



**Forsvarets høgskole**

**våren 2013**

**Masteroppgåve**

## **Flygarseleksjon i Forsvaret**

*Ein studie av seleksjonssystemet sin prediktive validitet*

**Kurt Svensson**



## Forord

Masterstudiet ved Forsvarets Høgskole har gjeve meg høve til å studera flygarseleksjonssystemet i Forsvaret på nært hald. Interesse for dette feltet har lege der latent sidan eg sjølv vart utsatt for dei same opptaksprøvene på 80-talet og etter det fekk høve til å fly både i Luftforsvaret, i U. S. Air Force og i SAS. Seinare har eg òg fått høve til å delta i flygarseleksjonsprosessen som intervjuoffiser. I det arbeidet har eg stadig undra meg over kva det er med opptaksprøvene som gjer at Forsvaret vel å satsa på visse kandidatar framfor andre.

Flygarseleksjonssystemet som vart nytta på 1980- og 90-talet er undersøkt for prediktiv validitet opp mot Euro-NATO Joint Jet Pilot Training (ENJJPT) -flygarprestasjonskriteria. Kandidatane som vart plukka ut til jagarflygarutdanninga vart strengt selektert på ulike evneområde. Dette har mellom anna bidrege til at suksessraten for norske flyelevar ved ENJJPT vart 99 %. Det har vore spanande å studera kva psykologitenesta i Forsvaret har lagt til grunn for val av seleksjonstestar og metodar. Studien har vist at metodane som vart nytta har bidrege til eit seleksjonssystem med god prediktiv validitet og effektivitet i høve produksjon av jagarflygarar til Luftforsvaret.

Eg ynskjer å retta ei stor takk til veileiar professor Monica Martinussen ved Universitetet i Tromsø/Forsvarets Høgskole for framifrå og effektiv veileiing. Eg vil òg takka bibliotekar Nina Riege ved FHS for naudsynt assistanse i høve innsanking av litteratur. Takk til Ole-Christian Lang-Ree, Sjefpsykologen for Forsvaret, for stønad med datahandsaming og gode innspel til oppgåva. I tillegg vil eg takka kona mi Anne for korrekturlesing dagane før levering.

Kurt Svensson

Forsvarets Høgskole 24. mai 2013

## Abstract

The purpose of this work was to conduct a scientific analysis of Norwegian fighterpilot selection and training. The sample used in this study were Norwegian students graduated from Euro-NATO Joint Jet Pilot Training (ENJJPT), Sheppard Air Force Base, TX, between 1989-2004 ( $N = 175$ ). The Norwegian student pilot (SP) graduation rate was 99 % throughout these years as only two SPs were eliminated from ENJJPT due to lack of performance.

This study demonstrated that several aptitude tests which were administered during the selection process had predictive validity for ENJJPT T-37 and T-38 Daily Flying grades, but not for T-37 and T-38 Checkride grades (flying exams). Nor did a measure of General mental ability (GMA) provide predictive validity for any performance criteria. However, a different measure of GMA, called *Alminnelig evnenivå/AE*, or general intelligence, which was mainly used for pre-screening purposes, demonstrated predictive validity for both academic results (flying theory) and daily flying performance with the T-38 advanced jet trainer.

The discussion suggests there are probable explanations to these findings. First, the checkride criterion seems to be challenged by reduced reliability that could affect Norwegian students more than others as relative rank in class is irrelevant for future assignment to Main Weapon System following ENJJPT Graduation. Second, the candidates were heavily screened on the GMA-measure leaving little variation left for statistical analysis due to the Restriction of Range phenomena. However, the other measure of general intelligence was not affected by this effect to the same degree.

The regression analysis demonstrated that most of the variance in daily flying performance, both on T-37 and T-38, was explained by selection test groups which measured Mechanical Comprehension, Spatial Orientation, Simultaneous Capacity and Psychomotor Abilities.

## Samandrag

Føremålet med studien var å undersøke i kva grad opptaksprøvene til Luftforsvarets Flygeskole (LFS) hadde prediktiv validitet for seinare flygarprestasjonar. Resultata oppnådd på desse prøvane vart korrelerte med utdanningsresultata for dei som gjennomførte jagarflygarutdanninga ved ENJJPT mellom 1989 – 2004 ( $N = 175$ ). Prediktorane var generell intelligens, teknisk innsikt og spatiale evner, psykomotoriske evner og simultankapasitet samt overføring/retningsevne. Desse prediktorane var avgjerande for om ein kandidat vart teke opp som flyaspirant ved LFS eller ikkje. Prestasjonskriteria var teoriexamensresultat, resultat frå dagleg flyging og eksamensflyging på T-37 og T-38. Desse prestasjonskriteria var avgjerande for det samla utdanningsresultatet og rangeringa i klassen.

Seleksjonstestane Figurmønster, Sortering, Retningsreversaller og Spegeltesten viste prediktiv validitet for flygarprestasjonar under dagleg flyging. Testgruppene Teknisk innsikt/romopplevelse og Simultankapasitet (psykomotoriske evner) hadde òg prediktiv validitet for dagleg flyging. I denne studien hadde ingen prediktorar prediktiv validitet for flygarprestasjonar under eksamensflyging. Prediktoren Alminnelig evnenivå (AE) hadde prediktiv validitet for teoriresultat og prestasjonar under dagleg flyging med T-38. Desse funna stemmer i hovudsak overeins med resultat frå tidlegare forskning på militær flygarseleksjon, men nokre avvik er diskutert.

Manglande samanheng mellom prediktorane og eksamensflyginga kan skuldast fleire tilhøve. Karaktersystemet var lagt opp på ein måte som kunne påverka reliabiliteten til kriteriet negativt. Eit slikt karaktersystem krev spesielt god standardisering mellom dei som vurderar elevprestasjonane. Mellommenneskelege faktorar mellom elev og evaluator kan òg ha spela inn og påverka resultata slik at vurderingane vart utført på ein mindre objektiv og usystematisk måte. Eksamensresultata viser i tillegg liten variasjon, noko som kan bidra til manglande korrelasjon. Eksamensresultat og rangering ved ENJJPT var dessutan mindre viktig for norske elevar, då dei var innforstått med at det var opplæring på jagarfly som venta berre ein fullførte, medan medelevar som presterte svakt ikkje ville få tilbod om det. Dette kan, i samspel med dei andre faktorene, moglegvis ha påverka både elev og evaluator sin strategi for korleis ein skulle handtera ei slik evaluering. I tillegg var deltakarane i studien strengt selektert. Fenomenet knytt til *Restriction of Range* har difor avgrensa korrelasjonane.

Utover dette avdekka studien at 99 % av kandidatane som fekk tilbod om jagarflygarutdanning, etter å ha bestått opptaksprøvene og LFS seleksjonsflyging, gjennomførte ENJJPT og fekk flygarving.

## Innhold:

<b>Forord</b> .....	<b>3</b>
<b>Abstract</b> .....	<b>4</b>
<b>Samandrag</b> .....	<b>5</b>
<b>1 Innleiing</b> .....	<b>7</b>
1.1 KVA HANDLAR STUDIEN OM? .....	7
1.2 KVIFOR ER DETTE INTERESSANT Å FORSKA PÅ?.....	7
1.3 PROBLEMSTILLING.....	8
1.4 FORVENTNINGAR TIL STUDIEN .....	10
<b>2 Forsking på flygarseleksjon</b> .....	<b>10</b>
2.1 JOBBANALYSE .....	10
2.2 VALIDERINGSSTUDIAR.....	11
2.3 META-ANALYSAR .....	13
2.4 RESTRICTION OF RANGE.....	13
2.5 FORSKINGSFUNN.....	14
2.5.1 Evner som har vist prediktiv validitet .....	15
<b>3 Flygarseleksjon i Noreg og jagarflygarutdanninga i USA</b> .....	<b>21</b>
3.1 GENERELT OM GJENNOMFØRING AV SELEKSJONEN .....	21
3.2 LITT BAKGRUNN I HØVE DEI NORSKE FLYGARTESTANE.....	21
3.3 SELEKSJONSPROSESSEN – EIN MODELL.....	22
3.4 SELEKSJONSPROSESSEN I PRAKSIS.....	22
3.4.1 Fase 0.....	23
3.4.2 Fase 1 .....	23
3.4.3 Fase 2 .....	23
3.4.4 Fase 3 - Luftforsvarets Flygeskole .....	24
3.4.5 Luftkrigsskule I.....	24
3.4.6 Euro-NATO Joint Jet Pilot Training .....	25
<b>4 Metode</b> .....	<b>31</b>
4.1 DELTAKARAR .....	31
4.2 PREDIKTORAR.....	31
4.2.1 Presentasjon av testane .....	32
4.3 KRITERIER .....	37
4.4 PROSEDYRE.....	39
4.5 STATISTISKE ANALYSAR .....	39
<b>5 Resultat</b> .....	<b>40</b>
5.1 KORRELASJONSANALYSE .....	40
5.2 MULTIPPEL HIERARKISK REGRESJONSANALYSE .....	46
<b>6 Diskusjon</b> .....	<b>50</b>
6.1 DISKUSJON AV PROBLEMSTILLING 1 .....	50
6.2 DISKUSJON AV PROBLEMSTILLING 2.....	51
6.3 DISKUSJON AV PROBLEMSTILLING 3.....	63
6.4 OM FORSKINGSSPØRSMÅLET .....	64
STYRKER OG SVAKHEITER VED STUDIEN .....	66
<b>7 Konklusjon</b> .....	<b>72</b>
7.1 PRAKTISK VERKNAD.....	74
<b>8 Referansar</b> .....	<b>76</b>

## 1 Innleiing

### 1.1 Kva handlar studien om?

Denne studien tek for seg seleksjon og utdanning av norske jagarflygarar. Føremålet har vore å undersøkje om det er samanheng mellom kandidatane sine observerte evner målt under opptaksprøvene til Luftforsvarets Flygeskole (LFS) sett opp mot resultata oppnådd under jagarflygarutdanninga i USA. Tidlegare nasjonale studiar har i all hovudsak vore avgrensa til å undersøkje seleksjonssystemet sin prediktive validitet opp mot bestått/ikkje bestått seleksjonsflyging ved LFS (sjå til dømes Martinussen & Torjussen, 1998). Då denne flyginga kun utgjer nokre få flytimar for å skilja ut kandidatane som skal få tilbod om å bli flygar i Luftforsvaret er det og eit behov for å studera opptaksprøvene opp mot meir langsiktige og reelle prestasjonskriteria frå flygarutdanninga. I denne studien er utdanningsresultatet til alle norske elevar som gjekk gjennom jagarflygarutdanninga ved Euro-NATO Joint Jet Pilot Training (ENJJPT) på Sheppard Air Force Base i USA mellom 1989 – 2004 samla inn. ENJJPT-utdanninga er ikkje undersøkt tidlegare og eg ynskjer difor å koma med eit bidrag til ei betre forståing av det norske flygarseleksjonssystemet ved å gjennomføra ei ytterlegare validering. Resultata frå dette arbeidet kan kanskje bidra til eit endå betre avgjerdsgrunnlag i høve utveljing av dei presumtivt best eigna kandidatane til jagarflygarutdanning i Forsvaret.

### 1.2 Kvifor er dette interessant å forska på?

Seleksjon og utdanning av jagarflygarar er ein omstendelig og kostbar prosess for Luftforsvaret. Noko forenkla kan ein seia at det er ynskjeleg å velgja ut dei individa med ibuande eigenskapar som gjer at den einskilde vil kunna gjennomføra utdanninga ved ENJJPT suksessfullt for deretter å verta operativ på kampfly heime i Noreg. Denne prosessen tek om lag 4-5 år. Etter at kampklar (*Combat Ready*) operativ status er oppnådd på F-16 fylgjer vanlegvis fleire år med spanande og krevjande operativ flyging ved ein jagarflyskvadron (331, 332 eller 338 skv). Den operative tenesta involverer både daglege flyoperasjonar i Noreg (dagleg treningsprogram og QRA<sup>1</sup>) og stundom skarpe operasjonar ute i verda, sist gong over Libya i høve Operation Unified Protector i 2011.

Føremålet med flygarseleksjon har minst tre dimensjonar. For det fyrste prøver ein å sikra at flygarkandidaten har størst mogleg sannsynlegheit for å fullføra utdanninga. Utdanninga i USA er særskild kostbar og det vil medføra auka kostnader viss kandidatar vert teke ut av flyprogrammet og sendt heim til Noreg fordi dei ikkje presterar godt nok. Jamnt over gode flygarprestasjonar vil truleg redusera faren for å verta eliminert frå utdanninga. For det andre

---

<sup>1</sup> Quick Reaction Alert – 2 stykk F-16 på høg beredskap med skarpe våpen for suverenitetsheving av norsk luftrom.

gjeld omsynet til flytrygging (Hunter & Burke, 1995). For Luftforsvaret er det viktig at kandidatar som klarar å gjennomføra utdanninga flyg på ein trygg og god måte gjennom heile flykarriera. På dette viset vil ein betre tryggja både personell og materiell over tid. I tillegg til desse to dimensjonane legg eg til grunn at det er i Luftforsvaret si interesse at kandidatane oppnår så gode resultat som mogleg. Det er rimeleg grunn til å anta at flygarar som presterar betre enn ein minste standard har eit overskot samanlikna med andre som presterar rett rundt minstekravet. Eit slikt overskot kan gje seg positive utslag både i høve grad av situasjonsoversikt dei til ei kvar tid har, men òg det å takla uventa og krevjande situasjonar som måtte oppstå under flyging.

Forsking knytt til flygarseleksjon har identifisert fleire menneskelege evner og ferdigheitar som bidreg til at det er større sjansar for å klara ei militær flygarutdanning. Resultat frå forskning på militær flygarseleksjon gjennom fleire ti-år har vist at utdanninga til, og jobben som, flygar ser ut til å krevja eit tilfredsstillande generelt kognitivt evnenivå eller generell intelligens (*g*) om ein vil, samt andre meir spesifikke evner som psykomotoriske evner, spatiale evner, einskilde personlege eigenskapar, diverse flyrelatert kunnskap og ferdigheitar (til dømes instrumenttyding, mekanisk/teknisk innsikt og forståing av slike konsept) samt reaksjonsevne (Carretta & Ree, 2003; Hunter & Burke, 1994; Martinussen, 1997; Martinussen & Torjussen, 1998; Martinussen & Hunter, 2008; Ree & Earles, 1992). I tillegg kan det sjå ut til at nokre av desse evnene aukar sannsynligheita for at ein kandidat får tilbod om å fly jagarfly etter utdanninga (Carretta, 1989). Med utgangspunkt i forskingsresultat innan militær flygarseleksjon, med hovudvekt på perioden rundt 80- og 90- talet, ynskjer eg å bidra med empiri og analysar som kanskje kan gje meir kunnskap om flygarseleksjonssystemet i Forsvaret.

### **1.3 Problemstilling**

Trass i at seleksjons- og utdanningssystemet har produsert jagarflygarar for Luftforsvaret over fleire ti-år er det framleis rom for forbetringar. Det er dessutan viktig å gjennomføra valideringstudiar for å kvalitetssikra seleksjonssystemet. Frå 80-talet og fram til i dag har kun eit fåtal norske flyelevar vorte heimsendt utan å fullføra utdanninga ved ENJJPT. Trass i at dette har vore eit lite antal samanlikna med andre land er det like fullt bortkasta tid og ressursar, både for den einskilde kandidat og for Forsvaret. I tillegg har fleire nyutdanna flygarar ikkje meistra overgangen til F-16 og operativ flyging ved jagarflyskvadron i Noreg etter utdanninga i USA.

Kva er det som påverkar utdanningsresultatet for norske ENJJPT-elevar? Av fleire moglege faktorar er det rimeleg å anta at resultat oppnådd på opptaksprøvene (med andre ord ulikt evnenivå) kan verka inn på flygarprestasjonane. Utover dette spelar truleg andre faktorar



---

som motivasjon og ulike aspekt ved utdanninga inn. Slike faktorar er halde utanfor i denne studien.

Forsking innanfor dette området har til no vist at fleire seleksjonstestar har prediktiv validitet for prestasjonar under flygarutdanninga (Carretta & Ree, 2003; Hunter & Burke, 1995; Martinussen, 1997). Er det slik at resultata oppnådd på opptaksprøvane til LFS òg har prediktiv validitet for flygarprestasjonane under jagarflygarutdanninga ved ENJJPT?

Forskingsspørsmålet er:

*I kva grad er det samanheng mellom resultata oppnådd på seleksjonstestane og prestasjonane under flygarutdanninga ved ENJJPT?*

Problemstilling 1:

*Har einskilde seleksjonstestar betre prediktiv validitet enn andre i høve ENJJPT utdanningsresultat?*

Problemstilling 2:

*I kva grad har dei ulike testgruppene (Gr. 1 - 4) prediktiv validitet i høve ENJJPT utdanningsresultat?*

Problemstilling 3:

*Kor godt predikerer testgruppene (Gr. 1 – 4) samla sett dei ulike prestasjonskriteria ved ENJJPT?*

## 1.4 Forventningar til studien

Resultata frå denne forskinga indikerar at prediktorar som måler evner og ferdigheitlar innanfor domena generell intelligens (*g*), andre meir spesialiserte kognitive/mentale evner (som til dømes spatiale evner), samt psykomotoriske koordineringsevner og flyrelatert kunnskap (til dømes instrumenttyding, mekanisk/teknisk innsikt og forståing av slike konsept) samsvarar positivt med flygarprestasjonar under utdanninga (sjå til dømes Hunter & Burke, 1995). Eg ventar at det same gjeld for utdanninga ved ENJJPT.

## 2 Forsking på flygarseleksjon

Føremålet med dette kapitlet er å gjera greie for korleis forskning på flygarseleksjon i hovudtrekk har føregått, kva metodiske utfordringar som er hefta ved dette arbeidet og kva resultat med relevans for problemstillinga som har kome fram.

### 2.1 Jobbanalyse

Det fyrste steget i ein seleksjonsprosess er vanlegvis å identifisera sentrale jobbprestasjonsfaktorar ved hjelp av ein jobbanalyse (Carretta & Ree, 2003; Hunter & Burke, 1995). Gjennom ein slik analyse vil ein kunna avdekka kva arbeidsoppgåver som er sentrale for å gjera ein god jobb og kva eigenskapar, kunnskapar, ferdigheitlar og evner som bør liggja til grunn for å meistra desse. Kunnskap om dette kan nyttast til å velgja prediktorar og kriterier til bruk i valideringsstudiar (Carretta & Ree, 2003). Dei prediktorane ein vurderar å nytta bør òg sjåast i samanheng med prediktorar som ein over tid har kunna stola på i høve prediksjon av flygarprestasjon (Hunter & Burke, 1995).

Midt på 90-talet gjennomførte NATO ein jobbanalyse for jagarflygarar. I denne studien, utført av Carretta, Rodgers og Hansen (1993), vart 43 erfarne jagarflygarar frå Canada, Noreg og USA spurt om kva evner og ferdigheitlar dei opplevde som viktige i jobben. Resultata viste at fylgjande eigenskapar vart vurdert som særst viktige: *situational awareness*<sup>2</sup>, *memorization*<sup>3</sup> og *reasoning*<sup>4</sup>. Det som vart vurdert som minst viktig var munnleg/skriftleg framstillingsevne og evne til personalleiing (*personnel management skills*). I eit større bilete kan ein seia at flygaren si *situational awareness* òg gjeld det å heile tida ha tilfredsstillande oversikt over kvar ein er, vèrtilhøve, kommunikasjon, spatial oversikt, det taktiske biletet og kva flyet og dei ulike sensorane gjer til ei kvar tid (Holoviak, Yauch, & Ercoline, 1996).

---

<sup>2</sup> *The state of constant mental readiness in order to respond to situational changes (Carretta et al., 1993).*

<sup>3</sup> *The ability to remember information, such as words, numbers, pictures, and procedures. Bits of information can be remembered by themselves or with other pieces of information (Carretta et al., 1993).*

<sup>4</sup> *The ability to combine separate bits of information and to apply general rules in order to derive logical answers or form conclusions (Carretta et al., 1993).*

## 2.2 Valideringsstudiar

I 1903 kasta Orville og Wilbur Wright «krone og mynt» om kven som skulle fly den fyrste flyturen. Seleksjonsmetodane knytt til flygarseleksjon har utvikla seg i positiv retning etter det (Torjussen & Hansen, 1999). I 1999 kosta det rundt 15 millionar norske kroner å utdanna *ein* jagarflygar i USA (Torjussen & Hansen, 1999). Kostnadane steig vidare i høve utsjekk til *Wingman*<sup>5</sup> (ca 6 mnd) og seinare oppnåing av full kampklar status (ca 12-18 mnd) i Noreg. RAF rekna i 1992 med at deira nasjonale jagarflygarutdanning (3 års lengde) kosta £ 3 millionar pr. pilot (Turnbull, 1992) (om lag 33 mill NOK 1992-kroner). Det er rimeleg å anta at det er tale om tilsvarande kostnader for Luftforsvaret. Med eit slikt kostnadsbilete vil det naturlegvis vera ynskjeleg å velgja ut jagarflygarkandidatar som med størst mogleg sannsynlegheit kan fullføra utdanninga. Kven skal ein velgja ut til å gjennomføra ei slik utdanning? Kva metode skal ein nytta og kor gode er eigentleg testane til å predikera utkoma? Kor trygg kan ein eigentleg vera på at ein velgjer rett kandidat basert på eit slikt system?

Med valideringsstudiar søkjer ein å samla evidens for at opptaksprøvene (prediktorane) faktisk måler dei evnene som er viktige for flygarprestasjonar (Bailey & Woodhead, 1996; Martinussen, 1997). I denne samanhengen er prediktorane dei ulike metodane som vert nytta ved seleksjon, som evnetestar og intervju (Martinussen & Hunter, 2008). Ein lyt dokumentera at dei testane som skal nyttast til seleksjon har prediktiv validitet for framtidige prestasjonar (kriterium). Ei slik undersøking går ut på å visa om det faktisk er ein samheng mellom prediktor og kriterium.

I den amerikanske publikasjonen *Standards for Educational and Psychological Testing* (1999) er det gjeve retningslinjer for kva krav ein bør stilla til testar som skal nyttast for å selektera personell. Her krevjer ein at det vert framskaffa kriterierelatert evidens (prediktiv validitet er en type kriterierelatert validitet). Det betyr at ein lyt demonstrera at testskår er systematisk relatert til eit eller flere utfallskriteria for jobbprestasjon. Valet av kriterium og korleis dette skal målast vert difor sentralt. Har ein ikkje kriterierelatert evidens å visa til kan ein til ein viss grad nytta innhaldsrelatert- eller begrepsrelatert evidens (Carretta & Ree, 2003). Sagt på ein meir pragmatisk måte lyt ein dokumentera at testen måler det som vert hevda at den måler, og samstundes, at den eigenskapen eller evna har betyding for utføringa av jobben.

<sup>5</sup> *Wingman* er den fyrste operative flygarstatus nye jagarflygarar oppnår. Ein *wingman* flyg saman med ein sertifisert formasjonsleiar.

