslutt skal alle data hentet fra Forsvaret slettes. Det skal sendes sluttmelding til FHS vedlagt masteroppgaven. Sluttmelding sendes til fhs.datautlevering@mil.no

Postadresse

Postboks 800 Postmottak
2617 Lillehammer
Norge

Besøksadresse

Oslo mil/Akershus
0015 OSLO
Norge

Sivil telefon/telefaks

/

Militær telefon/telefaks

99/0500 3699

Epost/ Internett

postmottak@mil.no
www.forsvaret.no

Organisasjonsnummer
NO 986 105 174 MVA

Vedlegg

1

Sven Gabriel Holtmark

Leder forskningsnemnda

Dokumentet er elektronisk godkjent, og har derfor ikke håndskreven signatur.