I ein lokal prediktiv valideringsstudie vert korrelasjonane mellom ulike prediktorar og ulike vurderingar av seinare jobbprestasjonar rekna ut (Martinussen & Hunter, 2008). Viss ein i valideringsstudien har tilgang på stor nok  $N$  vil ein kunna oppnå tilstrekkeleg statistisk styrke (*statistical power*) til å avdekka ein faktisk effekt. Ein er med det tryggare på at ein verkeleg effekt faktisk eksisterar og at samanhengane ikkje er tilfeldige. Ei anna utfordring med valideringsstudiar er knytt til val av kriterium. Dei valde kriteria må ha god reliabilitet og validitet i tillegg til å vera relevante for den aktuelle organisasjonen. Val av gode og relevante kriteria er heilt sentralt for at ein valideringsstudie skal kunna koma til praktisk nytte.

Det mest nytta kriteriet i høve prediksjon av flygarprestasjon har vore bestått/ikkje bestått seleksjonsflyging (i Noreg) eller *pass/fail basic flying school* internasjonalt (Hunter & Burke, 1995; Martinussen, 1989). Slike todelte kriteria fortel lite om kor godt eleven eigentleg presterte, anna enn at flyginga var bestått (Bailey & Woodhead, 1996). Ved å nytta meir reelle kriteria for flygarprestasjon vil ein kunna seia meir i høve prediktiv validitet. Moglegvis vil ein oppnå ein meir relevant test av prediktorane sidan det er dei meir reelle kriteria ein eigentleg ynskjer å predikera. Ein kan dessutan kanskje oppnå ei betre forståing av seleksjonsprosessen ved å nytta slike kriteria.

Kva kriterium bør ein i så fall gjera seg nytte av for best mogleg å måla reelle flygarprestasjonar? Hovudoppgåva til jagarflygaren er å oppnå, halda på og utnytta luftkontroll eller herredøme i lufta (Cobe, 1996). I praksisk betyr det at ein måler kampflygareffektivitet i Luftforsvaret med den einskilde F-16 pilot si oppnåing av status som kampklar i dei luftmaktrollene ein er meint å fylla. Fyrste hinderet mot kampklarstatus etter bestått LFS var jagarflygarutdanninga ved ENJJPT. Her utdanna kandidatane seg til flygarar. Sjølv om dei færraste av desse individa ville omtala seg sjølve som jagarflygarar like etter at ENJJPT var gjennomført er denne utdanninga den einaste som gjev norske flygarar moglegheit til å bli operativ på F-16 seinare.

I ein seleksjonsstudie Syversen (1960) utførte vart 164 norske flygarar som gjennomførte jagarflygarutdanning i USA og Canada i åra 1953-55 undersøkt. I denne studien vart karakterane gjennom utdanninga nytta som kriterier, med størst vekt på karakterar som dekkja den praktiske flyginga. Karakterane (resultata) frå kvar flytur vart oppgjeve som prosent med 100 % som oppnåeleg skår. Kriteria vart hovudsakleg delt inn i *Basic Flying*, *Advanced flying* og teorikarakterar. Eg har valgt kriterier ut frå studiar som involverar prestasjonskriterier under jagarflygarutdanning med spesiell vekt på studiar av UPT/SUPT-utdanninga i USA.

Prediktiv validitet vert ofte estimert som korrelasjon mellom prediktor og kriterium. I dei tilfella fleire prediktorar vert nytta kan multipel regresjonsanalyse nyttast for å rekna ut kor mykje varians i kriteriet som kan forklarast ved hjelp av dei aktuelle prediktorane (Ringdal, 2009).

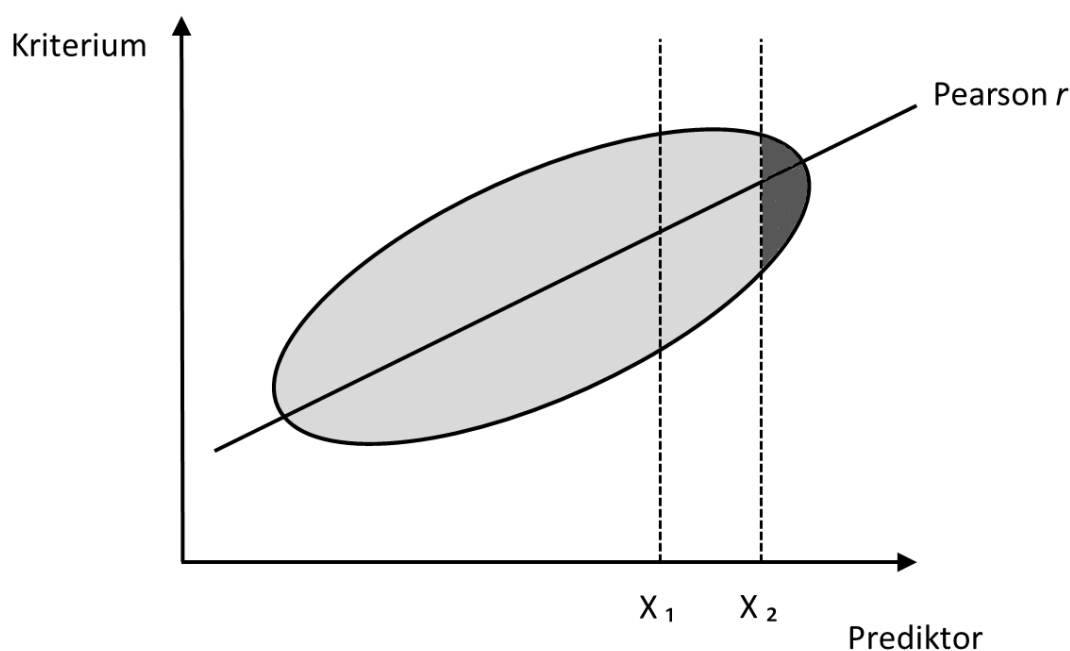
### 2.3 Meta-analysar

For små nasjonar kan det vera utfordrande å gjennomføra valideringsstudiar med tilstrekkeleg statistisk styrke då storleiken på utvalet ofte vil vera avgrensa. Ei mogeleg løysing kan vera å gjennomføra meta-analysar på tvers av allereie utførte studiar. Meta-analysar er ei fellesbetegning på ei kvantitativ oppsummering av funn (effektar) frå tidlegare undersøkjingar (Hunter & Burke, 1995). Meta-analysane bidreg med å rekna ut gjennomsnittet av korrelasjonar over fleire studiar innanfor same tema (til dømes flygarseleksjon). Små valideringsstudiar kan ofte oppnå avgrensa statistisk styrke og stort sprik i korrelasjon kvar gong ein undersøker ei ny lita gruppe. Ei føremon med meta-analyse er tilgang på meir data og større  $N$ . På dette viset kan nøyaktigheita på estimata på den *sanne* korrelasjonen mellom prediktor og kriterie for flygarseleksjon bli betre og ein kan jamvel undersøka i kva grad den prediktive validiteten let seg generalisera over ulike situasjonar (Martinussen, 1997).

### 2.4 Restriction of Range

Eit problem knytta til denne typen undersøkjingar er at jo strengare seleksjonen er, dess mindre vert gruppa det er mogeleg å inkludera. Ei utfordring er nemleg at gruppa som kunne vore ei kontrollgruppe i denne studien vart sendt heim etter kvart som seleksjonsprosessen gjekk framover. Dette fenomenet påverkar den observerte korrelasjonen mellom test og kriterium i resultata knytt til flygarseleksjonsstudiar og dette vert gjerne omtala som *restriction of range* (RR). RR fører til ei underestimering av den sanne prediktive validiteten, då fenomenet med fråfall av personell i høve ein seleksjonprosess avgrensar variansenitetsskåra, og med det kor store observerte korrelasjonar det er mogeleg å oppnå i analysar av seleksjon og seinare prestasjon. Fenomenet kan gjera sitt til at ein kjem fram til feilaktige konklusjonar og ein bør difor stadig vera merksam på RR (Carretta & Ree, 2003).

Utfordringa med RR kan illustrerast med figur 1. Figuren viser samanhengen mellom prøveresultat og kriterieskår både for heile gruppa og for den selekterte gruppa (til høgre i figuren). I figuren representerar  $X_1$  seleksjon til LFS (opptak som flyaspirant) og  $X_2$  seleksjon til jagarflygarutdanning (bestått LFS og selektert til ENJJPT-utdanning).



Figur 1: Restiction of Range

Korrelasjonar utrekna for gruppa lengst til høgre (mørkast felt) vert svakare enn om heile gruppa hadde vorte undersøkt. Variasjonen i denne gruppa vil fylgjeleg vera vesentleg redusert i forhold til om samtlige søkjarar til LFS, uavhengig av testresultat, vart sendt til USA på jagarflygarutdanning.

Dunbar og Linn (1991) har skildra ein metode som ofte vert nytta ved ynskje om å korrigera for RR. Slike korreksjonar kan vera nyttige om ein ynskjer å estimera den faktiske prediktive validiteten og med det nytten av testen, viss ein ynskjer å samanlikna nye og gamle seleksjonsinstrument eller samla inn samanliknbare validitetar i høve gjennomføring av meta-analysar. Desse korreksjonsmetodane har ofte temmeleg strenge føresetnader som ikkje passar heilt med ein faktisk seleksjonssituasjon der ein tek i bruk prediktorar stegvis og i kombinasjon med andre variablar (til dømes intervju). Dette gjer sitt til at korrigering for RR kan vera problematisk.

## 2.5 Forskingsfunn

Forskning i høve flygarseleksjon har vore klart dominert av den militære innsatsen. Det føreligg difor relativt lite forskingsresultat knytt til sivil flygarutdanning (Hunter & Burke, 1995; Carretta & Ree, 2003). Militæret har vore ute etter å identifisera dei gruppene av prediktorar som mest sannsynleg, basert på erfaring, gjev best prediksjon knytt til flygarprestasjon (Carretta, 1992a; Hunter & Burke, 1995).

Generelt har ulike evneprøvar hatt god prediktiv validitet i høve det å predikera arbeidsprestasjonar generelt, medan andre metodar, som til dømes ustruktureerte intervju og personlegheitstestar, har hatt lågare prediktiv validitet (Schmidt & Hunter, 1998). Desse funna var basert på meta-analysar av ei rekkje yrker og jobbkriterier. Likevel har det vist seg at ein kombinasjon av struktureerte intervju eller element frå personlegheitstestar (til dømes mål på integritet) med mål på generelle mentale evner (GMA) kan gje høg prediktiv validitet (frå  $r = .63$  til  $r = .65$ ) (Schmidt & Hunter, 1998).

Forsvaret valgte i høve innføringa av *1980-testbatteriet* ved seleksjon av flygarkandidatar til Luftforsvaret å nytta testar for å måla evner og ferdigheitar innanfor desse områda; generell intelligens, teknisk innsikt og romoppfatning, simultankapasitet og reversering- overføringsevner (Hansen, Heim, Nordby, Storsve, & Værnes, 1989; Hansen, 2006). Forskingsresultat hadde vist at evner innan desse områda hadde relevans for flygarprestasjonar.

Flygarseleksjon i det amerikanske flyvåpenet (United States Air Force - USAF) inkluderte testar som målte verbale evner, teknisk innsikt, *scale reading* (eit mål på oppfatningsevne), instrumenttyding og flykunnskap. Valet deira av seleksjonstestar er òg basert på resultat frå forskning innan flygarseleksjon. Ein finn tilsvarende testsamansetningar i dei fleste land der militær flygarseleksjon føregår (Hunter & Burke, 1995). Utover dette er det vanleg å nytta psykomotoriske testar, personlegheitstestar og intervju i høve militær flygarseleksjon.

## 2.5.1 Evner som har vist prediktiv validitet

### 2.5.1.1 Om generell intelligens (*g*) som prediktor

*Intelligence is the best predictor of job performance*

(Ree & Earles, 1992, s. 86)

Det er gjennomført ei rekkje studiar som har undersøkt den prediktive validiteten til ulike kognitive testar i samband med flygarseleksjon. I dette kapitlet presenterer eg funn knytt til intelligens og flygarprestasjon.

Som sitatet frå Ree og Earles (1992) over her uttrykkjer kunne desse forskarane visa at dess høgare lading ein test hadde på generell intelligens (*g*) dess høgare validitet hadde han som prediktor for flygar *training success* ( $N = 5500$ ). Den prediktive validiteten i høve *g* åleine var høg. Viss ein la til andre testar som målte andre meir spesifikke evner vart den prediktive validiteten forbetra, men berre i liten grad. Dei hevdar vidare at generell intelligens kunne predikera resultatet av opplæring i ny jobb like godt, på generell basis, uavhengig av jobbtype og

kompleksitet i utføringa. For flygarutdanning, der arbeidsprestasjonsmål omfatta flygarprestasjon og teorireultat, var  $g$  den klart beste prediktoren.

Allerie i 1919 gjorde Henmon funn knytt til generell intelligens og flygarprestasjonar. Han fann ein korrelasjon på  $r = .35$  mellom *Thorndike Mental Alertness Test* (eit mål på generell intelligens) og instruktørane sine vurderingar av flygarferdighet ( $N = 300$ ) (Henmon, 1919). Dette funnet tyder på at det kan vera samsvar mellom generell intelligens og flygarprestasjonar (Hunter og Burke, 1995). Ein liknande studie av Signori (1949) fann at av fem kognitive mål/testar var det resultatata frå prøven som målte generelt evnenivå (GMA) som hadde lågast korrelasjon ( $r = .06$ ) med bestått/ikkje bestått flygarutdanning under 2.VK ( $N = 319 - 366$ ). Noko seinare fant Want (1962) ein positiv korrelasjon ( $r = .23$ ) mellom eit mål på numerisk intelligens og utdanningssuksess ( $N = 117$ ), mens korrelasjonen mellom verbal intelligens og utdanningssuksess ikkje var signifikant. Av dei sju testane Want undersøkte var det eit mål på instrumenttyding som hadde høgast korrelasjon med kriteriet bestått/ikkje bestått flygarutdanning.

Då Olea og Ree (1994) samt Carretta og Ree (1994) undersøkte *Air Force Officer Qualifying Test* (AFOQT)-batteriet sin prediktive validitet i høve to flygarprestasjonsmål (bestått/ikkje bestått *Undergraduate Pilot Training* (UPT) og rangering i klassen etter fullført utdanning) var  $g$  – komponenten den beste prediktoren ( $N = 678 - 3900$ ). Det var dei  $g$ -lada papir-og blyanttestane frå AFOQT-testane som skilte seg ut som dei beste prediktorane, men tidlegare flyerfaring, psykomotoriske evner og haldning til risiko bidrog alle tre med trinnvise forbetringar i høve prediksjon av kriteria.

Hunter og Burke (1995) refererer til Ree og hans kolleger i USA og deira syn på at nær alle mål på generelle kognitive evner ladar høgt på  $g$  og at dei med det ser ut til å vera i stand til å predikera utdanningsprestasjonar basert på dette  $g$ -målet. Hunter og Burke hevdar på si side at det på den tida var avgrensa einigheit i litteraturen både om Ree sine påstandar var korrekte, og ein del diskusjon kring kva som eigentleg var meint med  $g$ . Sjølv trekkjer dei fram at empirien har vist tydeleg at mål på generelle kognitive evner, som verbale og numeriske resonneringsevner, kan koma med eit signifikant bidrag i høve prediksjon av flyelevar sine prestasjonar under utdanning og mogelegvis òg vera i stand til å predikera prestasjonar som operativ flygar seinare.

Trankell (1959) undersøkte 363 flygarar i SAS og fann signifikant korrelasjon mellom bestått flygarutsjekk (oppnådd SAS-flygar status) etter å ha gjennomført SAS sitt eige opplæringsprogram og mål på intelligens (verbal og induktiv) og teknisk innsikt ( $N = 363$ ).



I Noreg har Alminnelig evnenivå (AE) vore eit sentralt mål knytt til intelligens. Dette målet består av testar som måler teoretiske kunnskapar innan rekning, ordforståing og resonneringsevne (Hansen, 2006). Sidan alle menn over 18 år vert kalla inn til sesjon har ein god oversikt over AE i den mannlege befolkninga her til lands. Prøvene resulterte i ein samla karakter på ein skala frå 1-9, der 5 er gjennomsnittet for alle vernepliktige. I 2013 lyt kandidatar ha karakteren 5 eller høgare for å søkja på utdanning i Forsvaret og prøvene kan berre gjennomførast ein gong (Forsvaret, 2013). I fylgje Kjenstadbakk (2012) har undersøkingar vist at AE har signifikant prediktiv validitet i forhold til kriteria akademisk og militærfagleg prestasjon. Sundet, Barlaug, og Torjussen (2004) undersøkte AE-målet og fann mellom anna at AE-målet har høg reliabilitet og validitet. Basert på dette kan det vera interessant å undersøkje om AE har prediktiv validitet i høve meir spesifikke prestasjonar relatert til flyging.

#### **2.5.1.2 Om teknisk innsikt og spatiale evner som prediktor**

I 1947 kunne Fiske rapportera funn knytt til flygarseleksjon i US Navy. Testen som målte teknisk innsikt viste seg jamnt over å ha høgare korrelasjonar med eit bestått/ikkje bestått flygarutdanningskriterium enn det prøvar knytt til måling av generell intelligens hadde ( $N = 6000$ ). Melton (1947) kunne òg rapportera om høge korrelasjonar mellom teknisk innsikt og bestått/ikkje bestått flygarutdanning ( $N = 1500$ ) (Hunter & Burke, 1995).

I boka *Handbook of Pilot Selection* (1995) oppgjev Hunter og Burke at evner som er spesielt knytt til flyging er spatiale evner, det å kunna lesa tabellar og ymse instrument samt evne til å lesa av, tolka og forstå spesifikke flyinstrument. Dei samla resultatata knytt til flygarseleksjonsforskning viser at nokre mål på kognitive evner er betre enn andre i høve prediksjon av flygarprestasjon. Teknisk innsikt har over lang tid vist seg å kunna predikera utdanningsresultat betre enn mål knytt til verbale- og numeriske resonneringsevner. Det same gjeld i høve mål for instrumenttyding og flykunnskap. Av den grunn bør mål på slike evner vera ein del av eit effektivt flygarseleksjonssystem (Hunter & Burke, 1995).

I 1987 undersøkte Carretta spatiale evner som prediktor i høve flygarutdanning i United States Air Force (USAF). Han kunne rapportera at oppnådde resultat på ein spatial-evne-test ikkje samsvarte med bestått/ikkje bestått UPT flygarutdanning ( $N = 1939$ ). Likevel korrelerte denne prediktoren signifikant med kva flytype den einskilde nye flygar vart tildelt etter gjennomført flygarutdanning. Dette betydde at dess betre spatiale evner ein kandidat hadde dess større sjanse var det for å få opplæring på jagarfly eller jagar-bombefly framfor meir saktegåande fly som transport- eller tankfly etter fullført utdanning (Carretta, 1987).

I Martinussen si avhandling frå 1997 vart Forsvaret sin bruk av slike testar i høve flygarseleksjon undersøkt. Ei mindre meta-analyse ( $N = 244 - 977$ ) knytt til validering av det nasjonale testbatteriet utført av Martinussen og Torjussen (1998) viste at det for Luftforsvaret var tre gode prediktorar som var knytt til flygarprestasjon (bestått/ikkje bestått LFS inklusiv teorikarakter i fire fag): instrumenttyding, mekaniske prinsipp og flykunnskap (gjennomsnittleg  $r = .29$ ,  $r = .23$  og  $r = .22$ , høvesvis). Spasiale evner og generell intelligens korrelerte òg positivt men svakare (mellom  $r = .09$  og  $r = .14$ ). Dette viste seg å vera heilt i tråd med tidlegare funn i større meta-analysar om flygarseleksjon (Hunter & Burke, 1994; Martinussen, 1997).

### 2.5.1.3 Om simultankapasitet og psykomotorisk koordineringsevne som prediktor

Då Trankell (1959) undersøkte pilotsleksjonsystemet i SAS kunne han rapportera at eit mål på simultankapasitet, målt ved hjelp av ein tappe-test, var det målet, blant 14 vurderingsområder, som korrelerte høgast ( $r = .42$ ) med kriteriet bestått/ikkje bestått SAS-pilotutsjekk ( $N = 363$ ). Denne testen vert framleis nytta i høve militær flygarseleksjon i Noreg.

I 1989 fann Carretta at individuelle skilnader i testresultat knytt til psykomotoriske og kognitive evner var relatert til flygarprestasjonar under utdanning. Dei kandidatane som hadde betre psykomotoriske koordineringsevner, raskare reaksjonstid og betre og meir effektiv hukommelse hadde høgare sannsynlegheit for å gjennomføra flygarutdanninga ( $N = 478$ ). Ved å inkludera slike mål kunne ein redusera usikkerheit knytt til utvelging av kandidatar (Carretta, 1989). Nokre år seinare (i 1992) førte dette til at testar for å måla psykomotorisk koordineringsevne vart gjeninnført i USAF som ein del av vurderingsgrunnlaget i flygarseleksjonsprosessen (Hunter & Burke, 1995). Blower og Dolgin (1990) fann òg at kompleks psykomotorisk koordinering samt visuelle informasjons-prosesseringsevner var gode prediktorar i høve suksessfull gjennomføring av flygarutdanning.

Hunter og Burke (1994) og Burke (1993) gjennomførte omfattande meta-analysar ( $N = 2792 - 52153$ ) for å undersøka ulike prediktorar for flygarprestasjon. Dei fann at dei beste prediktorane var 1) *job sample* (til dømes i flysimulator) ( $r = .34$ ), 2) *Gross Dexterity* (eit psykomotorisk koordinasjonsmål på grovmotorikk ( $r = .32$ ), 3) reaksjonstid ( $r = .28$ ) og evna til å løysa tekniske problem ( $r = .29$ ). Eit eksempel på eit flygar *job sample* kan vera å la kvar kandidat gjennomgå ein flysimulatorprøve for å måla og avdekka flydugleik.

Griffin og Koonce (1996) rapporterte om to testar som ofte har blitt nytta til å måla psykomotoriske evner: *Complex Coordination test* og *Two-hand Coordination test*. Begge desse testane var signifikante prediktorar i høve bestått/ikkje bestått flygarutdanning, både for Air Force *pilots* og Navy *aviators*. I tillegg såg det ut til at oppdaterte og automatiserte versjonar av

denne type testar var like gode prediktorar som dei manuelle testane utvikla før 2.VK hadde vore ( $N = 3146 - 4779$ ).

Testar som måler psykomotorisk koordineringsevne har vist seg over tid å vera ein valid metode for å velja ut flygarkandidatar, både militært og sivilt, og har såleis vorte akseptert som ein heilt naudsynt del av ein flygar sitt evnesett (Hunter & Burke, 1995). Andre forskingsresultat støttar òg opp om at ein bør gjera nytte av mål på *multi-tasking* i høve seleksjon og prediksjon av flygarprestasjonar (Damos, 1993).

Martinussen si avhandling frå 1997 omhandlar òg ei større og meir «global» meta-analyse gjennomført i høve flygarseleksjon og bruk av psykologiske mål som prediktorar for flygarprestasjon ( $N = 3736 - 17900$ ). Forskaren fann her at tidlegare flyerfaring ( $r = .30$ ) og ein kombinert index som bestod av kognitive og/eller psykomotoriske testar ( $r = .37$ ) var dei beste prediktorane for flygarprestasjon (bestått/ikkje bestått). Dei nest beste prediktorane var testar som målte kognitive<sup>6</sup> ( $r = .24$ ) og psykomotoriske-/informasjonsprosesseringssevner<sup>7</sup> ( $r = .24$ ) samt flykunnskap ( $r = .24$ ). Personlegheit, generell/global intelligens<sup>8</sup> og skulekarakterar/matematikk- og engelsktestar oppnådde dei lågaste validitetane/korrelasjonane ( $r = .14$ ,  $r = .16$  og  $r = .15$ , høvesvis).

#### **2.5.1.4 Prediktorar knytt til militær flygarutdanning i USA spesielt (herunder UPT/SUPT og ENJJPT)**

Carretta (1992b) hevda at ein kunne forstå forholdet mellom seleksjonsfaktorar og flygarprestasjonar under utdanninga betre viss ein gjekk bort frå det todelte prestasjonskriteriet bestått/ikkje bestått utdanning. Prestasjonskriterier burde, etter hans syn, innehalda så mykje data som mogleg for å oppnå meir stabile prestasjonsindikatorar. Han kunne visa at nokre utdanningsresultat (frå UPT) påverka den vidare flygaropplæringa meir enn andre. I all hovudsak var desse kriteria teoriresultat (*Academic Grades*), resultat på daglege flyturar (*Daily Flying Grades*) og resultat oppnådd på eksamensflyturar (*Check Flight Grades*). Viss ein nytta slike kriterier kunne den relative prestasjonen til alle flygarkandidatane baserast på ein overordna prestasjon framfor spesifikke manøvrar i lufta eller einskilde eksamensresultat. Slike detaljrike kriterier kunne vera meir nyttig for forskinga på flygarseleksjon enn det dei todelte kriteria hadde vore. Carretta kunne òg rapportera at flygarprestasjonar basert på dei nemnte kriteria var sterkt

<sup>6</sup> *Cognitive tests included various paper-and pencil tests measuring mechanical comprehension, spatial orientation, time sharing, visualization, perceptual speed, instrument comprehension, and attention (Martinussen, 1997).*

<sup>7</sup> *Psychomotor/infoprocessing tests using PC or apparatus to measure complex coordination (Martinussen, 1997).*

<sup>8</sup> *Intelligence tests could have been included in the general heading of cognitive tests; however, these tests were placed in a separate category because they are intended to measure global intelligence rather than specific abilities (Martinussen, 1997).*

relatert til kva flytype den einstilte flyelev fekk etter gjennomført utdanning. Det har vist seg at dei amerikanske elevane som presterte best på flygarutdanninga fekk tilbod om å fly jagarfly, medan kandidatar som presterte dårlegast vanlegvis fekk tilbod om å fly bombe-, transport- eller tankfly ( $N = 755$ ) (Caretta, 1992b).

I 1995 gjennomførte Carretta og Ree ei omfattande undersøkjing av det amerikanske AFOQT-batteriet sin validitet for å kunna predikera multiple flygarprestasjonar under UPT-utdanninga i USAF ( $N = 7563$ ). Daglege flykarakterar (*Daily flying grades*), karakterar frå eksamensflyging (*Check flight grades*) og teoriprestasjon frå både T-37 og T-38 vart nytta som kriterier. Testen som målte flykunnskap og instrumenttydingstestane var dei beste prediktorane i høve T-37 *Daily flying* og *Check grades* ( $r = .25 / r = .19$  og  $r = .18 / r = .19$  høvesvis). Testen som målte evne til *Scale Reading* (eit mål på oppfatningsevne) var den mest prediktive testen for T-38 *Daily flying* og *Check flight* ( $r = .08$  og  $r = .13$ ) og den viste samstundes god validitet for flyging med T-37 ( $r = .16$  og  $r = .17$ ). Validitetane som her er oppgjevne er ikkje korrigert for RR. Testen *Arithmetic Reasoning*, som var eit godt mål på generell intelligens, var den beste prediktoren for gjennomsnittleg resultat på dei skiftlege eksamenane i flyteori (*Academics*).

I ein nyare studie undersøkte Carretta (2011) prediktiv validitet for ulike prediktorar i forhold til prestasjonane til amerikanske flyelevar under flygarutdanning i USA ( $N = 883$ ). Flygarutdanninga føregjekk på Specialized Undergraduate Pilot Training (SUPT), tilsvarande SUPT ved ENJJPT, der kriteriet var ulike mål på T-6 prestasjonar (det vil seie *Basic flying*, fasen før T-38 *Advanced Flying*). Carretta nytta igjen *Academics* (teori), *Daily Flying Grades* (dagleg flyging) og *Checkride Grades* (eksamensflyging) og rangering i klassen i tillegg til bestått/ikkje bestått utdanning som kriteria. Han kunne rapportera at seleksjonstestane som var nytta viste like god prediktiv validitet for kriteria, trass i mindre justeringar over tid for testane og sjølve flyprogrammet etter innføringa av *Pilot Candidate Selection Method* (PCSM) -systemet i 1993. Carretta meinte at den vedvarande prediktive validiteten skuldast faktorar som over tid har vist samanheng med flygarprestasjon under utdanninga, som kognitive evner, kunnskap om jobben, psykomotoriske evner og tidlegare flyerfaring.

I 2000 undersøkte Carretta mål som inkluderte andre område enn evnetestar og fann at ein organisasjon ikkje bør leggja vekt på *Officership*<sup>9</sup> som seleksjonskriterium. Dei to store flygarrekrutteringsorganisasjonane *Air Force Academy* (AFA) og *Reserve Officer Training Corps* (ROTC) hadde i stor grad lagt *Officership* til grunn for seleksjon. Det viste seg at dette

<sup>9</sup> Dette er eit samleomgrep som inkluderar militær framferd, initiativ, leiareigenskapar med meir.

målet ikkje kunne bidra med prediktiv validitet i høve framtidige flygarprestasjonar slik som evnetestar kunne det.

### **3 Flygarseleksjon i Noreg og jagarflygarutdanninga i USA**

I dette kapitlet skal eg presentera flygarseleksjonssystemet i Forsvaret med vekt på den praktiske gjennomføringa og innhaldet slik seleksjonen vart utført på 1980- og 90-talet og fram til rundt år 2000.

#### **3.1 Generelt om gjennomføring av seleksjonen**

Opptaksprøvene til LFS vart gjennomført som gruppetestar. Berre Spegeltesten, DMT (*Defence Mechanism Test*) og intervjuet vart gjennomført individuelt. Opptaksrådet/uttakskommisjonen hadde fullmakt til å ta opp flyaspirantar til LFS. Slike vedtak vart fatta etter at kandidatane hadde bestått opptaksprøvene og gjennomført intervjuet. I tillegg til dei psykologiske undersøkingane måtte kandidatane gjennomgå fysiske testar og medisinske undersøkingar.

#### **3.2 Litt bakgrunn i høve dei norske flygartestane**

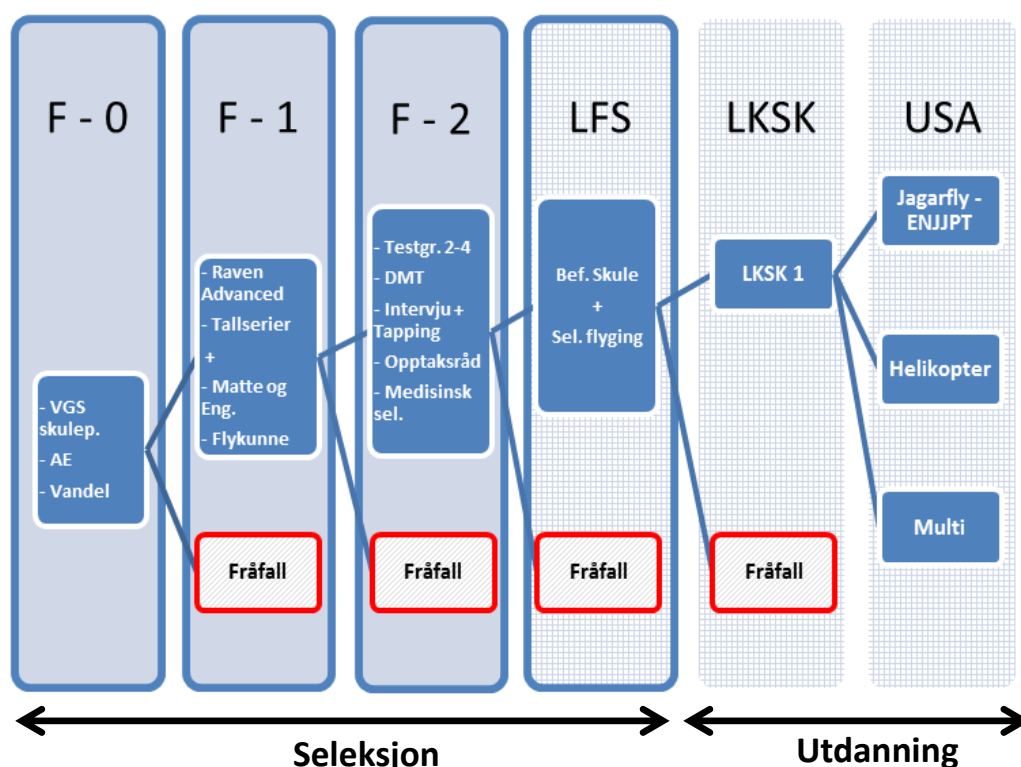
Etter 2. VK hadde Forsvaret eit testbatteri som vart utvikla i 1949 og brukt fram til 1960 då Syversen analyserte batteriet sin prediktive verdi (Hansen, 1987). Syversen (1960) kom fram til to seleksjonsfaktorar. Den eine var teoretisk prega (T-sum for teoretiske evner) og den andre hadde med praktisk flyging å gjera (P-sum for praktiske evner) (Syversen, 1960). Konklusjonen var mellom anna at ein med strengare seleksjon kunne forbetra kvaliteten på flygaren. Dei neste 10 åra vart det einskilde prøveresultatet, alminneleg sumskår, T- og P-sum og realskulekarakteren nytta som det viktigaste grunnlaget for opptaksvurderinga (Syversen, 1960; Hansen, 1987). Forsvarets Psykologitenester byrja etter kvart å utvikla enkle psykomotoriske prøvar (Hansen, 1987). I midten av 1970-åra hadde FPT bygd opp eit testbatteri som inneheldt både psykomotoriske testar og teoriprøvar for flygaropptak. Storsve analyserte batteriet for 1974- og 1975- flykulla. Dette førte til ei vidare utbygging av testbatteriet slik at ein på sumaropptaket i 1978 fekk prøvd ut eit stort og nytt testbatteri (Hansen, 1987). Saman med vidare utprøvingar leia dette fram til utforming og innføring av *1980-batteriet* (Hansen et al., 1989). Det er dette testbatteriet som ligg til grunn for utvelging av flygarane som er undersøkt i denne studien.

Innføring av databaserte testar som erstatning for delar av *1980-batteriet* vart starta opp i 1996 (samtale med sjefpsykologen i Forsvaret, 6. des 2012). Fram til og med 1998 vart *1980-batteriet* og datatestane køyrt parallelt. Vedtak om opptak til LFS vart alltid gjort på grunnlag av

dei gamle papir-og blyant- og apparattestane åleine før dei databaserte testane overtok for nokre av prøvane frå 1999 (Torjussen & Hansen, 1999).

### 3.3 Seleksjonsprosessen – ein modell

Seleksjon av flygarar i Luftforsvaret gjekk gjennom fleire fasar. Prosessen vert illustrert med fylgjande modell:



Figur 2: Forsvaret sin flygarseleksjonsmodell

### 3.4 Seleksjonsprosessen i praksis

Prosessen startar ofte med eit stort antal søkjarar til LFS. Fase 0 inneber ei utvelging basert på data som allereie føreligg frå sesjon, altså før flygaropptaksprøvene startar opp. Fase 1 inkluderar testar som er lite ressurskrevjande. Då det føregår eit fråfall av kandidatar undervegs vert det lagt meir ressursar til seleksjonprosessen etter kvar fase.

På grunnlag av dei faktoranalysane som vart gjennomført vart dei teljande prøvane frå 1980-batteriet delt inn i ulike område (Hansen et al., 1989):

Testgruppe 1) Generelt evnenivå (Fase 1-prøvar)

Testgruppe 2) Teknisk innsikt og romopplevelse (Fase 1-prøvar)

Testgruppe 3) Simultankapasitet (Fase 2-prøvar)

Testgruppe 4) Overføring/retningssevne (Fase 2-prøvar)

Dei kandidatane som oppnådde tilfredsstillande snittresultat innanfor kvart område vart sendt vidare til det siste prøveområdet. Det bestod av DMT og psykologintervju (gjennomført mot slutten av Fase 2). I løpet av intervjuet måtte kandidaten i tillegg gjennomgå ein tappetest for å måla simultankapasitet (Hansen et al., 1989).

#### 3.4.1 Fase 0

Skulepoeng og karakterar frå vidaregåande skule (VGS) saman med AE-resultat frå sesjon vart nytta for å velja ut kandidatane som skulle kallast inn til fase 1 av opptaksprøvene (Martinussen, 1989). I tillegg måtte ein del formelle krav innfriast. Søkjaren måtte vera mellom 18 og 24 år, ha god helse og godt syn, vera mellom 163 – 193 cm og inneha generell studiekomptanse frå VGS, samt god vandel (Martinussen, 1989). Det vart ikkje stilt spesielle krav til skulekarakterar (Torjussen & Hansen, 1999). Krav til 5 eller betre på AE for å koma inn på befalsskule i Forsvaret kom seinare medan general S. Diesen var Forsvarssjef (2005-2009).

#### 3.4.2 Fase 1

I LFS-opptaket sin fyrste fase (øg omtalt som *Regionale opptak* på 80- og 90-talet) ynskte Luftforsvaret å testa kandidatane sin generelle intelligens, tekniske innsikt, spatiale evner/romopplevelse, flykunnskap, matematiske kunnskap og engelskforståing (Hansen & Torjussen, 1999). Resultatet oppnådd på desse prøvane var teljande for å koma vidare til fase 2 (øg omtalt som *Sentrale opptak*). Det var satt visse *cut-off* grenser. Dei som skåra under dette nivået gjekk normalt ikkje vidare til neste fase. Resultat frå fase 1 testane danna to av fire testgrupper; testgruppe 1 – Generelt evenenivå og testgruppe 2 – Teknisk innsikt/romopplevelse. Testgruppe 1 og 2 på *1980-batteriet* inneheldt testane Raven Advanced Matrices, 43 KF Talseriar, 370 Mekaniske oppgåver, U3B-54 Plateforming, 391 Figurmønster, 402 Figurformer og romoppfattingsdelen av prøve 344. Testane er omtalt nærare i metodedelen (kapittel 4).

#### 3.4.3 Fase 2

Testgruppe 3 og 4 på *1980-batteriet* vart administrert i løpet av eit to veker langt sentralt opptak der både kandidatar, psykologar og anna personell involvert i flygarseleksjon var samla. Fase 2 inneheldt fylgjande enkeltprøvar: 350 Sorteringsprøven, 310 Retningsreversaller, 410 Tallsøker,

300 Instrumenttyding, 330 Romoppfatning, 360 Spegltesten og tidsvurderingsdelen av prøve 344. Testane er omtalt nærare i metodedelen (kapittel 4).

#### **3.4.3.1 Fase 2 Opptaksråd og medisinsk undersøking**

Kandidatar som hadde bestått testane i testgruppe 1 – 4, samt DMT og psykologintervju med tappetest, vart kalla inn til opptaksråd/uttakskommisjon. Dette rådet stod fritt til å intervju kandidaten vidare etter behov og koma med avklarande spørsmål. Samansetninga i rådet var slik at det var eit formelt og vedtaksdyktig organ (Torjussen & Hansen, 1999). For å koma til opptaksråd/uttakskommisjon burde kandidaten skåra over eit visst nivå. Det tyder ikkje at den einskilde nødvendigvis måtte oppnå høgast mogleg skår, men heller skåra bra på alle evneområda testane var meint å dekkja (Torjussen & Hansen, 1999). Her vart den einskilde kandidaten anten akseptert som aspirant til LFS eller dimmitert. Dei som vart anbefalt tekne opp som aspirantar måtte i tillegg gjennomgå ei grundig medisinsk undersøkjing ved Flymedisink Institutt i etterkant for å få høve til å møta opp på skulen (Martinussen, 1989).

#### **3.4.4 Fase 3 - Luftforsvarets Flygeskole**

LFS hadde både eit befalsskule- og eit flygarseleksjonsføremål. Kandidatar som gjennomførte LFS mottok vitnemål for gjennomført befalsskuleutdanning på same vis som andre befalsskular i Forsvaret. Nokre flyaspirantar hadde gjennomført befalsskule tidlegare. Dei møtte direkte til sjølve seleksjonsflygingsperioden ved LFS.

Ein sentral funksjon for LFS var seleksjonsflyginga. Flyaspirantane måtte bestå eit flyprogram med propellflyet Saab Safari for å få tilbod om flygarutdanning i USA. Reint statistisk har det vist seg at om lag 50 % av kandidatane som vert akseptert som flyaspirantar faktisk består seleksjonsflyginga (samtale med sjef Seleksjonscenteret, jan 2013). For å bestå måtte aspirantane mellom anna visa at dei overkom strenge krav til progresjon i høve flyging. Flyprogrammet var kort og intensivt og det vart såleis lagt opp til ei bratt læringskurve.

#### **3.4.5 Luftkrigsskule I**

Etter fullført LFS gjennomførte flyelevane Luftkrigsskule I, del 1, eventuelt Luftkrigsskule I kvalifiseringskurs på ordinært vis. Denne utdanninga var både praktisk og akademisk retta og var ikkje knytt til flyging. Militære fag knytt til leing, luftmakt, strategi og militærhistorie, samt språk og realfag var sentrale. Etter bestått Luftkrigsskule gjekk ferda vidare til USA og flygarutdanning. For jagarflygarelevane gjekk reisa til ENJJPT i Texas lokalisert ved 80th Flying Training Wing, Sheppard Air Force Base.



### 3.4.6 Euro-NATO Joint Jet Pilot Training

#### 3.4.6.1 Oppdraget

Oppdraget til *80th Flying Training Wing* var å gjennomføra *Undergraduate Pilot Training* (UPT) (U. S. Air Force, 2008). UPT, omtala som *Specialized UPT* (SUPT) frå midt på 90-talet (Carretta, 2000), er *Air Education and Training Command* (AETC) sitt omgrep på flygarutdanninga dei utfører for United States Air Force. UPT/SUPT er eit militært flygarutdanningsprogram der opplæringa består av ein serie progressive fasar, der kvar ny fase vert suksessivt meir kompleks og krevjande enn den førre fasen (R. C. Air Force, 2011).

13 nasjonar<sup>10</sup> gjekk i 1981 saman om å etablere ei felles flygarutdanning (U. S. Air Force, 2013a). Over ein 30-års periode har skulen utdanna over 6400 flygarar for NATO. Programmet gjekk gjennom tre fasar: fase I - teori, fase II – flyging med T-37 og fase III – flyging med T-38 (Carretta, 1992b). Seinare har T-6 erstatta T-37 i fase II. Flyskulen har eit høgt aktivitetsnivå. Kvart år vert det flydd meir enn 55000 flyoppdrag for å utdanna omlag 200 nye flygarar (U. S. Air Force, 2013b). Utdanninga tek 55 veker (U. S. Air Force, 2001) og det offisielle språket er engelsk (U. S. Air Force, 2008).

#### 3.4.6.2 Generelt om flyturane under utdanninga

AETC Instruction 36-2205 gav dei overordna retningslinjene i høve flyturane elevane skulle fly. (U. S. Air Force, 2003). Detaljinnhaldet i kvar flytur og kva som skulle øvast på var regulert av ENJJPT *Syllabus* (U. S. Air Force, 2008). Elevane måtte gjennomføra flyturane *syllabus* la opp til og alle trenings-*events* måtte føregå i rett rekkjefylgje. *Flight Commander*<sup>11</sup> hadde ansvaret med å tildela elevane ein fast flyinstruktør (*Instructor pilot – IP*) og samstundes sørge for at den einskilde flyelev flaug med eit avgrensa antal instruktørar totalt sett gjennom kvar fase. *Flight Commander* hadde òg ansvaret med å sikra at flyelevane som tilhørte hans/hennar flyavdeling (*flight*) flaug med jamn frekvens slik at ein unngjekk uheldige opphald i treninga (U. S. Air Force, 2003).

#### 3.4.6.3 Phase I - Academics

I fase I av utdanninga vart det undervist i flyteori og -prosedyrar for T-37 og T-38 (Carretta, 1992b). Teoriundervisninga dekkja mellom anna fag som *Aerospace Physiology*, *Aircraft Systems*, *Basic Instrument*, *Mission Planning*, *Navigation* og *Weather* (Carretta, 2011). Prestasjonen på teorieksamen i dei ulike faga vart målt i prosent, der 85 % eller betre var bestått

<sup>10</sup> Belgia, Canada, Danmark, Tyskland, Hellas, Italia, Nederland, Noreg, Portugal, Spania, Tyrkia, Storbritannia og USA.

<sup>11</sup> Instruktør i ein leiarfunksjon med ansvar for dagleg drift og oppfølging av ein *flight* med eit knippe elevar.

(U. S. Air Force, 2003). Det fyrste forsøket på kvar teorieksamen vart ståande som teljande resultat (U. S. Air Force, 2010a).

#### 3.4.6.4 Phase II - T-37

I fase II vart flyelevane utdanna på T-37. Dei fekk 123 flytimar (U. S. Air Force, 2001) med instruksjon innanfor flydisiplinane *Contact, Instrument, Low-level Navigation and Formation* (U. S. Air Force, 2009). Elevane fekk fly åleine (solo) i disiplinane kontakt- og formasjonsflyging. Opplæringa innanfor kvar flydisiplin vart lagt opp med aukande vanskegrad og det var krav om jamn progresjon gjennom heile fase II (U. S. Air Force, 2002).

Prestasjonsnivået den einskilde elev måtte tilfredsstilla i løpet av fase II var nedskrive på ein detaljert måte i ENJJPT *Syllabus*. Desse standardane vart omtala som *Course Training Standard* (CTS) (U. S. Air Force, 2002) og vart oppdatert og godkjent av ENJJPT *Steering Committee* (styret som representerar deltakarnasjonane) med jamne mellomrom (U. S. Air Force, 2008). Siste flytur innanfor kvar flydisiplin var ein praktisk eksamensflytur (*Checkride*).

Eksamensflyturane (*checkrides*) skulle sikra at flyelevane tilfredsstilte CTS i høve flydugleik innan kvar flydisiplin slik det var nedfelt i *syllabus* (U. S. Air Force, 2002; U. S. Air Force, 2010b). Sertifiserte evaluatorar (Checkpilotar) vurderte utføringa av den einskilde manøver opp mot spesifikke kriterium (U. S. Air Force, 2010b). Fyrste forsøk på kvar manøver, eller flygeelement, vart ståande som teljande i høve endeleg resultat sjølv om eleven i høve kontinuasjonseksamen gjorde det betre eller dårlegare på den aktuelle manøveren. På eksamensflyging skulle ikkje eleven motta instruksjon. Det skulle vera ein rein evalueringstur. Checkpiloten kunne berre bidra med assistanse i særskilde tilfelle der nye og uventa situasjonar oppstod i lufta, altså situasjonar ein normalt ikkje ville forventa at flyeleven kunna meistra på eigenhand (U. S. Air Force, 2002).

I tilfelle stryk på ein *checkride* ville flyeleven få høve til eit nytt forsøk. Det vart kalla *Progress Checkride* (*Initial Progress Check - IPC*). IPC Checkpiloten skulle på ein eksamenstur nr. 2 framleis evaluera, men i større grad enn på ein ordinær *checkride*, vurdere kandidaten sitt potensiale til å fullføra utdanninga (U. S. Air Force, 2000). Resulterte ein IPC òg i stryk fekk eleven ein siste sjanse, kalla *Elimination Checkride* (omtala som *Final Progress Check - FPC*). I høve FPC måtte skvadronssjef eller NK fly med kandidaten for å vurdere om han/ho kunne motta og respondera tilfredsstillande på flyinstruksjon, om det var potensiale for å gjennomføra gjenverande utdanning på T-37 eller T-38, samt vurdere det overordna potensialet til å lukkast med vidare opplæring på andre flytypar etter eventuell fullføring av UPT/SUPT (U. S. Air Force,

2000). Kvar bestått flyeksamen resulterte i vidare opplæring i neste flydisiplin, og etter halvgått utdanningsperiode, flyging med T-38 *fast jet*.

#### 3.4.6.5 Phase III - T-38

Etter fullført fase II vart flyelevane flytta over til T-38 flyskvadronen. T-38 er ein *Fighter-trainer*. Det var spesielt denne flytypen som gjorde det mogleg for elevane å utvikla flygarferdigheitane vidare mot å meistra flyging med raske fly (U. S. Air Force, 2009). I byrjinga av undersøkingsperioden fekk elevane 137 flytimar på T-38. Etterkvart vart dette timetalet noko redusert. I 1993 fekk elevane 133,7 timar og i 2000 130 timar i lufta med T-38 (Manning, 2007). Elevane fekk opplæring innanfor flydisiplinane *Contact*, *Instrument*, *2-ship Low-level Navigation* (to fly), *Basic Formation* (to fly) og *Advanced Formation* (fire fly) (U. S. Air Force, 2002). Dei fekk òg fly åleine (solo) i disiplinane kontakt- og formasjonsflyging med ein instruktør i det andre flyet. For T-38 var det nedskrive eigne CTS som det vart målt mot i høve flyeksamen. Siste flytur innanfor kvar flydisiplin vart òg her avslutta med ein praktisk eksamensflytur. Forholda rundt eksamensflyging i fase III var lagt opp på same måte som for fase II.

I høve rangering innad i kvar klasse vart resultatata frå T-38 flyeksamen i *Contact*, *Instrument*, *Basic Formation* (to fly) og *Advanced Formation* (fire fly) rekna med. *2-ship Low-level Navigation* (to fly) var ikkje ein teljande eksamen (U. S. Air Force, 2000). Elevar som vart rangert blant dei 10 % beste i kvar klasse var tenkt å få tilbod om sitt fyrsteval av flytype (til dømes: F-16, F-15, F-111, B-1, B-2, T-37/T-38 instruktør (FAIP<sup>12</sup>)) (U. S. Air Force, 2000; U. S. Air Force, 2003). Denne rangeringa var kanskje mest aktuell for amerikanske elevar som faktisk hadde fleire flytypar å velgja mellom. For norske elevar var dette valet i utgangspunktet avklart, uavhengig av utdanningsresultat og rangering i klassen, då F-16 var hovudvåpensystemet i Luftforsvaret og såleis det einaste reelle alternativet.

#### 3.4.6.6 Standardisering og kvalitetssikring

I arbeidet med å standardisera og kvalitetssikra ENJJPT-utdanninga la ein opp til eit omfattande regulerings- og kontrolleringsregime. I tillegg til CTS med tilhøyrande *syllabus* som viste kva flytrenoing den einskilde elev skulle gjennomgå var det i høve ENJJPT *Plan of Operations* òg lagt vekt på både overordna og meir detaljerte planar for flyinstruksjon samt retningsliner for korleis instruksjonen skulle føregå. Det var og skrive ned kor mange flytimar ein gjennomsnitts ENJJPT-elev skulle ha til disposisjon for å oppnå tilfredsstillande flydugleik innanfor kvar

flydisiplin og -fase (U. S. Air Force, 2008). I høve standardiseringsarbeidet vart det utarbeidd standard leksjonsplanar, studieguidar, elevarbeidsbøker, standardiserte instruktør-guidar og eit omfattande eksamensregime, òg ut over *checkrides*. Denne innsatsen var meint å bidra til kvalitetssikring av utdanninga (U. S. Air Force, 2008).

I tillegg til standardisering knytt til avvikling av eksamen, både i teori (*Academics*) og flyging (*Checkrides*) vart òg flyinstruktørane kvalitetssikra. For å verta sertifisert som instruktør ved ENJJPT for T-37 eller T-38 måtte instruktørkandidatar gjennomgå instruktørutdanning (Pilot Instructor Training – PIT). Skulen hadde eigne avdelingar som stod for dette. Det var òg ei eige standardiserings- og evalueringsavdeling (*Stan/Eval*), adskilt frå flyskvadronane, der flygarar (*flight examiners*) vart opplært og peika ut til å evaluera særskilte instruktør-*checkrides*. Denne ordninga var meint å bidra til kvalitetsikring og standardisering av instruktørstanden før dei fekk høve til å fly med og instruera elevar. I tillegg vart nyutdanna instruktørar monitorert ekstra tett den fyrste tida dei var i virke. Flygar- og instruktørrettigheitane til den einskilde instruktør måtte fornyast av *Stan/Eval*-avdelinga kvar 18. mnd (instruktøreksamen knytt til teori, instrumentflyging, instruksjonsflyging og handtering av nødprosedyrar).

Ved ENJJPT 89th Flying Training Squadron var skvadronsleiinga oppteken av at nasjonar som hadde elevar i programmet var representert i Checkpilotavdelinga der elevane sine flygarprestasjonar vart evaluert. Standardisering i høve opplæring og utøving av rolla som Checkpilot for elevane vart forsøkt oppnådd gjennom interne prosessar på flyskvadronane i perioden eg var nestkommanderande. Det viktigaste kriteriet var personelege eigenskapar knytt til det å vera ein objektiv og rettvis Checkpilot. Ut over dette hadde alle sertifiserte Checkpilotar pliktig oppmøte på kvartalsvise *Check-pilot meetings*. Føremålet med desse samlingane var å sikra at karaktersettingsfilosofi, praktisk gjennomføring av *checkrides* og vedtekne standardar var konsistente over tid (U. S. Air Force, 2010b).

I tillegg til at instruktørar og Checkpilotar var underlagt eit standardiseringsregime som her er nemnd, måtte òg alle flyelevane ved ENJJPT fylgja eit eige *Student Standardization Program*. Programmet la opp til eit eige vekebasert briefing- og testregime. Den delen av utdanninga inkluderte orienteringar og ulike testar på dagleg- og vekebasis knytt til flyteori og ulike flyprosedyrar. Innsatsen hadde som mål å sikra eit høgast mogeleg kunnskapsnivå, både i høve teori og praktisk bruk av flyprosedyrar (U. S. Air Force, 2003).

---

<sup>12</sup> First Assigned Instructor Pilot (FAIP). Nyutdanna flygar som vart verande att som instruktør etter fullført utdanning.

I høve standardisering av ENJJPT og programmet totalt sett var det lagt opp til rutinemessige inspeksjonar frå UPT/SUPT- inspeksjonslag (*AETC/HQ Inspection teams*) utsendt frå hovudkvarteret til Air Education and Training Command.

#### 3.4.6.7 Graduation

Etter at klassen hadde fullførte T-38-fasen mottok den einsskilde eleven *flygarvingen* som synleg bevis på oppnådd flygarstatus (U. S. Air Force, 2008). I høve denne seremonien vart det delt ut ulike bestemannspremiar. *The Flying Award* og *Daedalian Award* var dei to prisane som var nærast knytt til flydugleik. Fyrstnemnde gjekk til den flyeleven som oppnådde høgast karaktarsnitt på *Checkrides* og *Daily Flying grades* gjennom utdanninga. Sistnemnde gjekk til den flyeleven som oppnådde høgast snitt i flydisiplinen *Formation*, der både *Checkride grades* og *Daily Flying grades* telte med (U. S. Air Force, 2008). Det var òg premie til beste elev i flyteori (*Academics*) (U. S. Air Force, 2008).

#### 3.4.6.8 Rangering og tildeling av flytype etter fullført utdanning

Dei ulike ENJJPT deltakarlanda avgjorde sjølve kva flytype den einsskilde nye flygar skulle fly etter fullført utdanning. Då amerikanarane var *Host Nation* for programmet og flyginga føregjekk i USA var mesteparten av reglane og retningslinjene amerikanske. For å betre forstå kva kriterier og kva vekting som låg til grunn ynskjer eg å trekkja fram det amerikanske MASS-rangeringssystemet (*Merit Assignment Selection System*) som vart nytta (U. S. Air Force, 2000). Amerikanske elevar som graduerte som topp 10 % i sin klasse skulle normalt få sitt fyrste-val av flytype. Ved ENJJPT måtte elevane ha fullført T-38 *contact, instrument, basic formation* (2 fly) og *advanced formation* (4 fly) for å kunna vera med å konkurrera om flytype (F-16, F-15, F-111, A-10, B-1, osv) i sin klasse (U. S. Air Force, 2000). I praksis innebar det at resultat oppnådd med T-38 i *Navigation* – flydisiplinen ikkje påverka rangeringa.

MASS-utrekninga tok omsyn til ei heil rekkje faktorar knytt til utdanningsresultatet. Rangeringa var meint å gje eit heilskapleg bilete av flyeleven sin dugleik basert på dokumenterte prestasjonar. I tillegg skulle *Flight Commander* (FLT/CC) gje sin eigen vurdering av kvar elev sin flydugleik, leiareigenskapar, militær framferd, haldningar og potensiale til å fullføra opplæring på andre flytypar etterpå (U. S. Air Force, 2010b). MASS-skår var meint til å hjelpa avgjerdstakarane med å vedta kva opplæring dei nye flygaren skulle få etter ENJJPT (jagarfly eller ikkje-jagarfly). Eit eksempel på ENJJPT *Merit Ranking* og vekting av elevane sine T-37 og T-38 flyprestasjonar er presentert under. Opp gjennom 90-talet telte flyginga 70 % og teori 10 %. FLT/CC si subjektive vurdering utgjorde dei siste 20 %. Innanfor flyrelatert skår telte T-37

prestasjonar 25 % og T-38 prestasjonar 75 %, der *Checkride*-resultat vog tyngst (U. S. Air Force, 2000):

<b>FLYING</b>		<b>70%</b>
Checkride Maneuver T-Score		<b>50%</b>
<b>T-37</b>	<b>T-38</b>	
Contact (16%)	Contact (24%)	
Advanced Contact (24%)	Instrument (28%)	
Instrument (24%)	Basic Formation (28%)	
Formation (20%)	Advanced Formation (20%)	
Low Level (16%)		
Daily Performance T-Score		<b>20%</b>
Flight Line Tests (25%)		
Daily Maneuver Scores (75%)		
<b>ACADEMICS T-Score</b>		<b>10%</b>
<b>FLIGHT COMMANDER RANKING T-Score</b>		<b>20%</b>
<b>TOTAL</b>		<b>100%</b>

Figur 3: ENJJPT Merit Ranking

Flyprestasjonane har av naturlege årsaker alltid hatt størst innverknad på rangeringa av flyelevane ved ENJJPT og prestasjonar på T-38 har hatt større betydning enn prestasjonar på T-37. I 2010 hadde fordelinga endra seg til 40 % T-6 og 60 % T-38 (U. S. Air Force, 2010b).

Dei norske flygarane som gjennomførte ENJJPT mellom 1989 og 2004 var i utgangspunktet kvalifisert for og tiltenkt å fly F-16 i Noreg etter endt utdanning. Rett nok var det stundom nokre individ som av ulike årsaker ikkje fullførte dei nasjonale F-5- og F-16-kursa, men dei fekk oftast tilbod om å fly annan flytype (til dømes C-130, Sea King eller DA-20). Den einskilde si rangering i klassen var mindre viktig for dei norske elevane. Så lenge Luftforsvaret stadig hadde behov for å trenar opp nye F-16 flygarar var det forutbestemt at alle skulle få tilbod om F-16-opplæring etter fullført utdanning.

## 4 Metode

### 4.1 Deltakarar

Alle jagarflygarar som graduerte frå ENJJPT i perioden 1989 - 2004 vart undersøkt i denne studien. Den spesifikke perioden vart valt fordi kandidatane gjekk gjennom nær identiske opptaksprøvar og fordi flyprestasjonane deira vart registrert ved hjelp av same vurderingsverktøy ved ENJJPT. Samstundes viste det seg at andre nasjonale flygarseleksjonsstudiar knytt til flyprestasjon òg kjem frå denne perioden. På det viset har ein tilgjengelege studiar det kan vera relevant å samanlikna med.

I alt 177 kandidatar frå Luftforsvaret vart sendt til Texas for å utdanna seg til jagarflygarar i dette tidsrommet. 175 gjennomførte utdanninga og fekk flygarstatus i Luftforsvaret ( $N = 175$ ), blant dei to kvinner. Den nøyaktige alderssamansetninga er ukjend då fødselsdato ikkje er oppgjeve i datasettet. Sidan eit av opptakskrava til LFS er alder mellom 18 og 24 år vil dei fleste av deltakarane vera omlag to år eldre medan dei utdanna seg i USA.

Fråfallet undervegs i seleksjonprosessen i perioden 1989-2004 var høgt, og det same er tilfelle i dag (samtale med sjef Seleksjonssenteret, jan. 2013). Undersøkjingar har vist at om lag 10 % av dei som møtte opp til fyrste fase i seleksjonen vart seinare teke opp som flyaspirantar ved LFS (Martinussen & Torjussen, 1998). Fråfallet i seleksjonsflygingsperioden ved LFS har over tid vore om lag 50 %. Dei siste åra har Luftforsvaret sendt seks flyelevar til ENJJPT pr. år. Det er svært mange søkjarar samanlikna med talet på utdanningsplasser og eit stort antal kandidatar lyt selekterast bort.

### 4.2 Prediktorar

Prediktorar i denne oppgåva var testgruppene 1 – 4. Desse gruppene var basert på dei testane som var teljande og med det gjort gjeldande for opptak til LFS i undersøkingsperioden. Nokre testar var ikkje teljande for opptak, men likevel ein del av prøvebatteriet. Testgruppene var:

Testgruppe 1 - Generelt evnenivå

Testgruppe 2 - Teknisk innsikt og romopplevelse

Testgruppe 3 - Simultankapasitet

Testgruppe 4 - Overføring/retningssevne

Gjennomsnittleg testresultat frå gruppe 1 – 4 var hovudgrunnlaget for opptak til LFS og var meint å seia noko om den einskilde kandidat sine flygarevner eller potensiale for flyging (Hansen et al., 1989). Testgruppene vart inndelt på denne måten på grunn av faktoranalysar som

vart utført i høve innføringa av *1980-batteriet* og var avgjerande for opptak som flyaspirant (Hansen et al., 1989). Gjennomsnittresultat under verdi 5 (Stanine 1-9 skala) vart vanlegvis underkjent, men stundom kunne òg resultat *over* 5 verta underkjent (Forsvaret, 1989). Luftforsvaret nytta seg av desse testgruppene for seleksjon til LFS og denne oppgåva nyttar difor dei same testgruppene som hovudprediktorar for flygarprestasjonar under jagarflygarutdanninga.

Ei årsak til at ein valde å operera med testgrupper var at dei ulike testane som inngjekk i dei respektive gruppene var meint å måla meir eller mindre dei same kognitive evnene. Då er det rimeleg å nytta gjennomsnittsskåren innanfor kvar testgruppe. Skåren vil vera meir reliabel enn ein enkelt test åleine. Ei samling av fleire liknande evnetestar bidreg til at den psykologiske testinga vert meir robust (Bailey & Woodhead, 1996). Ved å vurdera testgruppene i forhold til kvarandre i tillegg til skår pr. gruppe sikrar ein at kandidaten ikkje har spesielt svake områder. Dette kunne vorte oversett om ein kun såg på totalsnittet over alle testane. Då desse prediktorane ikkje har vore vurdert mot ENJJPT prestasjonskriteriar tidlegare, gjer det sitt til at dette er interessant å undersøkje vidare.

Gjennomsnittet av testgruppe 1 – 4 er òg tatt med som prediktor. Det er valgt fordi det kan gje ein samla effekt av alle testane, eller nytten av testbatteriet samla i forhold til prediksjon. Det er vanleg i høve den praktiske vurderinga av testresultata at ein ser både på enkeltgrupper og på totalgjennomsnittet av testane.

I tillegg til testgruppe 1 – 4 og gjennomsnittet av desse, ynskjer eg å undersøkje Alminneleg evnenivå (AE) frå sesjon som ein prediktor i høve flygarseleksjon. Som nemnt tidlegare har undersøkjingar vist at AE har signifikant prediktiv validitet i forhold til kriteria akademisk og militærfagleg prestasjon (Kjenstadbakk, 2012). Dessutan er AE eit mål på eit individ sin generelle intelligens og Forsvaret nyttar dette målet for seleksjon til befalsutdanning i dag (samtale med Sjefpsykologen for Forsvaret, mai 2013).

#### **4.2.1 Presentasjon av testane**

I dei neste avsnitta vil eg kort presentera dei testane/prøvene som var med i *1980-batteriet*. Denne skildringa er i all hovudsak henta frå Militærpsykologiske meddelelser (MPM) nr. 15: Psykologiske metoder ved flygeruttak, del A: Beskrivelse av testene (Hansen, 1987), MPM nr. 16: Del B: Utprøving av nytt testbatteri (Hansen, Heim, Nordby, Storsve, & Værnes, 1989), hovudoppgåva *Seleksjon av flygere* til Martinussen (1989) og artikkelen *Forsvaret, best i test* av Torjussen og Hansen (1999) publisert i Tidsskrift for Norsk Psykologiforening.



Gr. 1: GENERELT EVNENIVÅ vart utrekna som eit gjennomsnitt av dei to testane Raven Advanced Matrices og 43 KF Tallserier:

Raven Advanced Matrices: Denne prøven er basert på J. C. Raven sine *Progressive Matrices* frå 1938. Raven meinte han kunne måla eit individ sin kapasitet til å gjera samanlikningar, analogiar og evna til logisk tenkning. Testen omfatta 36 oppgåver som gjekk ut på å finna rett svar blant fleire figurar og finna den som passa inn i eit større mønster der ein bit manglar. Prøven var normert for søkjarar til LFS og oppgåvene hadde aukande vanskegrad.

43 KF Tallserier: Testen måler evna til numerisk resonnering. Kvar oppgåve i testen bestod av ein ufullstendig talserie. Kandidaten skulle finna ut kva prinsipp talserien var bygd opp etter og føra opp dei to manglande sifra i serien. Prøven hadde aukande vanskegrad.

Gr. 2: TEKNISK INNSIKT OG ROMOPPLEVELSE vart rekna som gjennomsnitt av dei fem testane 370 Mekaniske oppgåver, U3B-54 Plateforming, 391 Figurmønster, 402 Figurformer og romoppfatningsdelen av prøve 344. Sistnemnde vart etterkvart omplassert til testgruppe 3 (samtale med sjef Seleksjonssenteret, jan 2013).

370 Mekaniske oppgåver: I ein periode vart det nytta tre testar som målte teknisk innsikt; U3A på testar på sesjonsnivået, U3-49 på befalsskulenivået og *Prøve 370* for LFS og krigsskulenivå (Hansen et al., 1989). Her skulle ein ut frå fleire figurar finna fram til kva mekanske/tekniske prinsipp som låg til grunn og ut frå dette svara på spørsmål knytt til kvar figur. Oppgåvene hadde aukande vanskegrad.

U3B-54 Plateforming: Prøven var ein tredimensjonal spatialtest som ein meinte målte «manipulerande visualisering». Det vil seia ei kognitiv handsaming av visuelle objekt som involverte evna til å førestilla seg korleis eit mønster vart sjåande ut etter breiting eller rulling utført etter fastlagde reglar.

391 Figurmønster: På grunnlag av dei undersøkingane som vart gjort med *Spatial Relation Test* vart denne versjonen av Figurmønster laga og innført i *1980-batteriet*. Den varianten som Forsvaret nytta vart laga av Ivar Hansen. I opphavet kom denne type testar frå undersøkingar knytt til ein spatialfaktor leia av Spearman i 1920-åra. Slike testar måler rompersepsjon eller romoppfatning og inngjekk vanlegvis i testbatteri som målte mekanisk eller teknisk innsikt. Oppgåva bestod av kvadratfigurar som mangla ein del. I ei rekkje til høgre for oppgåvefiguren var det fire figurar der ein av dei var den manglande delen. Oppgåva vart løyst ved å finna den rette delen. Skår var antal rette svar minus talet på feil svar i løpet av 5 min.

402 Figurformer: På denne testen var ein del figurar satt opp på venstre sida av eit ark. Oppgåva gjekk ut på å finna att desse figurane blant ei mengd andre figurar som var rotert og plassert til høgre på arket. Skår var antal rette svar minus talet på feil i løpet av 10 min.

Prøve 344 Tidsvurdering og romoppfatning: Denne prøven bestod av to oppgåver som skulle løysast samstundes. Testen vart i opphavet laga for å kunna avdekka evna til å estimera tid mest mogeleg korrekt i forstyrrende situasjonar. Lengda på 20 tidsintervall skulle estimarast. Medan tidsestimeringa føregjekk skulle kandidatane løysa oppgåver knytt til romoppfatning. I romoppfatningsoppgåva vart kandidaten presentert for bilete av klossar som var satt oppå og inntil einannan. Kandidatane skulle telja berøringspunkt mellom ein spesiell klosse og dei andre som var plassert rundt. For flygaruttaka var det romoppfatningsaspektet som vart sett på som det viktigaste og tidsintervallvurderinga meir som det forstyrrende simultanelementet i prøven. I høve innføringa av *1980-batteriet* vart det vedteke at både romoppfatningsdelen og tidsdelen skulle nyttast. Rom- og tidsvurderingsdelen vart innført som to separate mål. Oppgåva vart med det å vurdera tidsintervall og telja klossar som var i kontakt med kvarandre samstundes. Skår var antal rette svar minus talet på feil.

323 Flykunnskap: Dette var ein test i kunnskap om fly og flyging av fleirvalgstypen. Testen var ein del av *1980-batteriet*, men var ikkje teljande for opptak til LFS. Den vart nytta meir som eit grunnlag for psykologintervju med tanke på motivasjon enn for seleksjon (Martinussen, 1989). Tanken var at motivasjon for yrket skulle ha samanheng med interesse for fly og flyging (Hansen, 1987). Nasjonale undersøkingar viste at prøven ikkje skilte i det heile teke mellom dei som vart flygarar og dei som ikkje greidde det (Hansen et al., 1989). Samstundes viser resultatata frå ei nasjonal meta-analyse at denne prøven hadde prediktiv validitet for kriteriet bestått ikkje bestått LFS med gjennomsnittleg  $r = .22$  ( $N = 973$ ) (Martinussen & Torjussen, 1998). Denne testen er framleis ikkje teljande for opptak til LFS.

Matematikk- og engelskprøve: Kandidatar til LFS måtte besvara to engelskprøvar og ein matematikkprøve. På same måte som Prøve 323 Flykunnskap betydde desse prøvane lite i forhold til seleksjon, men fungerte meir som tilleggsinformasjon (Martinussen, 1989). I dag vert prøvane omtala som 127 Engelsk og 255 Regneproblem.

Gr. 3: SIMULTANKAPASITET var prega av psykomotoriske evnetestar. I denne testgruppa var det fire testar (Hansen, 2006):

350 Sorteringsprøven: Dette var ei komplisert koordineringsoppgåve. Prøven var meint til å selektera personell med god konsentrasjonsevne, noko som gjorde at dei kunne reagera raskt og

nøyaktig ut frå fleire samtidige instruksar i ein tidspressa situasjon. Kandidatane skulle her leggja ut farga brikker (fire fargar) i ein pappboks med ni hol fordelt på tre rekkjer. Kwart hol var gjeve to kjenneteikn; eit fargeband og eit fargenamn som stort sett var i konflikt med kvarandre. Brikkene skulle leggjast ut etter ordrar gjeve på lydband, eks: «blå –tre bank – blå», det vil seia at ei blå brikke skulle leggjast i rekkje nummer tre der fargenamnet var blå. Oppgåvene vart gjevne i aukande tempo. Prøven vart vidareført på *1980-batteriet* utan endringar. Skår var antal rette svar minus talet på feil med ei prøvetid på 3 minutt og 18 sekundar.

310 Retningsreversaller: Med denne prøven tok ein sikte på å sila ut dei kandidatane som hadde problem med å halda greie på retningane i horisontal- eller vertikalplanet. Det kan vera eit fatalt problem for flygarar. Problema oppstår gjerne når flygaren er trøyyt eller på annan måte befinn seg i ein stressfull situasjon. Riis hadde indirekte berørt problemet om retningsforvirring i ein diskusjon frå 1955 om romorienteringsproblem hjå flygarar i høve risiko for ulykker (Hansen, 1987). Prøven har vore teljande for flygaropptak sidan midten av 70-talet og vart vidareført på *1980-batteriet* utan endringar. Oppgåvearket var inndelt i 16 ruter der kvar rute hadde eit namn etter kvar dei var plassert (venstre-venstre, venstre-høgre/ned-opp, opp-opp). Ordrar over kvar eit fargestiftmerke skulle teiknast vart gjeve over lydband. Tempoet vart auka etter kvart. Skår var antal rette svar minus talet på feil over ei prøvetid på 3 minutt og 48 sekundar.

410 Tallsøker: Prøven gjekk ut på å finna svarte tal i rekkjefylgje frå 1 til 100, der tal av ulik storleik var spreidd utover eit ark. På svararket skulle kandidatane skriva ned det raude talet som stod i same rute som oppgåvetalet. Prøven vart assosiert med evna til korttidsminne, rask dømmekraft, høgt tempo og effektivitet. Desse evnene vert redusert etterkvart som ein vert eldre. Prøven vart bestemt å ta inn i *1980-batteriet* under simultankapasitetsområdet. Skår var antal rette svar minus talet på feil med ei prøvetid på 10 minutt.

300 Instrumenttyding: Denne testen skulle måla kandidatane si evne til å visualisera korleis eit fly oppfører seg i det tredimensjonale rom, samt evna til å lesa av ulike instrument og kunna kombinera informasjonen frå dei. Kandidatane vart presenterte for to instrument; eit kompass og ein kunstig horisont. Basert på informasjon frå desse skulle dei avgjera kva bilete av dei fem avbilda flya dei to instrumenta var knytt til. Testen lada høgt på ein romfaktor og ein resonneringsfaktor. Prøven vart vidareført på *1980-batteriet* utan endringar.

Ein anna test som var assosiert med denne gruppa var «Tappetesten». Denne testen vart administrert under psykologintervjuet. Testen var i all hovudsak ein sjekk på simultankapasitet (Hansen et al., 1989). Resultat frå denne prøven er ikkje teke med i denne studien då

administreringa av testen i for liten grad var standardisert, samt at resultatata vart registrert ulikt (Martinussen, 1989). Det vart difor ikkje nytta ressursar på å henta inn data frå «Tappetesten» for deltakarane.

Gr. 4: OVERFØRING/RETNINGSEVNE kan seiast å dekke evna til det å overføre kunnskapar til den praktiske flygarsituasjonen, eller det å kunna utføre prosedyrar i lufta (Hansen et al., 1989). Testgruppa bestod av prøvane 330 Romoppfatning, 360 Spegltesten og tidsvurderingsdelen av prøve 344.

330 Romoppfatning: Dette var ein test der kandidaten skulle plassera seg i eit tenkt flysete. Visuelt vart bilete av eit stilisert fly presentert i ulike posisjonar (framfrå, bakfrå, opp eller ned, venstre eller høgre sving). Oppgåva gjekk ut på å forflytta flyet til ein annan stad i biletet, og etter spesifikke reglar teikna ned kva stikkebevegingar som måtte nyttast for å få dette til. Biletet vart vist i 2 sekundar og det vart gjeve 8 sekundar til å avgje svar (Hansen, 2006). Skår var antall rette svar.

360 Spegltesten: Denne prøven var ein individualtest. Her skulle kandidaten føra ein metallpenn langs eit mønstra spor. For å fylgja sporet fekk ikkje kandidaten mogelegheit til å sjå på sporet direkte, men måtte sjå via ein spegel. Det vart registrert feil viss kandidaten kom bort i kanten av sporet, hadde for hardt trykk, løfta pennen opp frå sporet eller la handa si ned på apparaturet. I tillegg vart rundetida målt. Kvar gong kandidaten gjorde ein feil kom ein «buzzerlyd». Ein meinte at spegltesten lada høgt på ein koordineringsfaktor som gjekk på evna til å koordinera synsinntrykk og handrørsle. Reint konkret vart kandidatane målt på kor raske dei var til å korrigerer feil undervegs. Prøven vart vidareført på *1980-batteriet* meir eller mindre uendra. Etter kvart vart sjølvte mønsteret endra frå stjerneform til eit spor som ikkje ville vera like lett å trenar opp på førehand (versjon 361).

Tidsvurderingsdelen av prøve 344: Testen vart i opphavet laga for å kunna avdekke evna til å estimera tid mest mogleg korrekt i forstyrrende situasjonar. Lengda på 20 tidsintervall skulle estimarast på denne testen samstundes som kandidaten skulle løysa oppgåver knytt til romoppfatning (jfr. prøve 344 Tidsvurdering og romoppfatning i høve gjennomgangen av Gr. 2).

Alminnelig evnenivå: AE målet bestod av resultatet frå testar som ikkje høyrte til testgruppe 1 - 4. Desse var Rekneoppgåver (U4-54), Figurmønster/Matriseprøve (U5-54) og Ordforståelse (U6-54) (Hansen, 2006). I dag består AE av dei tre testane Regning, Matriser og Ordprøve (Forsvaret, 2013).

### 4.3 Kriterier

Med utgangspunkt i tidlegare studiar der prestasjonskriterier under jagarflygarutdanning i USA og Canada var involvert, med spesiell vekt på studiar av flygarprestasjonar ved UPT/SUPT-utdanningsinstitusjonar i USA, har eg vald fylgjande kriteria:

1. *T-37 / T-38 Academics* – gjennomsnittleg eksamensresultat oppnådd gjennom teoridelen av utdanninga oppgitt i prosent (0 – 100) korrekt besvarelse for både T-37 og T-38. Det var krav til 85 % korrekte svar ved eksamen (*multippel choice*-type) i mellom anna aerodynamikk, meteorologi, T-37/T-38 fly- og motorlære, instrumentlære, flyplanlegging og navigasjon med meir (Carretta, 2011).
2. *T-37 Daily Flying Grades* – gjennomsnittleg karaktervurdering i prosent (0 – 100) i høve dagleg flyging og opplæring/instruksjon innanfor flydisiplinane *Contact, Instrument, Formation og Navigation*. Prestasjonsvurderingane vart utført av sertifiserte ENJJPT T-37 flyinstruktørar frå dei ulike nasjonane som deltok i programmet.
3. *T-38 Daily Flying Grades*– gjennomsnittleg karaktervurdering i prosent (0 – 100) i høve dagleg flyging og opplæring/instruksjon innanfor flydisiplinane *Contact, Instrument, 2-ship & 4-ship Formation og 2-ship Navigation*. Prestasjonsvurderingane vart utført av sertifiserte ENJJPT T-38 flyinstruktørar frå dei ulike nasjonane som deltok i programmet.
4. *T-37 Checkride Grades* – gjennomsnittleg oppnådd eksamenskarakter i prosent (0 – 100) i høve *Checkride* innanfor flydisiplinane *Contact, Instrument, Formation og Navigation*. Prestasjonsevalueringane vart utført av sertifiserte ENJJPT T-37 Checkride-pilotar frå dei ulike nasjonane som deltok i programmet.
5. *T-38 Checkride Grades*– gjennomsnittleg oppnådd eksamenskarakter i prosent (0 – 100) i høve *Checkride* innanfor flydisiplinane *Contact, Instrument, 2-ship & 4-ship Formation og 2-ship Navigation*. Prestasjonsevalueringane vart utført av sertifiserte ENJJPT T-38 Checkride-pilotar frå dei ulike nasjonane som deltok i programmet.

T-37 kriteria representerer *Basic Jet Flying* (fase II) og T-38 kriteria representerer *Advanced Jet Flying*. T-38 er ein *fast jet* og ligg nærare opptil seinare flyging med jagarfly enn det T-37 gjer, både i høve maksimal flyhøgde, -hastighet og rekkevidde.

I høve *Air Education and Training Command Instruction (AETCI) 36 – 2205* (U. S. Air Force, 2003) skal rangering av elevane føregå ved å nytta skår på flygarprestasjonar og teori. Subjektive vurderingar frå Flight Commander skal maksimalt utgjera 20 % av total poengsum (jfr. figur 3).

Kvar flymanøver og ulike aktivitetar utført i flyet (til dømes: oppstart og avgang, flyging til treningsområda, loop eller landing) vart tildelt ein poengsum basert på korleis eleven presterte. Eleven fekk null poeng viss han/ho var ukunnig (*U – Unable*), ni poeng for trygg prestasjon (*F – Fair*), 12 poeng for tilfredsstillande prestasjon (*G – Good*) og 15 poeng for sær god prestasjon (*E – Excellent*) (korrespondanse med Mr Michaels, Randolph AFB. AETC/A3F, april 2013). Dess meir krevjande ein manøver var dess meir vektning hadde han (U. S. Air Force, 2010b). Ein *Checkride* vurdert som feilfri, det vil seia utan trekk (*zero downgrades*), ville koma ut med ein skår på 100 %. For å oppnå eit slikt resultat måtte eleven motta karakteren *Excellent* på alle element som var til vurdering. Skår på ein *Checkride* vart då det prosentvise forholdet mellom poeng mottatt i høve kor mange poeng det hadde vore mogeleg å få ((poeng mottatt/poeng mogeleg) x 100) (korrespondanse med Mr Michaels ved HQ/AETC/A3F, april 2013). Fyrste forsøk på kvar manøver vart ståande som teljande. Ein elev som strauk kunne oppnå eit betre resultat på eksamen enn ein elev som bestod den same eksamensturen, fordi sistnemnde kan ha fått fleire *downgrades*.

Karakterane for manøvrar utført i høve dagleg flyging hadde same poengverdi, men vart utdelt på ein annan måte. I høve opplæring og instruksjon vil det vera mest normalt at elevar byrjar på dei låge karakterarne og arbeider seg gradvis opp mot *Good*, som var eit godkjent prestasjonsnivå. Sagt på ein litt forenkla måte: flyinstruktørane kunne ikkje venta at ein uerfaren flyelev kunne landa jettfly perfekt på dei fyrste forsøka. Vanleg progresjon ville vera at ein elev presterte på U-nivået før instruktørane etterkvart vart overtydd om at han faktisk kunne meistra manøveren på ein trygg måte kvar gong. Fyrst godt uti programmet ville den tilfredsstillande karakteren *G* normalt oppnåast. På den andre sida burde Checkpilotane forventa at elevane presterte tilfredsstillande på fyrste forsøk, sidan dei trass alt var meldt opp til eksamen. Gjennomsnittskarakterane som vart oppnådd i høve *Daily Flying* (60 – 75 %) var difor vesentleg lågare enn *Checkride* karaktersgjennomsnittet (90 – 100 %).

Det var to ulike måtar å gje karakterar på i høve flyturene; ein absolutt og ein relativ karakterskala (U. S. Air Force, 2000). Den fyrstnemnde skulle nyttast i høve vurdering av den einskilde manøveren. Korleis eleven hadde prestert på denne manøveren tidlegare skulle ikkje takast omsyn til. Vurderinga skulle kun ta omsyn til faktisk prestasjon der og då sett i forhold til krava. Relativ karakterskala skulle nyttast i høve flyturen sin hovudkarakter. Her skulle instruktøren sjå på prestasjonane i ein større samanheng. På *checkrides* skulle ein kun nytta den absolutte karakterskala. Det er heilt avgjerande at instruktørane sine vurderingar av elevprestasjonar føregår på ein systematisk og påliteleg måte for at ein skal unngå kjende

feilkjelder knytt til slike personvurderingar (Martinussen & Hunter, 2008). Dei ulike måtane å gje karakterar på lyt difor haldast frå einannan.

#### 4.4 Prosedyre

Deltakarane sine utdanningsresultat frå ENJJPT vart samla inn ved Air Education Training Command HQ, Randolph Air Force Base, Tx av sjef Luftforsvarets utdanningsinspektorat (LUI). Innsamla ENJJPT-data vart sendt vidare til Sjefpsykologen for Forsvaret. Resultata frå flygarsleksjonen for deltakarane vart henta inn frå Seleksjonscenteret ved Rygge Flystasjon som oppbevarer seleksjonsdata for Sjefpsykologen. Sjefpsykologen for Forsvaret kopla datalistene saman, aidentifiserte dei og gjorde dei tilgjengeleg for analyse. Datamaterialet kan ikkje identifisera einskildpersonar og er derfor ikkje melde- og konsesjonspliktig etter personopplysningslova.

#### 4.5 Statistiske analysar

Analysane i denne studien vart utført ved hjelp av *Statistical Package for the Social Sciences* (SPSS, versjon 20). Programmet vart nytta til å analysere moglege samanhengar mellom prediktorar nytta ved flygarseleksjon i Luftforsvaret og dei multiple prestasjonskriteria som vart målt gjennom utdanninga ved ENJJPT. Ei korrelasjonsanalyse vart utført for å avdekke samanhengar mellom prediktorane og kriteria. Hierarkisk multiple regresjonsanalysar vart gjennomført for å kasta lys over i kva grad fleire prediktorar til saman predikerte eit kriterium (Martinussen & Hunter, 2008). Regresjonsanalysane kunne òg nyttast for å visa kor stor del av variansen i den avhengige variabelen som i statistisk forstand kunne forklarast av regresjonsmodellen. Statistiske *power* tabellar viste at eit utval på 175 deltakarar er godt nok for å avdekke korrelasjonar rundt  $r = .20$  og  $r = .30$  med signifikansnivå på .05-nivået for ein tohala test. Statistisk *power* er i dette tilfellet høvesvis .76 og .98, ein akseptabel styrke i høve tabellen.

Det var ikkje mogleg å henta komplette testresultat for alle deltakarane ( $N = 175$ ). Seleksjonscenteret på Rygge flystasjon viser til at testresultat gjennom fleire år vart manuelt ført på *gule ark* arkivert i personlege mapper ved senteret. På eit ukjent tidspunkt gjekk ein over til å føra opp resultata digitalt. I same periode gjekk seleksjonprosessen gradvis over frå å nytta papir- og blyanttestar med manuell registrering til databaserte testar. Basert på dei *gule arka* i arkivet og nokre digitale testresultat har ein klart å henta inn komplette testgrupperesultat for eit antal deltakarar for testgruppe 1 ( $N = 164$ ), testgruppe 2 ( $N = 147$ ), testgruppe 3 og 4 ( $N = 157$ ) og for AE ( $N = 167$ ). Det er ingen god grunn til å hevda at *missing* data er systematisk fråverande. Meir sannsynleg er det at ein her ser noko mangelfulle rutinar i høve registrering av prøveresultat og at manglande data har forsvunne på tilfeldige måtar.

## **5 Resultat**

### **5.1 Korrelasjonsanalyse**

Tabell 1 gjev ei oversikt over korrelasjonar mellom individuelle testar nytta i høve opptaksprøvene (fase 1 og 2) til LFS opp mot ENJJPT -prestasjonskriteria. Interkorrelasjon mellom dei ulike ENJJPT kriteria og deskriptiv statistikk er utelete av plassomsyn. Desse er oppgjevne i tabell 2.



Tabell 1. Bivariate korrelasjoner for individuelle testar opp mot ENJJPT kriteria (N = 140 - 175)

Variablar	1	2	3	4	5	6	7	8	9	10	11	12	13	14
<b>Prediktorar</b>														
1. Raven Advanced														
2. Tallserier	.23**													
3. Mekaniske oppg.	.19*	.04												
4. Figurmønster	.27**	.23**	.13											
5. Figurformer	.11	.27***	.03	.29***										
6. Plateforming	.31***	.05	.28***	.33***	.18*									
7. Romoppfatning (344)	.12	.10	.00	.22**	.11	.06								
8. Sortering	.11	.02	-.07	.12	.08	-.04	.05							
9. Retn.reversaller	.10	.17*	-.04	-.02	.07	.09	.07	.34***						
10. Tallsøker	.11	.18*	-.07	.14	.08	.06	.12	.04	.14					
11. Instrumenttyding	.17	.15	.09	.30***	.12	.22**	.28***	.08	.14	.02				
12. Romoppfatning (330)	.17*	.16*	.05	.21*	.08	.14	.05	.26**	.18*	.07	.16*			
13. Tidsvurdering (344)	-.06	-.04	-.04	-.04	.12	.03	.07	-.02	.04	.03	.06	.09		
14. Speltesten	.19*	.08	.10	.18*	.10	.23**	.12	.15	.20*	.09	.18*	.03	.11	
<b>Kriterier</b>														
15. T-37/T-38 Academics	.00	.14	.00	.20*	-.09	-.14	.01	.20*	.21**	.20*	-.00	.16	-.00	.03
16. T-37 Daily Flying	.01	.10	-.02	.20*	.04	.06	.09	.26**	.29***	.06	.14	.05	.09	.25**
17. T-38 Daily Flying	.05	.04	-.06	.23**	.11	.10	.20*	.31***	.30***	.11	.10	.05	.04	.32***
18. T-37 Checkrides	-.00	-.02	-.04	.11	.10	-.05	.02	-.03	.10	.11	.04	.08	.05	.06
19. T-38 Checkrides	.05	-.00	.09	-.06	.07	-.01	.00	.02	.19*	.03	.17*	-.01	-.00	.14

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$  (to-hala)

Tabell 1 viser at testane Sortering, Retningsreversaller og Spegltesten er dei tre testane som oppnår høgast korrelasjon med kriteria. Korrelasjonen er høgast i forhold til *T-38 Daily flying* med  $r = .20$  til  $r = .31$ . Testane Figurmønster og Romoppfatning (344) korrelerte òg med dagleg flyging. Dei andre testane korrelerte ikkje med daglege flygarprestasjonar.

Testane Sortering og Retningsreversaller korrelerte i tillegg med *Academics*. Det same gjorde testane Figurmønster og Tallsøker (begge  $r = .20$ ). Retningsreversaller og Instrumenttyding var dei einaste testane i *1980-batteriet* som korrelerte med *Checkride*-kriteria. Desse to testane korrelerte med *T-38 Checkride* med  $r = .19$  og  $r = .17$ , høvesvis.

Tabell 2 gjev ei oversikt over deskriptiv statistikk og korrelasjonar mellom ulike prediktorar som vart nytta ved opptaket til LFS og flygarprestasjonskriteria som vart nytta i høve jagarflygarutdanninga ved ENJJPT i perioden 1989 - 2004.

Tabell 2. Gjennomsnitt, standardavvik og bivariate korrelasjoner for variablene i studien (N = 147 - 175)

Variablar	M	SD	0	1	2	3	4	5	6	7	8	9
<b>Prediktorar</b>												
0. Alminnelig evnenivå (AE)	7.5	1.04										
1. Gen.evnenivå (Gr.1)	6.4	1.18	.21**									
2. Tekn.Ins/Romopplev. (Gr.2)	6.9	1.01	.34***	.39***								
3. Sim.kapasitet (Gr.3)	6.9	0.92	.29***	.21**	.12							
4. Presisjon/Retn.evne (Gr.4)	6.1	1.05	.15	.12	.27**	.28***						
5. Gjennomsnitt av Gr. 1 - 4	6.6	0.70	.35***	.71***	.68***	.59***	.63***					
<b>Kriterier</b>												
6. T-37/T-38 Academics	96.2	2.28	.18*	.09	.04	.26**	.08	.15				
7. T-37 Daily Flying	63.1	6.84	.11	.07	.22**	.22**	.21**	.20*	.49***			
8. T-38 Daily Flying	65.8	9.82	.21**	.08	.26**	.27**	.20*	.25**	.34***	.78***		
9. T-37 Checkrides	94.3	2.41	-.01	-.03	.05	.09	.10	.02	.25**	.37***	.10	
10. T-38 Checkrides	93.5	2.29	-.04	.03	.08	.10	.07	.05	.14	.19*	.19*	.20**

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$  (to-hala)

Prediktoren AE frå sesjonsprøvene korrelerte med to ENJJPT-kriteria, *T-37/ T-38 Academics* og *T-38 Daily Flying* med høvesvis  $r = .18$  og  $r = .21$ . I tillegg korrelerte dette målet med gjennomsnittet av Gr. 1 – 4 med  $r = .35$ .

Testgruppe 1 - Generelt evnenivå korrelerte ikkje med ENJJPT-kriteria. Gr.2 Teknisk innsikt/romopplevelse, Gr.3 Simultankapasitet og Gr.4 Presisjon/retningsevne korrelerte med *T-37-* og *T-38 Daily Flying* kriterier. Desse korrelasjonane var i spennet  $r = .20$  til  $r = .27$  og var statistisk signifikante.

Tabell 2 viser vidare at ingen prediktorar korrelerte med kriterier *T-37-* og *T-38 Checkride*. Det var berre teori- og flygarprestasjonar undervegs i sjølve utdanninga som korrelerte signifikant med desse kriterier. I tillegg viser tabellen at prediktoren simultankapasitet korrelerte med kriteriet *T-37/ T-38 Academics* med  $r = .26$ .

Gjennomsnittresultatet av seleksjonstestgruppe 1 – 4 korrelerte med *T-37-* og *T-38 Daily Flying* (høvesvis  $r = .20$  og  $r = .25$ ). Gjennomsnittsskår av dei teljande testane for opptak til LFS hadde ingen samanheng med ENJJPT-teorieresultat eller *Checkride*-prestasjonar.

## 5.2 Multipel hierarkisk regresjonsanalyse

For å vurdere i kva grad dei ulike prediktorane til saman predikerte ENJJPT-resultat vart multipel hierarkisk regresjonsanalyse gjennomført. Variabelen AE vart utelatt då denne ikkje var ein del av avgjerdsgrunnlaget i høve opptaket til LFS slik som Gr. 1 – 4 var det.

I den fyrste regresjonsanalysen vart prediksjon av ENJJPT *T-38 Daily Flying* prestasjonar undersøkt. Korrelasjonsanalysen avdekkar at prediktorane korrelerte med *T-38 Daily Flying*-kriteriet. *T-38* prestasjon har mest å seia i høve rangering av elevane og desse samanhengane vert difor presentert fyrst. I tabell 3 er *T-38 Daily Flying*-kriteriet den avhengige variabelen og Gr. 1 - 4 er analysert steg for steg.

Tabell 3. Hierarkisk multipel regresjonsanalyse for *T-38 Daily Flying*.

Variablar	<i>T-38 Daily Flying</i>	
	$\Delta R^2$	$\beta$
Steg 1	.01	
Gr.-1 Generelt evnenivå		-.05
Steg 2	.06**	
Gr.-2 Teknisk innsikt/Romoppfattelse		.23**
Steg 3	.06**	
Gr.-3 Simultankapasitet		.23**
Steg 4	.00	
Gr.-4 Presisjon/retningsevne		.07
$R^2$	.13**	
$N$	147	

Note: Alle koeffisientar er tekne frå siste steg i regresjonsanalysen.

\*  $p < .05$ ; \*\*  $p < .01$

Tabell 3 viser at både Gr.2 og Gr.3 forklarar 6 % av variansen for *T-38 Daily Flying* prestasjonar på signifikant vis. Heile modellen med alle tre prediktorane inkludert forklarar 13 % av variansen. Generelt evnenivå kan forklara 1 % av variansen. Denne har ein svak negativ  $\beta$ , men det kan vera tilfeldig. Gr.4 Presisjon/retningsevne forklarar ingenting i tillegg etter at det er kontrollert mot dei tre fyrste prediktorar. Både Gr.2 og 3 var statistisk signifikante med høvesvis  $\beta = .21$  og  $\beta = .22$ .

I den andre regresjonsanalysen vart prediksjon av ENJJPT *T-37 Daily Flying* prestasjonar undersøkt. Korrelasjonsanalysen viste at prediktorane korrelerte signifikant med dette kriteriet. I tabell 3 er *T-37 Daily Flying*-kriteriet den avhengige variabelen og Gr. 1 - 4 er analysert steg for steg.

Tabell 4. Hierarkisk multipel regresjonsanalyse for *T-37 Daily Flying*.

Variablar	<i>T-37 Daily Flying</i>	
	$\Delta R^2$	$\beta$
Steg 1	.01	
Gr.-1 Generelt evnenivå		-.00
Steg 2	.04*	
Gr.-2 Teknisk innsikt/Romoppfattelse		.17
Steg 3	.04*	
Gr.-3 Simultankapasitet		.18*
Steg 4	.00	
G.-4 Presisjon/retningsevne		.09
$R^2$	.09**	
$N$	147	

Note: Alle koeffisientar er tekne frå siste steg i regresjonsanalysen.

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

Tabell 4 viser at Gr.2 Teknisk innsikt/Romoppfattelse og Gr.3 Simultankapasitet forklarar 4 % av variansen for *T-37 Daily Flying* prestasjonar. Heile modellen med alle tre prediktorar inkludert forklarar 9 % av variansen. Gr.4 Presisjon/retningsevne forklarar ingenting i tillegg etter at det er kontrollert mot dei to fyrste prediktorane. Beta for Gr.3 var statistisk signifikant med  $\beta = .18$ .

I den tredje analysen vart prediksjon av ENJJPT *T-37-* og *T-38 Academics* prestasjonar undersøkt. I tabell 3 er *T-37-* og *T-38 Academics*-kriteriet den avhengige variabelen og Gr. 1 - 4 er analysert steg for steg.

Tabell 5. Hierarkisk multippel regresjonsanalyse for *T-37* og *T-38 Academics*.

Variablar	<i>T-37- og T-38 Academics</i>	
	$\Delta R^2$	$\beta$
Steg 1	.01	
Gr.-1 Generelt evnenivå		.07
Steg 2	.00	
Gr.-2 Teknisk innsikt/Romoppfattelse		-.03
Steg 3	.07**	
Gr.-3 Simultankapasitet		.26**
Steg 4	.00	
G.-4 Presisjon/retningsevne		.00
$R^2$	.08*	
$N$	147	

Note: Alle koeffisientar er tekne frå siste steg i regresjonsanalysen.

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

Tabell 5 viser at Gr.3 Simultankapasitet forklarar 7 % av variansen for *T-37-* og *T-38 Academics* prestasjonar. Heile modellen med alle tre prediktorar inkludert forklarar 8 % av variansen. Dette betyr at testane som høyrte til i Gr.3 Simultankapasitet forklarar det meste av det som kan forklarast med denne modellen. Modellen sett under eitt var statistisk signifikant og stigningstalet  $\beta$  var .26 for steg 3.



I den siste regresjonsanalysen vart prediksjon av ENJJPT *T-37-* og *T-38 Checkride* prestasjonar undersøkt. Trass i at ingen prediktorar korrelerte med dette kriteriet ynskjer eg likevel å presentera dette regresjonsanalyseresultatet. I tabell 6 ser ein Gr. 1 - 4 analysert steg for steg med *T-37-* og *T-38 Checkride* – gjennomsnittet er den avhengige variabelen.

Tabell 6. Hierarkisk multippel regresjonsanalyse for *T-37-* og *T-38 Checkride*.

Variablar	<i>T-37- og T-38 Checkride</i>	
	$\Delta R^2$	$\beta$
Steg 1	.00	
Gr.-1 Generelt evnenivå		- .01
Steg 2	.01	
Gr.-2 Teknisk innsikt/Romoppfattelse		.06
Steg 3	.02	
Gr.-3 Simultankapasitet		.12
Steg 4	.00	
G.-4 Presisjon/retningsevne		.06
$R^2$	.03	
$N$	147	

Note: Alle koeffisientar er tekne frå siste steg i regresjonsanalysen.

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

Tabell 6 viser resultatata for prediksjon av T-37 og T-38 eksamensflyging. Denne forklaringsmodellen var ikkje statistisk signifikant.

## 6 Diskusjon

### 6.1 Diskusjon av problemstilling 1

*Har ein skilde seleksjonstestar betre prediktiv validitet enn andre i høve ENJJPT utdanningsresultat?*

Figurmønster (eit mål på spatiale evner), Sortering og Retningsreversaller (mål på simultankapasitet) og Tallsøker var dei testane som hadde prediktiv validitet for ENJJPT teoriprestasjon. Dei to testane i Gr.1 som var meint å måla generell intelligens korrelerte ikkje med dette kriteriet. Carretta og Ree (1995) fann at *Arithmetic Reasoning test* (eit mål på *general cognitive ability*) var den beste prediktoren for UPT/SUPT *Academic Average* ( $N = 7563$ ). Prediktoren AE viste prediktiv validitet for ENJJPT teoriresultat ( $r = .18$ ).

I høve flygarprestasjonar viste testane Figurmønster, Romoppfatning (344), Sortering og Retningsreversaller samt Spegltesten korrelasjon i spennet  $r = .20$  til  $r = .32$  i høve dagleg flyging, både med T-37 og T-38, men T-38 spesielt. Dette resultatet stemmer godt overeins med at både spatiale evner, psykomotoriske evner og simultankapasitet har vist prediktiv validitet mot flygarprestasjonskriterer i forskinga. Funnet kan i tillegg tyda på at dei andre prøvane som inngår i dei same seleksjonstestgruppene måler desse evnene i mindre grad. I høve prediktiv validitet kan ein basert på dette funnet indikera at det er Figurmønsterprøven som dominerer Gr.2, medan testane Sortering og Retningsreversaller gjer det i Gr.3 og Spegltesten i Gr.4, sett opp mot ENJJPT *Daily Flying* karakterar.

Martinussen og Torjussen (1998) gjennomførte ei nasjonal validering av *1980-batteriet* i høve prediktiv validitet for LFS flygarprestasjon (hovudsakleg bestått/ikkje bestått LFS) ( $N = 244 - 977$ ). Meta-analysen viste at testane som hadde best gjennomsnittleg prediktiv validitet var Instrumenttyding ( $r = .29$ ), Mekaniske oppgåver ( $r = .23$ ) og den ikkje-teljande testen Flykunnskap ( $r = .22$ ).

Kunne ein kanskje teke bort dei andre testane og berre basert Gr. 1 – 4 på dei få testane som er nemnt her? Konklusjonar av den typen kan vera forhasta. Hansen et al. (1989) og Hansen (2006) omtalar slike problemstillingar og anbefalte i si tid å fortsetja med fleire testar innanfor kvar testgruppe. Når alt kjem til alt dreier det seg om at ein kandidat kan vera uheldig i høve gjennomføring av *ein* prøve i ei testgruppe, men meistra betre på dei andre innanfor same gruppe. Gitt at testane måler noko av dei evnene det er behov for, kan ein på ein betre måte sikra at kandidatar som eigentleg er gode nok ikkje vert selektert bort. Dei andre testane som ikkje vert

nemnd her, då dei manglar statistisk signifikans, kan likevel seinare visa seg å ha signifikans. Dette vil til dømes gjelda viss ein skulle få tilgang på større  $N$  ved eit seinare høve.

Trass i at ingen av prediktorane i studien har prediktiv validitet for T-38 *Checkride*-prestasjonar ved ENJJPT, er det to enkelttestar som faktisk har det: Retningsreversaller og Instrumenttyding. Dette finn eg ikkje noko god forklaring på anna enn at testen Retningsreversaller mellom anna krevde evne til å halda orden på retning både i horisontal- og vertikalplanet sjølv om ein er sliten eller befinn seg i ein pressa situasjon. Testen auka tempoet etter kvart og mot slutten var det særskilt krevjande å løysa oppgåvene raskt og presist nok. Dei flyelevane eg har tala med oppgjev ofte at *checkrides* vart opplevd som stressande, spesielt viss noko nytt og uventa oppstod i lufta, eller anna usikkerheit dukka opp. Gode evner knytt til retningssans vil kanskje kunna vera til hjelp for å halda orden på posisjonering medan ein flyg med høg fart og samstundes kanskje er uroleg og oppkava. Denne evna kan vera spesielt relevant for instrumentflyging, då ein i stressande situasjonar ikkje har høve til å sjå ut for å bli orientert om kvar ein er. Instrumettydingstesten var lada høgt med ein spatialfaktor og ein resonneringsfaktor (Hansen et al., 1989). Begge faktorane kan vera med på å forklara samanhengen med T-38 *Checkride*-prestasjon, slik som det er forsøkt i høve Retningsreversaller. Ei anna forklaring kan naturlegvis vera at det her er tale om ei form for *sampling error*. På den andre sida har testen vist prediktiv validitet for bestått/ikkje bestått LFS (Martinussen & Torjussen, 1998). Seleksjonsflyginga ved LFS er òg knytt til flyging under pressa- og stundom stressfulle situasjonar. Det same kan seiast å gjelda for ENJJPT *Checkrides*. Samstundes viser testen ingen prediktiv validitet mot T-37 *Checkrides* som òg kan vera like prega av stressframkallande situasjonar. For å seia noko meir om desse samanhengane lyt dette undersøkjast nærare.

## 6.2 Diskusjon av problemstilling 2

*I kva grad har dei ulike testgruppene (Gr. 1 - 4) prediktiv validitet i høve ENJJPT utdanningsresultat?*

Gr.1 Generelt evnenivå ser ikkje ut til å visa prediktiv validitet i høve ENJJPT prestasjonskriteria trass i at andre undersøkjingar indikerar at det faktisk er ein samheng mellom denne type testar og militærflygarprestasjon (Carretta og Ree, 2003; De Kock og Schlechter, 2009). For å forstå dette resultatet betre lyt ein fyrst avklara kva som ligg i omgrepet generelt evnenivå. Det er her tale om eit mål på generell intelligens (Martinussen, 1997). I høve Cronbach (1970) søkjer ein å måla intelligens som ein *general mental ability*. Denne evna er verdifull i nær sagt all form for

tankeverksemd og testar laga for å prøva å måla desse evnene vert ofte kalla intelligenstestar (Cronbach, 1970).

Så lenge ein opererer med omgrep som generell intelligens lyt det òg vera noko som vert omtala som spesifikk intelligens. Vanleg oppfatning i dag er at intelligens kan delast opp i fleire faktorar. Det er mogleg at dei ulike faktorane har *eitt* og same utspring, eller basis, i form av ein overordna generell intelligens, omtala som *g* eller *General Mental Ability* (GMA) (Carretta & Ree, 2003; Schmidt & Hunter, 1998). For Ree og Earles (1992, s. 86) var intelligens eit sentralt mål då dei hevda at «Intelligence is the best predictor of job performance» uavhengig av kva jobb det er tale om. Dette støttar òg Tett, Jackson, og Rothstein (1991) då dei uttalar at generelle kognitive evner er ein viktig faktor knytt til jobbprestasjon uansett jobb.

Viss det likevel er slik at Gr.1 faktisk måler noko anna enn *g*, kan det forklara kvifor ein ikkje finn samanheng med flygarprestasjonar. I den vidare diskusjonen tek eg utgangspunkt i at testgruppe 1 måler generell intelligens og at dette målet kan jamførast med *g*. Dette valet er fyrst og fremst basert på at ein til no ikkje har indikasjonar på at Ravens Matriser, som inngår i Gr.1, *ikkje* måler generell intelligens (samtale med Martinussen, mars 2013).

Generelle kognitive evner (*g*) spelar ei sentral rolle for å predikera flygarsuksess (De Kock & Schlechter, 2009). Carretta og Ree (1994; 2003) har jamvel reist spørsmål om det eigentleg gjev meining å måla spesifikk intelligens (*s*) sidan denne er så «metta» med *g* at det kan vera bortkasta tid. Testar som måler *s* bidreg ikkje nemneverdig til prediktiv validitet utover det testar som måler *g* gjer (De Kock & Schlechter, 2009). Dette bør ein ta omsyn til. Ein kan difor ikkje utan vidare konkludera med at dette ikkje gjeld for norske forhold. Samstundes har meta-analysar vist at testar som måler meir spesifikke kognitive evner har betre prediktiv validitet for flygarprestasjonar enn testar som måler meir generell intelligens (Hunter & Burke, 1994; Martinussen, 1997).

Statistiske analysar og deskriptiv statistikk knytt til Gr.1 har vist at kandidatane har oppnådd eit gjennomsnitt på 6,4 på desse prøvane ( $N = 164$ ). Årsaka til at psykologitenesta i si tid tok inn Raven *Advanced* i testbatteriet var for å nytta ein intelligenstest som skilte betre mellom dei beste kandidatane enn det AE testane frå sesjon gjorde (Hansen et al., 1989). Det same gjaldt testen Tallserier som òg var normert for søkjarar til LFS. Rekneoppgåver og ein norsk versjon av Raven matrisetestar inngjekk i sesjonsprøvene for å måla AE. Desse testane var normert for vernepliktige i høve sesjon og sidan dei var mindre krevjande enn LFS-testane kunne ikkje AE-resultatet skilja like godt på kandidatar som fekk høg skår.

Kva har dette å seia for denne studien? Det ser ut til at deltakarane har oppnådd gode resultat på testar som er meint å måla intelligens, både på sesjon (AE:  $M = 7,5$ ,  $SD = 1,04$ ,  $N = 167$ ) og Gr.1 testane ( $M = 6,4$ ,  $SD = 1,18$ ,  $N = 164$ ). Gjennomsnittleg skår på AE blant norske soldatar er meint å vera 5 på Stanine skala ( $SD$  på Stanine-skalaen er 2). Gjennomsnittet blant deltakarane i studien ligg med det litt over 1 standardavvik over gjennomsnittleg AE blant sesjonspliktige. Halvparten av flygarane i denne studien oppnådde ein AE-skår på 8 eller 9 ( $N = 83$ ). Medianskår var 7 ( $N = 167$ ).

Tabell 2 viser at prediktorane AE og Gr.1 korrelerar svakt men signifikant med kvarandre ( $r = .21$ ,  $N = 157$ ). Ei mogleg forklaring på dette kan være at AE og Gr.1-testane kanskje måler noko ulike aspekt ved intelligens. Samstundes kan nokre kandidatane ha vore uheldige med dagen då Gr.1-testane vart gjennomført, eller kanskje sesjons- og LFS-testane vart administrert på ulikt vis. Dessutan kan kandidatane ha hatt ulik motivasjon i høve gjennomføring av sesjon og opptak til LFS. Desse aspekta kan verka negativt inn på testreliabilitet.

Skulle kandidatane vera uheldige i høve gjennomføringa av Gr.1-testane ville det vera ugunstig, både for den einskilde og for Luftforsvaret. Kandidatane som hadde låg skår på Gr.1 (men høg skår på AE) ville truleg ikkje vorte teke opp som aspirantar basert på Gr.1-resultata åleine. Desse kandidatane fullførte trass alt utdanninga og fekk flygarstatus i Luftforsvaret. Ein kan kanskje hevda at det vil vera god grunn for å vurdera AE saman med Gr.1-resultata i høve flygarseleksjon. Søkjarane til LFS er ei prestasjonsorientert gruppe (Hansen, 2006). Det er rimeleg å anta at motivasjonen for å gjera sitt beste ville vera tilstades blant søkjarane under gjennomføring av LFS-opptaksprøvene, kanskje i større grad enn i høve dei obligatoriske sesjonsprøvene. Det kan gje god grunn til å leggja meir vekt på Gr.1-resultata, gitt at dei måler dei same evnene som AE.

Utover dette er det selektert strengt på Gr. 1 (Hansen, 2006) og ein kan difor få større problem med RR på desse testane i forhold til AE som kandidatane berre indirekte var selektert på. Streng seleksjon gjer at RR òg verkar inn og avgrensar interkorrelasjon mellom AE og Gr.1.

Basert på dei skåra som deltakarane i denne studien har oppnådd, både på AE og Gr.1-testane, verkar det åpenbart at det generelle evnenivået, eller intelligensmålet blant kandidatane har vore både høgt og tilfredsstillande i forhold til å gjennomføra utdanninga. Då denne testgruppa ikkje korrelerar med utdanningsresultatet bør ein leita etter forklaringar. Ei mogleg forklaring kan vera at utdanninga krev generell intelligens over eit visst nivå. For kandidatane som har skåra over eit slikt minstemål vert det kanskje irrelevant kor høgt over minstemålet dei har skåra? Då kandidatane vart strengt selektert på Gr.1 kan ein tenkja seg at dei kom over eit

minstenivå. Den manglande samanhengen med flygarprestasjonar vil i eit slikt tilfelle gje meining. Viss dette var tilfelle kan det vera andre aspekt med intelligensmålet, eventuelt andre evner eller eigenskapar, som i større grad spelar inn på prestasjonane for desse kandidatane.

I artikkelen *Intelligence is not the best predictor* vert det prøvd å tilbakevisa betydninga av intelligens, eller *g* om ein vil, i høve jobbprestasjonar (McClelland, 1993). McClelland hevdar at Ree og Earles (1992) tok feil i vurderinga av generell intelligens som prediktor for seinare jobbprestasjonar. McClelland hevdar at andre ikkje-intellektuelle variablar som utdanning og motivasjon er betre prediktorar enn generelle kognitive evner (intelligens). Dessutan stiller han spørsmål ved korleis ein måler ein person sin intelligens (McClelland, 1972). Ree og Earles (1993) trekkjer fram Sternberg og Wagner som legg vekt på *tacit knowledge* (kunnskap som er vanskeleg å overføra mellom menneske, skriftleg eller munnleg) som ein betre prediktor for jobbprestasjon enn psykometrisk målt *g*. Dette avviser Ree og Earles (1993) med å visa til at Sternberg, Wagner og McClelland har basert sine konklusjonar på for små utval ( $N = 29 - 104$ ) og at funna deira difor kan skuldast tilfeldige utvalsfeil (*sampling error*). Sjølv viser dei til at eigne funn knytt til *g* som prediktor for jobbprestasjonar er sikrare, då dei kan visa til stor  $N$  (1550 – 78000). Trass i dette hevdar Lohman (1997) at så lenge ein er ueinig om val av målemetode og at det eksisterar ulikskapar i den sosiale-, politiske- og personelege kontekst, vil det òg eksistera ueinigheit om korleis menneskelege evner eigentleg er satt saman.

Eg vil difor diskutera vidare moglege årsaker til at Gr.1 ikkje ser ut til å henga saman med ENJJPT-prestasjonar medan AE gjer det. AE korrelerar signifikant både med gjennomsnittleg teoriresultat ( $r = .18$ ) og T-38 dagleg flyging ( $r = .21$ ).

Gr.1 korrelerer svakt med AE. Dette kan tyda på at AE har målt noko anna enn Gr.1. Kanskje AE har målt meir av andre spesifikke evner (*s*) som har betre prediktiv validitet for flygarprestasjonar? Viss det er tilfelle kan det vera ei moglege forklaringar til at AE i større grad korrelerar med ENJJPT-prestasjonar enn det Gr.1 gjer. Basert på resultatata så langt kan ein kanskje seia at AE inneheldt andre testar enn reine mål på generell intelligens (i alle fall i periodar) som òg har målt en del spesifikke evner som var relevante for flygarprestasjonar? Er det i tillegg mogeleg at Gr.1 måler meir resonneringsevne medan AE måler meir generell intelligens? Slik dei to testane i testgruppe 1 er skildra kan ein undrast over det, men det fins ikkje belegg for ein slik påstand her. Summert kan det verka som at ein her har indikasjonar på at AE og Gr.1 har målt noko ulike sider av intelligens. I tillegg har testane hatt ulik vanskegrad.

Det ligg dessutan ei mogeleg forklaring i den strenge selekteringa på Gr.1. Omlag 10 % av søkjarane kalla inn til opptak vart tekne opp som flyaspirantar og omlag halparten av dei fekk

tilbud om flygarutdanning, der ca. halvparten av dei att fekk tilbud om ENJJPT-utdanning. Under slike forhold vert det særskild lite variasjon att i gruppa i forhold til Gr.1 testskår og flygarprestasjonsskår. Av den grunn lyt korrelasjon med kriteriet verta avgrensa. Diskusjonen kan oppsummerast med å seia at RR lyt ta ein del av skulda for tilsynelatande fråvær av samanheng for testgruppe 1. Deltakarane sine høge skår på testgruppe 1 kan dessutan ha verka inn på ein måte som har gjort at korrelasjonar ikkje har dukka opp som forventa. For AE var situasjonen noko ulik. Kandidatane vart ikkje direkte selektert på AE slik dei vart det basert på Gr.1-resultata (jfr. vanlegvis krav om skår over 5 for denne testgruppa, og den var normert for LFS-søkjjarar) og i datasettet finn ein AE skår spreidd over nær heile skalaen (3 – 9). AE ser dermed ut til å vera mindre råka av streng selektering og RR enn Gr.1.

Diskusjon knytt til Gr. 2 – 4 og flygarprestasjon:

Mål på spesifikke evner går gjerne under omgrepet *measures of spesific abilities* (Cronbach, 1970). Resultat frå forskning på militær flygarseleksjon gjennom fleire tiår har vist at utdanninga til- og jobben som militær flygar, ser ut til å krevja spesifikke evner som presisjon og hurtighet i høve psykomotoriske koordineringsevner, spatiale evner og diverse flyrelatert kunnskap/evner (t.d. instrumenttyding, mekanisk/teknisk innsikt og forståing av slike konsept) (Carretta & Ree, 2003; Hunter & Burke, 1994; Martinussen, 1997; Martinussen & Torjussen, 1998; Martinussen & Hunter, 2008). Testgruppe 2 – 4 vert oppfatta som ei samling av evnetestar som søker å dekkja desse områda.

Testgruppe 2 – 4 korrelerar alle signifikant med T-37 og T-38 *Daily Flying* ( $r = .20$  til  $r = .27$ ). I tillegg predikerte Gr.3 Simultankapasitet teorireultatet med  $r = .26$ . Gruppe 2 og 3 hadde høgast korrelasjon med T-38 flyginga ( $r = .26$  og  $r = .27$ ). Gr.4 hadde litt svakare korrelasjon. Dette resultatet er ikkje overraskande. Likevel korrelerte heller ikkje desse prediktorane med prestasjon på eksamensflyginga (*T-37* og *T-38 Checkride*).

Carretta og Ree (1994) undersøkte den prediktive validiteten til United States Air Force sitt flygarseleksjonssystem (PCSM) for utdanningsresultat ved UPT/SUPT-flyskulane og fann at observerte korrelasjonar (rundt  $r = .20$ ) kan vera redusert frå ein sann korrelasjon på nærare .60. Dette skuldast stor RR, og difor er det viktig å ikkje konkludera på bakgrunn av dei observerte korrelasjonane ein får fram i analysane. Dei sanne korrelasjonane ligg *alltid* høgare enn dei observerte. Det er eit sentralt poeng som òg gjeld for resultatata i denne studien. Carretta og Ree (1994) understreka at dei har gjort tidlegare funn som indikerar at testar som måler meir spesifikke evner er lada med *g*. Undersøkjinga viste at testar som måler spesifikke evner, som til dømes psykomotoriske og spatiale evner, bidreg lite i høve tilleggsvaliditet utover det papir-og

blyanttestane gjer åleine. Mine funn indikerar det motsatte. Her bidrar Gr.1 lite eller ingenting i høve prediktiv validitet samanlikna med testgruppe 2 til 4. Samtidig er det viktig å understreka at Gr.2 - 4 ser ut til å måla generell intelligens i tillegg til dei meir spesifikke evnene. Store studiar ( $N = 5500$ ) har demonstrert at dess høgare generell intelligens ( $g$ ) lading ein test har dess meir valid er han som prediktor i høve flygar *training success* (Ree & Earles, 1992). I høve gjeldande faktor-analytisk teori er kognitive evner organisert på ein hierarkisk måte. Dette legg Ree og Carretta (1997) vekt på. Evnetestar måler ikkje berre spesifikke evner, som til dømes spatiale evner, dei måler òg ein  $g$ -faktor (Martinussen, 1997).

Meta-analysar knytt til flygarseleksjon (Hunter & Burke, 1994; Martinussen, 1997) har vist at testar som måler generell intelligens har hatt lågare prediktiv validitet enn testar som måler kognitive evner, der mål på  $g$  inngår saman med andre meir spesifikke evner. Schmidt og Hunter (1998) trekkjer fram at  $g$  i kombinasjon med andre mål (*work sample, personal integrity, og structured interview*) hadde høgast prediktiv validitet for flygarprestasjon. Carretta og Ree (2003) hevdar at ein kvar seleksjonstest, eller mål på ei evne, måler både  $g$  og ei spesifikk evne eller spesifikk kunnskap. Olea og Ree (1994) legg likevel vekt på at der er  $g$  som er den beste prediktoren og at dei spesifikke evnene bidreg lite, noko Carretta og Ree (2003) støttar. Denne  $g$ -faktoren kan jamvel estimerast (Ree & Earles, 1992). Wheeler og Ree (1997) fann at prediktiv validitet for USAF sin psykomotoriske test (BAT) for flygarseleksjon kom både frå ein  $g$ -komponent og ein psykomotorisk komponent.

Tabell 2 indikerar at Gr.2 og 3 kan vera lada med  $g$  då korrelasjon mellom Gr.1 og Gr.2 / 3 er  $r = .34$  og  $r = .29$ , høvesvis. Ein kunne undersøkjia dette nærare ved å gjennomføra ein faktoranalyse på testresultat til heile søkargruppa. Det vart vurdert til å liggja utanfor problemstillinga.

Oppsummering av diskusjon i høve problemstilling 1:

Sjølv om denne studien ikkje kan visa til prediktiv validitet for generell intelligens målt ved Gr.1-testane for flygarprestasjonar under utdanninga bør ein likevel ikkje underslå betydninga av denne testgruppa. Det kan vera slik at desse kandidatane har hatt ein Gr.1-skår som har vore betre enn det utdanninga krevde, og at det fylgjeleg er ei forklaringa på at ein ikkje har korrelasjon med prestasjonskriteria. For å få fram ein slik samanheng måtte ein kanskje akseptert kandidatar med lågare Gr.1-skår og forsøkt å utdanna dei ved ENJJPT. Viss dei med lågare skår vart teke ut av programmet oftare enn dei med betre skår, eller at dei jamnt over hadde svakare prestasjonar, ville det kunna bidra til at ein samanheng dukka opp. Det kan òg vera at utvalet sin storleik er for liten i forhold til den statistiske styrken ein treng for å kunna avdekka så pass små



korrelasjonar, sidan utvalet si samansetning er så avgrensa av RR. Cook og Ward (1996) trekkjer dessutan fram at medan nokre flygarevner kan trenast opp er det truleg ikkje gjeldande for evner knytt til intelligens. Er det tilfelle vil kandidatane vera «låst» til eit intelligensnivå og det kan difor vera interessant for ein organisasjon å velgja dei med gode evner òg innanfor dette domenet.

Gr.2 – 4 viste prediktiv validitet for flygarprestasjonar. Det er antyda at kandidatane sin *g*-skår kan ha verka inn på desse testresultata. Det vert i alle høve hevda at dei fleste testar av denne typen (*aptitude tests*<sup>13</sup>) måler meir enn berre ei einskild evne (Ree & Carretta, 1997).

Hunter og Burke (1995) føreslår å gå pragmatisk til verks når det gjeld diskusjonen om *g*-faktoren si betydning i høve flygarseleksjon. Sjølv om det fins stønad for at *g*-faktoren godt kan vera *det* mest effektive psykometriske mål ein kan ta i bruk for seleksjonsføremål, bør ein heller vurderer ein miks av ulike gruppefaktorar for å oppnå ein mest mogleg valid og rettvis flygarseleksjon. I og med at den einskilte flygarkandidat møter opp til seleksjonsprosessen med ulik bakgrunn og erfaring, vil ei breiare tilnærming opna for betre balanse mellom dei ulike seleksjonskriteria. Ei slik tilnærming vil òg gjera seleksjonsprosessen lettare å forsvara utad, utanfor den psykometriske arenaen seier Hunter og Burke.

Diskusjon knytt til prediktiv validitet for teoriresultat:

Det er signifikant samheng mellom AE / Gr.3 Simultankapasitet og gjennomsnittleg teoriresultat. Dette resultatet er vanskeleg å tolka. Kva er det Gr.3-testane måler som kan bidra til å forklara at dei som skårar høgt her òg skårar betre enn andre på teori prøvane? Gr.3 var prega av psykomotoriske evnetestar, men òg evna til å konsentrere seg slik at ein kan reagere raskt og nøyaktig under tidspress (Hansen et al., 1989; Hansen, 2006). I tillegg inngjekk Tallsøker i denne gruppa. Denne testen korrelerte òg signifikant med teorikriteriet ( $r = .20$ ). Tallsøker-testen vart assosiert med evnene til korttidshukommelse, rask dømmekraft, høgt tempo og effektivitet. Slike evner ser ut til å slå positivt ut for prestasjonar på fleirvalsprøvar i flyteori ved ENJJPT. Dette går på tvers av resultatet frå ei stor amerikansk undersøkjing frå UPT/SUPT-utdanninga ( $N = 7563$ ) der generell intelligens (*g*) hadde prediktiv validitet for teoriresultatet (Carretta & Ree, 1995). Ei mogleg forklaring kan vera knytt til at dei norske flyelevane vart selektert strengare og på ein annan måte enn dei amerikanske elevane. Ein indikasjon på dette finn ein ved å samanlikna seleksjonsflygingsprogramma i dei to landa. Amerikanske flyaspirantar flaug 23 timar i eit einmotors propellfly med instruktør. Her oppnådde 12 % ikkje-bestått medan dei andre

<sup>13</sup> An aptitude test is one intended to predict success in some occupation or training course (Cronbach, 1970)

kandidatane gjekk vidare til UPT/SUPT-utdanning (Weeks, Zelenski, & Carretta, 1996). På LFS vart om lag 50 % av flyaspirantane dimitert i løpet av seleksjonsflyginga med Saab Safari med om lag same timetal.

Diskusjon knytt til «*Checkride*-problemet»:

Studiane til Carretta (1992a,  $N = 755$ ), Carretta (2005,  $N = 994$ ), Carretta og Ree (1994,  $N = 678$ ), Carretta og Ree (1995,  $N = 7563$ ) og Carretta (2011,  $N = 883$ ) oppnådde signifikante korrelasjonar mellom evnetestar og flygarprestasjon, både i forhold til *Academics*, *Daily Flying* og *Checkride* resultat for amerikanske UPT/SUPT flyelevar. I denne studien er det ikkje påvist samanheng mellom prediktorane AE / Gr. 1 – 4 og ENJJPT *T-37* og *T-38 Checkride* karaktergjennomsnittet til norske elevar.

Eg ynskjer å presentera nokre perspektiv som i hovudsak er basert på egne erfaringar i rolla som instruktør, Checkpilot og instruktørevaluator ved den multi-nasjonale flyskulen ENJJPT.

For det fyrste kan mellommenneskelege forhold mellom elev og Checkpilot ha påverka prestasjonane. For ein flyelev er det å bli evaluert av ein Checkpilot ein over-gjennomsnittleg pressa situasjon. Evaluatoren sit rett bak/ved sida og vurderar taust alt eleven gjer, medan dei høyrer andedrettet til kvarandre gjennom interkommunikasjonsanlegget. Vart den einskilde norske elev sine *checkride*-prestasjonar forstyrra av dette vil det kunna vera ei mogeleg forklaring på avviket knytt til eksamensflyging. I tillegg kan det tenkjast at det mellom Checkpilot og elev kan oppstå spente situasjonar knytt til evalueringa og risiko for å ikkje bestå, noko som kan påverka eleven sine prestasjonar under flyturen. På den andre sida verkar det urimeleg at stressande og spente situasjonar knytt til eksamensflyging skulle påverka dei norske elevane sine prestasjonar meir negativt enn deira amerikanske medelevar.

At praktisk talt alle norske elevar klarar å gjennomføra utdanninga, medan andre nasjonar har vesentleg lågare gjennomføringsrate er ei kjent sak. Det kan mogelegvis påverka korleis ein Checkpilot angreip evalueringssituasjonen. I ein slik evalueringsetting kan ein tenkja seg til at nokre Checkpilotar ser for seg ei enkel og grei evaluering med lite etterarbeid, då den norske eleven mest sannsynleg kom til å fly ein grei tur og bestå eksamen. Ein kan òg tenkja seg at evaluatarar såg på situasjonen meir som eit høve til å prøva ut kandidaten. Ei «er du så god som dette ryktet skal ha det til?» -haldning kan bidra til at eleven vert prøvd hardare enn normalt. Begge desse tilnærmingane kan føra til at eleven sine prestasjonar vart påverka på ein måte som

gjorde at prestasjonsvurderingane og karakterane vart tildelt på ein mindre systematisk måte enn det 100 % objektive vurderingar kunne ha gjort.

Halo-effekt påverkar korleis personar vurderar kvarandre. Slike vurderingar er vanskelege å stola på. Hovudinstrykket av ein person ser ut til å påverka alle andre vurderingar som vert gjort slik at ein ikkje får eit objektivt bilete av dei ulike eigenskapane ein person eigentleg har (Thorndike, 1920). For å unngå at slike uynskte mellommenneskelege faktorar skal påverka resultatet er ein avhengig av at organisasjonen meistrar å standardisera Checkpilotane sine vurderingar av flygarprestasjon slik at det ikkje vert rom for subjektive oppfatningar der halo-effektar kan verka inn på ein forstyrrande måte.

Interkorrelasjon på  $r = .19$  mellom *Daily Flying Average* og *Checkflight Average* er lågt. Carretta (1995) kan visa til interkorrelasjon for amerikanske elevar på  $r = .41$  og  $r = .64$ . Dette kan tyda på at dei amerikanske elevane har handtert, eller vore i ein situasjon som har ført til at avviklinga av flyeksamen har vorte gjennomført på ein måte som betre reflekterer dei daglege flygarpresentasjonane enn det dei norske erfarte.

For det andre kan karaktersystemet, og svak standardisering i høve dette systemet, ha påverka eksamensresultatet. Eit element frå metodedelen kan vera relevant i høve denne diskusjonen. Karakter oppnådd på fyrste manøverforsøk vart ståande som teljande, uavhengig av korleis eleven presterte på kontinuasjonseksamen. Som eg nemnde i metodedelen vil dette karaktersystemet kunna opna for at ein elev som stryk ein *checkride*, men består forsøk nummer to, kan oppnå betre skår enn ein elev som består på fyrste forsøk. Dette kan opplevast unaturleg. Enkelte checkpilotar kan difor verta freista til å trekkja ned karakterar på manøvrar som er prestert mellom karakteren E og G (*eit nedtrekk på ein manøver frå E til G er omtala som ein downgrade* og dess fleire *downgrades* dess mindre poeng kan oppnåast) for ein eksamenskandidat som har flydd til stryk. Dette kan vera ein faktor både i høve det å legitimera at flyturen vart vurdert til ein stryk (understreka med ytterlegare *downgrades*) men òg for å bidra til at ein kandidat som stryk får ein skår som betre står i forhold til eit ikkje bestått resultat. Såleis vil det kunna bidra til at flyelevar som faktisk står på fyrst forsøk sjeldnare får mindre poeng enn kandidatar som stryk. Dessutan kan halo-effektar spela inn og påverka ein checkpilot si vurdering av ein ikkje-bestått eksamen. Dette gjeld spesielt viss eleven i det store og heile presterte veldig bra men var uheldig og gjorde ein tabbe. Eit slikt karaktersystem er altså særskilt avhengig av at vurdering av flygarprestasjon på eksamen er fri for subjektivitet.

Elevane si eiga oppleving av karaktersystemet kan òg vera ei medverkande årsak til at ein ikkje finn samanheng mellom prediktorane og ENJJPT-eksamensflyging. Til dømes vil det

kunna vera slik at ein amerikansk student vil ha eit incentiv om å få færrest mogleg *downgrades* då ein slik situasjon kan resultera i eit poengmessig større trekk enn ein stryk som skuldast *ein* stor feil. Systemet er lagt opp slik at poeng oppnådd på *checkrides* tel mest og færre poeng fører til lågare rangering. Dess lågare rangering dess lengre bak i køen kjem ein i høve tilbod om ei jagarflykarriere etter utdanninga. Samstundes vil det for norske elevar vera mest interessant å bestå eksamen, uavhengig av kor mange *downgrades* ein måtte få. På den måten vil ein kunna unngå å ta eksamen på nytt. Alle elevane var innforstått med at ein konsekvens av å stryka kontinuasjonseksamen markerte starten på ein elimineringsprosess. Dette kan ha ført til at norske elevar ville fokusera på å gjera alle manøvrar så sikkert og trygt som mogleg slik at resultatet totalt sett vart godt nok til å bestå. Elevar som satte omsynet til rangering i klassen høgast kan tenkja motsatt og velgja ein annan strategi. Eit omsyn til rangering kan stimulera til å prøva å unngå *downgrades*. I tilfelle stryk kan eleven likevel oppnå ein høgare poengsum totalt sett enn ein medelev som valde ein eksamensstrategi som endte i fleire poengtrekk. Ulike incentiv og motivasjon for eksamensprestasjon kan difor potensielt ha påverka resultatet, i allehøve i den grad eleven sjølv hadde innverknad.

Sjølv om ENJJPT har lagt opp til eit standardiseringsregime for å unngå subjektive vurderingar av elevprestasjonar, vil ikkje ein slik innsats føra til 100 % objektivitet. Fleire utfordringar er knytt til slik innsats. Mellom anna vart Checkpilotar heile tida skifta ut i høve rutinemessige personellrotasjonar ved skulen. Frekvente rotasjonar kan vera utfordrande for kontinuitet og organisatorisk minne. Ein kan gå ut frå at Checkpilotane hadde ulike personlege eigenskapar og individuelle trekk som kunne verka inn på evalueringssituasjonen. Ulike behov og personlege preferansar kan gjera sitt til at subjektivitet kan koma inn og forstyrre personvurderingar og såleis påverka datagrunnlaget. Eg har sjølv erfart at instruktørar ofte har lik oppfatning ved vurdering av om ein manøver er utført på ein trygg måte (*safe*) eller ikkje. Slike vurderingar vert gjort meir likt enn vurderingar knytt til om ein spesifikk flymanøver på ein *Checkride* er *Good* eller *Excellent* (til dømes ei innflyging og landing med mykje vind og anna samtidig flytrafikk). Sjølv om regelverket gjev støtte til slike vurderingar er det likevel rom for skjønn. Vurderinga om ein manøver var *G* eller *E* var i det store og heile ikkje viktig å treffa korrekt på kvar gong. Begge karakterane førte uansett fram til bestått eksamen men *E* resulterte i fleire poeng. Det same kan ikkje seiast om vurderingar knytt til kva som er trygg og sikker flyging. Under krevjande flysituasjonar får flygaren kanskje berre *ein* sjanse til å gjera ei korrekt vurdering. Slike vurderingar er dermed viktigare å vera einige om mellom checkpilotane. Samstundes er det talet på *dowgrades* som i størst grad er med på å påverka total poengskår og

rangering i klassen for den einskilde flyelev. Dei tilhøva som her er peika på kan påverka kor systematisk og objektiv karaktergjevinga på flyeksamen har vore. Viss subjektive vurderingar har fått spelerom kan reliabiliteten til dette prestasjonskriteriet ha vorte svekka.

For det tredje er det godt mogleg at andre ytre faktorar i større grad kan spela inn på prestasjonsresultatet for *checkride* enn dagleg flyging med instruktør. Medan ein akademisk eksamen på høgskule eller universitet vanlegvis føregår på ein nøytral arena utan ytre forstyrrende element, føregjekk *checkrides* i releativt tett samkvem med evaluator. I tillegg vart det flydd under skiftande vèr- og vindforhold og nye trafikksituasjonar oppstod heile tida. Eksamensflyginga kan vera spesielt utsatt for slike varierende forhold, då eleven normalt ikkje vil motta assistanse i høve handtering av nye og uvante situasjonar i lufta. På ein *checkride* bør eleven i utgangspunktet meistra nye og ukjente situasjonar åleine. Det er likevel opna for at Checkpiloten kan assistera i situasjonar ein åpenbart ikkje vil kunna forventa at eleven skal meistra. Kva type assistanse som vert tilbudt kan variera frå Checkpilot til Checkpilot. Språkproblem kan spela inn og føra til ekstra friksjon, noko som igjen kan føra til misforståing mellom elev og evaluator. Terskelen for å spurja evaluatoren repetitivt for å oppklara ei mogeleg misforståing kan dessutan vera høgare enn ved dagleg flyging saman med ein instruktør som eleven kjenner godt. Slike ytre faktorar kan tenkast å påverka prestasjonane slik at flydugleik som er tydeleg i dagleg flyging i mindre grad kjem til syne på eksamen.

For det fjerde kan det vera for få deltakarar representert i studien. I tillegg er standardavviket for *checkride* (og *academics*) kriteria vesentleg mindre enn SD for *Daily flying*-kriteria. Det er difor ei mogelegheit at variasjonen i desse prestasjonskriteria er så avgrensa at det vert utfordrande å finna samsvar. Nokre svake, ikkje signifikante, positive korrelasjonar kjem fram i analysen ( $r = .07$  -  $r = .10$ ). Med større  $N$  kunne desse korrelasjonane som no framstår som tilfeldige potensielt vorte sikrare, slik at sjølv små samanhengar ikkje ville vorte forkasta. Korrigert for RR ville dessutan korrelasjonane vorte sterkare.

Det har no vore diskutert fleire aspekt med eksamensflyginga ved ENJJPT. Desse aspekta kan vera mogelege forklaringar til at ein observerer ulik prediktiv validitet for prediktorane i høve dagleg flyging og eksamensflyging. Fleire av desse faktorane kunne åleine, eller i samspel, verka inn på at vurderingar av flygarprestasjonar vart meir usystematisk utført enn det ideelt sett burde vore. *Checkride*-resultata kan ha vorte meir tilfeldige enn det resultata av dagleg instruksjonsflyging ser ut til å ha vorte. Kva *checkride*-kriteriet i denne studien eigentleg har målt er ikkje mogeleg å seia sikkert, men kriteriet var meint å måla flygarprestasjonar på ein objektiv og systematisk måte. Skilnaden mellom resultata frå store amerikanske studiar og denne

er tydeleg og fleire moglege forklaringar er lansert i diskusjonen over. Nokre moglege skilnader mellom amerikanske og norske aspekt knytt til rangering under utdanninga er òg diskutert. Dessutan skil ENJJPT seg ut ved å vera ein multinasjonal organisasjon der ulike kulturar og språk møtes. Om lag halvparten av elevar og instruktørar/checkpilotar er frå europeiske land. Dei ulike kulturelle aspekta, både mellom amerikansk og europeisk kultur, men òg mellom ulike europeiske kulturar, kan jamvel verka inn på måtar ein her ikkje meistrar å skildra på ein fullgod måte. Det kulturelle aspektet vert likevel trukke fram som ein faktor i høve denne diskusjonen.

Avslutningsvis ynskjer eg å trekkja fram ei mogleg årsak til dette avviket. Norske flyelevar vart selektert på ein annan måte enn amerikanske flyelevar. I Noreg nytta ein psykologintervju som eit supplement til evneprøvene, medan ein i USA ofte nytta uttale om kandidaten si *Officership* som eit viktig avgjerdsgrunnlag i seleksjonsprosessen. Vektinga av denne eigenskapen ved ein kandidat vart gjort utan at dette kunne visa prediktiv validitet for flygarprestasjonar (Carretta, 2000). Etter å ha jobba ved ENJJPT nokre år er mi erfaring at amerikanarane, på alle nivå i organisasjonen, trekkjer fram godt *Officership* som ein av dei viktigaste eigenskapane i høve dagleg virke i deira militære organisasjon. Dette kjem tydeleg til uttrykk mellom anna i dei årvisse tenesteuttalelsane deira (*Officer Performance Reports*). Korleis ein eventuell høg skår på *Officership* vil kunna påverka eksamensresultat kan ein ikkje seia nøyaktig. Ein kan likevel ikkje sjå heilt bort frå at denne faktoren på eit eller annan vis kan ha påverka samspelet mellom evaluator og elev i *checkride*-situasjonen, trass i at evalueringa var meint å berre vurdera flygarprestasjonar. Oppsummert kan ein kanskje seia at ei mogleg tolking av resultatet er at *checkride*-kriteriet av ymse grunnar kan ha svakare reliabilitet for dei norske elevane enn for dei amerikanske.

Lesarar som er interessert i temaet og ynskjer å sjå meir detaljerte analyseresultat og vidare diskusjon av desse, vil finna meir om dette i vedlegga. Utdanningsresultata ved ENJJPT er brote vidare ned i flydisiplinane instrument-, kontakt- og formasjonsflyging. I vedlegg A finn ein korrelasjonsresultata for prediktorar som korrelerte med T-37 og T-38 *Daily Flying* knytt til dei ulike flydisiplinane.

### 6.3 Diskusjon av problemstilling 3

*Kor godt predikerar testgruppene (Gr. 1 – 4) samla sett dei ulike prestasjonskriteria ved ENJJPT?*

Gjennomsnittresultatet av testgruppe 1 – 4 korrelerte kun med *T-37-* og *T-38 Daily Flying* med høvesvis  $r = .20$  og  $r = .25$ . I den fyrste regresjonsanalysen kjem det fram at Gr.2 og Gr.3 utgjer eit hovudbidrag i høve forklart varians for den avhengige variabelen *T-38 Daily Flying*. Alle fire gruppene kan i statistisk forstand til saman forklara 13 % av variansen.

Regresjonsresultatet ser ut til å avvika frå anna forskning. Carretta og Ree (2003) har presentert regresjonsanalysar som antyder at generell intelligens ( $g$ ), ein faktor som vert hevda ligg høgare opp i evnehierarkiet hjå den einskilde kandidat, kan utgjera heile 67 % av forklart varians for flygarprestasjonar under utdanning. Andre spesifikke evner forklarte 33 %, herunder 9 % for flyrelatert kunnskap (teknisk innsikt og instrumenttyding) og 4 % for spatiale evner). I denne studien var ikkje apparatustestar knytt til psykomotoriske evner teke med. Analysen kan difor ikkje jamførast direkte med regresjonsanalysane i denne studien, men her er like fullt ein indikasjon på avvik i høve Gr.1 og forklart varians for *T-38* prestasjon.

I den andre regresjonsanalysen var *T-37 Daily Flying* den avhengige variabelen. Igjen var Gr.2 og Gr.3 dei stega i modellen som bidrog mest til forklart varians. Modellen kunne samla sett forklara 9 % av dei daglege *T-37* prestasjonane.

I den tredje regresjonsanalysen vart modellen undersøkt mot gjennomsnittleg teorireultat for *T-37* og *T-38*. Denne modellen var òg signifikant, om enn på eit lågare nivå (05-nivå). I denne regresjonen var Gr.3 den einaste som signifikant bidrog til ein forklart varians på 8 %.

Ikkje overraskande viste den siste regresjonsanalysen, med gjennomsittleg *Checkride*-resultat for *T-37* og *T-38* samla, at ingen av testgruppene kunne bidra til forklart varians. Modellen som heilskap var ikkje signifikant.

Dei statistiske utrekningane for korrelasjonsanalysar og regresjonsanalysar føregår på ulikt vis og det er difor viktig å gjennomføra begge analysane for å avdekka fleire aspekt ved valideringsstudiar som denne. I høve *checkride*-kriteriet har ein såleis undersøkt fleire aspekt og kan med det seie, med større tryggleik, at den prediktive validiteten for dei fire prediktorane er fråverande for dette kriteriet.

Samla sett har regresjonsanalysane demonstrert at sjølv om testgruppene har prediktiv validitet for flygarprestasjon under utdanninga, har variablane avgrensa forklaringskraft i høve sentrale prestasjonskriterier totalt sett. Analysane avdekka at Gr.1 – 4 kun kan bidra med opp til

13 % av forklart varians i T-38 prestasjonskriteriet. Det er difor andre faktorar som òg spelar inn i høve forklaring av variasjon i desse prestasjonskriteria. Hunter og Burke (1995) hevdar at testbatteri ein har sett til no i beste fall kan forklara opp mot 50 % av variasjonen i flygarprestasjon og at ein difor har mykje å gå på i høve det å forbetra evnetestane.

#### 6.4 Om forskingsspørsmålet

*I kva grad er det samanheng mellom resultatane oppnådd på seleksjonstestane og prestasjonane under flygarutdanninga ved ENJJPT?*

Studien indikerar at det er varierende samanheng mellom seleksjonsprediktorane og flygarprestasjonskriteria. Ein finn prediktiv validitet for testgruppe 2 – 4 for T-37 og T-38 *Daily Flying* prestasjon. Dette stemmer godt overeins med funn frå andre studiar utført på denne tida.

Studien har avdekka at det tilsyneleatande ikkje er signifikant samanheng mellom Gr.1 testane som er meint å måla generell intelligens og ENJJPT-prestasjonar. Dette var noko overraskande i starten av arbeidet, men moglege forklaringar er diskutert. Det kan tenkjast at deltakarane jamnt over har hatt betre skår på dette intelligensmålet enn det som har vore påkrevd for å meistra utdanninga. AE-målet ser derimot ut til å ha prediktiv validitet for både ENJJPT teori og flyging med T-38. Dessutan ser det ut til at *g* og andre meir spesifikke evner er vevd saman, slik at kandidatar med høg *g* òg skårar høgt på testar som måler sentrale evner knytt til teknisk innsikt, spatiale evner og psykomotoriske evner. I tillegg er det prediktiv validitet for Gr.3 Simultankapasitet i høve teorireultat.

Carretta og Ree (2003) rapporterte at kandidatar som skåra høgt på USAF sitt seleksjonssystem (PCSM), som våre eigne seleksjonstestar elles kan samanliknast med, hadde større sannsynlegheit for å gjennomføra UPT/SUPT-utdanninga. I tillegg nytta dei færre flytimar, oppnådde høgare rangering i klassen og hadde med det større sannsynlegheit for å få tilbod om utdanning på jagarfly etterpå.

Analysane har avdekka at testgruppene 2 – 4 har prediktiv validitet for flygarprestasjonar ved ENJJPT i høve dagleg instruksjonsflyging. Det er rimeleg å anta at desse prestasjonane påverka kandidatane si sannsynlegheit for å fullføra ENJJPT-utdanninga. Det er vidare rimeleg å anta at kandidatar som skåra høgt på desse testgruppe har forbrukt mindre flytimar, då dei har prestert betre enn dei med lågare skår. Betre prestasjonar vert her assosiert med behovet for ekstra-turar for å oppnå tilfredsstillande flygarprestasjon og progresjon i utdanninga. Ein studie av UPT/SUPT-utdanninga stadfester denne samanhengen ( $N = 1082$ ) (Duke & Ree, 1996).



Forholdet mellom evner og vilje/motivasjon til å gjennomføra utdanninga er diskutert i liten grad. Studiar har vist at personlegheitstrekk som til dømes sterk vilje til å prestera, kan spela inn på flygarprestasjonane (Carretta & Ree, 2003). Schmidt og Hunter (1998) fann at eit mål for *conscientiousness* (som mellom anna inkluderar prestasjonsvilje, pliktoppfylling og sjølvdisiplin) som går inn i personlegheitstestar som måler *Big Five* kombinert med *g*, hadde prediktiv validitet for utdanning og seinare jobbprestasjon. Trass i at slike personlege eigenskapar ikkje er ein del av studien lyt det nemnast. Det er rimeleg å anta at me alle har både positive og negative erfaringar knytt til korleis motivasjon og vilje kan påverka prestasjonane våre til dagleg. Jamført med regresjonsanalysen, der T-38 *Daily Flying* var den avhengige variabelen (tabell 3), kunne prediktorane samla sett berre forklara 13 % av forklart varians. Åpenbart er det andre faktorar som òg spelar inn og bidreg til den registrerte variasjonen i dette prestasjonskriteriet. Det kan difor vera ein grunn til at ein har inkludert andre prediktorar, som til dømes personlegheitstestar og intervju med psykolog, som ein del av flygarseleksjonssystemet.

Samstundes som det her vert indikert samanhengar mellom seleksjonstestar og flygarprestasjon under utdanning lyt ein òg trekkja fram motrøyster. Cook og Ward (1996) legg fram forskning som indikerar at det kan vera grunn til å tvila på diverse flygartestbatteri sine prediktive kapasitetar. Årsakene kan vera knytt til at testane ikkje er pålitelege nok, at me har mangelfull forståing av kva behov flygaren har for dei ulike evnene og at ein har for liten innsikt i korleis jobben til flygaren vert utført. Damos (1996) hevda at det store fleirtalet av militære seleksjonstestbatteri er designa nettopp for å predikera suksess i *utdanninga*. Ho kunne knapt finna studiar som undersøkte slike testar sin prediktive validitet for operativ flygarprestasjon, som trass alt er eit viktigare kriterium. Etter ein gjennomgang av flygarseleksjonslitteraturen fann Damos (2011) at det var to evner som såg ut til å vera nærast kopla til det å lukkast som militær flygar. Det var god oppfatningsevne (*perceptual speed*) og spatiale evner.

Den positive samanhengen som er svart på i høve forskingsspørsmålet i denne studien er ikkje uviktig sjølv om operativ flygarprestasjon ikkje har vore eit kriterium. Fyrste steg i ein flygarkarriere er trass alt å fullføra flygarutdanninga. Dessutan har det vist seg særskilt krevjande å etablere reliable operative prestasjonskriteria (Damos, 1996). Eit problem med slike analysar er mellom anna korleis ein skal nytta operative prestasjonsmål som er knytt til utføring av jobben i krig eller væpna konflikt. Talet på kor mange fly ein flygar har skote ned, eller kor mange gonger han sjølv har unngått å bli skote ned eller angripen på anna vis, er sentrale mål. Det er likevel så pass mange andre faktorar som spelar inn på utkomet av kampflyoperasjonar at det ville vera umogeleg å føra kontroll med alt. Samstundes trekkjer Cook og Ward (1996) fram at

sjølv om det har vist seg at til dømes psykomotoriske evner er ein viktig eigenskap gjennom heile flygarkarriera, er denne evna kanskje ikkje like sentral i alle aspekt i utøvinga av jobben som jagarflygar. Eigenskapar knytt til planlegging, gjennomføring og avgjerdstaking i høve deltaking eller leiing av kompliserte kampflyoperasjonar vert etter kvart vel så viktig som reine flygarevner.

Då Howse og Damos (2011) undersøkte kva flygarevner som har vist seg som dei mest sentrale for å lukkast som militær flygar, understreka dei at det her var fleire utfordringar knytt til det å definera desse flygarevnene. I tillegg dukka andre metodologiske utfordringar opp undervegs som gjorde sitt til at berre avgrensa konklusjonar kunne trekkast opp. Mellom anna vart konklusjonen deira negativt påverka av mangelfulle teoriar knytt til flygarpersonlegheit som kunne geleida dei.

### **Styrker og svakheiter ved studien**

Den største svakheita med studien er at den undersøker kandidatlar som er strengt selektert. Dessutan har seleksjonen føregått gjennom fleire fasar over relativt lang tid. Dette gjer sitt til at andre faktorar kan koma inn og påverka prestasjonane. Undersøkjinga her er gjort utan å inkludera kandidatane som falt frå undervegs i seleksjonsprosessen. Til dømes kunne resultata til dei andre som òg kom inn på LFS som flyaspirantar, men som etter kvart falt frå under seleksjonsflyginga, vorte inkludert for å gje eit endå betre bilete av seleksjonssystemet si prediktive validitet.

Denne situasjonen har ført til at studien vart hardt råka av avgrensing knytt til *Range Restriction*. Problemet med RR aukar dessutan for kvart seleksjonsledd (Hunter & Burke, 1995). Eg har valgt å ikkje henta inn data frå fleire kandidatlar enn dei 175, RR har difor ført til at ein heile tida har vore tvungen til å arbeida med observerte korrelasjonar. Sidan slike korrelasjonar ligg lågare enn dei faktiske, har det vore umogeleg å skildra kor sterk prediktiv validitet prediktorane eigentleg har hatt. Difor lyt ein her nøya seg med å seia at det er evidens for prediktiv validitet, men at det er uvisst kor sterk den er. Eit mindre strengt seleksjonssystem hadde ført til at fleire kandidatlar kunne prøvd seg på flygarutdanninga. Det er nok likevel ikkje realisme i denne tilnærminga. Eit slikt seleksjonssystem ville dessutan vore særskilt kostbart, mellom anna på grunn av større fare for auka fråfall under utdanninga i USA. Metoden kunne gjeve eit meir korrekt bilete av seleksjonssystemet si prediktive validitet, men det er truleg ikkje verdt innsatsen. Så lenge undersøkjingar konsekvent støttar opp om eit seleksjonssystem sin relevans vil det ikkje vera behov for ytterlegare styrking av den prediktive validiteten, då systemet ser ut til å verka i praksis.

Utover dette er det òg fleire andre metodiske utfordringar ein møter på i høve gjennomføring av lokale valideringsstudiar. Valet av kva variablar ein skal nytta som prediktorar og val av gode kriteria er heilt sentralt. Både prediktor og kriterium lyt ha tilfredsstillande reliabilitet og validitet i tillegg til at dei helst bør vera relevante for den aktuelle oranisasjonen. Dessutan er ein avhengig av å ha tilstrekkeleg stor  $N$ .

For å bøta på nokre av desse utfordringane kan meta-analysar nyttast blant anna for å auka  $N$ . Samstundes kan meta-analysar òg vera hefta med utfordringar knytt til reliabilitet. Denne analysevarianten som samlar valideringstudiar kan kritiseras for å pressa saman ulike studiar som ein miks av «epler og pærer» av prediktorar som faktisk har lite til felles (Hunter & Burke, 1995).

Utfordringar knytt til reliabilitet og validitet som kan ha påverka resultatet i denne studien negativt er:

1. Det burde kanskje vore tilgang på større  $N$  for å avdekka faktisk prediktiv validitet for seleksjonssystemet. Amerikanske studiar av UPT/SUPT-utdanning eg har hatt tilgang til har operert med  $N = 478 - 7563$ . Carretta og Ree (2003) trekkjer fram at dess strengare seleksjon som har føregått og dess mindre variablane er til å stola på, dess større utval vert ein avhengig av. Basert på den strenge seleksjonen som her har føregått kan det ha ført til at  $N$  vart for liten til å avdekka heile bilete.

Det kunne vore mogleg å auka  $N$ . Det ville samstundes kunna gå på bekostning av validitet og reliabilitet i studien. For å oppnå større  $N$  måtte kandidatar som hadde gjennomført databaserte testar (i fase 2) vorte innlemma. Trass i at denne type testar var meint å måla dei same eigenskapane som dei gamle apparatustestane, ville eit slikt val kunna påverka validitet og reliabilitet til studien. Ved å inkludera kandidatar frå tida før *1980-batteriet* vart implementert ville nye spørsmål knytt til reliabilitet og validitet òg kunna oppstå. Det ville i tillegg vore mogleg å inkludera kandidatar som hadde gjennomført utdanninga etter 2008 trass i at desse flyelevane flaug ein annan flytype (T-6 erstatta T-37 ved ENJJPT rundt år 2008). Ei slik utviding ville òg truleg bidrege til lågare reliabilitet og ny usikkerheit for resultatene. Basert på dette vart eit datagrunnlag med  $N = 175$  valt.

2. Usikkerheit knytt til validitet og reliabilitet ved dei valde ENJJPT- kriteria, spesielt i høve risiko for subjektivitet i vurdering av flygarprestasjonar. Det er heilt avgjerande at instruktørane sine vurderingar av elevane sine prestasjonar føregår på ein systematisk og påliteleg måte slik at ein unngår ei heil rekkje kjende feilkjelder knytt til slike personvurderingar (Martinussen & Hunter, 2008). Dette gjeld både dagleg flyging og eksamensflyging.

Flyinstruktører som har oppdaga at ein elev har det som skal til kan ( gjerne omtala som at ein flyelev har *hands*<sup>14</sup>), i samkvem med dei andre instruktørane som flyg med eleven til dagleg, sjå mellom fingrane på mindre avvik (avvik som ikkje gjeld trygg og sikker flyging) der eleven kan ha vore uheldig, eller har hamna i ein uvant situasjon, og likevel gje eleven gode nok karakterar til å gå vidare til neste treningsbolc. Slike avvik kan ein kanskje akseptera sidan instruktørane fylgjer elevane sine prestasjonar kvar dag og såleis er i tett oppfylgjing med kontroll på den einskilde sin progresjon.

Det er funne ut at viss ein person er god til å gjera *ein* ting vert dei andre oppgåvene personen utfører ofte vurdert meir positivt. Dette vert omtala som ein slags *smitteeffekt* (Martinussen & Hunter, 2008). Syversen (1960) trekkjer i sin studie av utdanning av jagarflygarar i Canada på 50-talet fram at vurdering av elevane sine prestasjonar i flyet, både av vanlege instruktører og Checkpilotar kan vera påverka av hyppige instruktørskifter, den individuelle instruktøren si *frame of reference* og ein halo-effekt. I høve *checkride*-bedømminga kan i tillegg sjølv evalueringssettinga, med sine kunstige aspekt samanlikna med dagleg instruksjonsflyging, vera krevjande å kontrollera for ein organsisasjon.

Validitet og reliabilitet i forhold til eksamensflyging kan vera utsatt for faktorar diskutert tidlegare, som kan ha påverka kor stabilt og standardisert dette kriteriet var på denne tida. Det er vanskeleg å undersøkje dette kriteriet sin sanne reliabilitet. Ein flyelev får ikkje gjennomført same flyeksamen med same evaluator to gonger der begge startar evalueringa frå identisk nullpunkt og der flyginga i tillegg føregår under like tilhøve. I kva grad eksamensflyturane målte flygarprestasjon er diskutert allereie. Fleire faktorar kan spela inn og gjera sitt til at det her vert målt òg andre aspekt med kandidaten sine prestasjonar, som til dømes meistring av stress, grad av flekisibilitet i høve det å meistra nye og krevjande situasjonar og korleis ein best handterar ein spent evalueringssituasjon med eit ukjent menneske.

Sjølv om storleiken på korrelasjonane mellom variablar vert avgrensa av reliabiliteten (Carretta & Ree, 2003) reduserer ikkje kriteriereliasiteten korrelasjonane i den grad som RR gjer (samtale med Martinussen, feb 2013). Carretta og Ree (1995) hevdar at korreksjon for RR og manglande kriteriereliasitet kan vera i størrelseorden  $r = .25$  til  $r = .58$ . Denne studien kan ikkje antyda kor store korreksjonar som her kunne vore aktuelle, anna enn å visa til studiar det er mogeleg å gjera samanlikningar med. Det kunne vore interessant å undersøkje kriteriereliasitet noko meir i høve *checkride*-kriteriet.

<sup>14</sup> Med omgrepet *hands* meines her at kandidaten flyg på ein behageleg og presis måte. Flyparameteravvik vert korrigerast raskt og naturleg. Flygarsjargong.

3. Studien har ikkje teke omsyn til at det har gått ei viss tid frå kandidatane gjennomførte opptaksprøvene fram til dei var ferdig utdanna flygarar. I løpet av denne tida gjekk kandidatane gjennom befalsskule (viss befalskule ikkje allereie var gjennomført), seleksjonsflyging med Saab Safari og Luftkrigsskule I. I løpet av denne tida har den einskilde kandidat mognast og utvikla seg vidare både i høve motivasjon for yrket og andre tilhøve i livet. Denne tidsfaktoren kan vera relevant for studien, men er valt bort fordi slike data ikkje var tilgjengelege.

4. Standardavviket for *checkride* og *academics* -kriteria er vesentleg mindre enn SD for *Daily flying*-kriteria. Det er difor ei moglegeheit at variasjonen i kriteriet er så avgrensa at det kan ha bidratt til at samanhengar med prediktorane ikkje dukka opp.

5. Studien kunne analysert AE vidare. AE viste prediktiv validitet for både teoriresultat og flyging med T-38. Flygarprestasjonane på T-38 hadde mest å seia for rangering og utdanningsresultatet. Ein kunne difor valgt å inkludera dette intelligensmålet i regresjonsanalysane. Likevel vart dette valgt bort då kandidatane ikkje vart selektert utifrå AE, anna enn indirekte, men utifrå Gr.1 primært.

Nokre moglege styrkar med studien:

1. Dei faktorane som har bidratt mest til reliabilitet i denne studien er at alle 175 kandidatane gjennomførte like opptaksprøvar. I tillegg gjekk dei gjennom det same flyprogrammet ved ENJJPT. Mindre justeringar av seleksjonstestane og ENJJPT- *Syllabus* undervegs er vurdert til ikkje å utgjera nokon praktisk skilnad for resultatata. Dette kan vera ein styrke for resultatata i studien. For å oppnå høg reliabilitet i ein studie som denne lyt ein oppnå det same resultatet kvar gong. Kven som helst kan gå gjennom datamaterialet som her er nytta og få same analyseresultatata.

2. Ein styrke med studien er at det ikkje vert nytta eit todelt kriterium (bestått/ikkje bestått utdanning). Då ein har nytta meir reelle kriterium for flygarprestasjon under utdanninga aukar det relevansen til studien. Ideelt sett skulle ein nytta kriterium frå operativ flyging med F-16. Likevel ligg dei valde kriteria nærare opp til prestasjonskriteria frå operativ flyging enn det seleksjonsflygingskriteria ved LFS gjer. Ei utfordring ved bruk av kriterium som er henta ei tid etter at seleksjonstestane vart gjennomført, er samstundes at andre faktorar i større grad kan ha verka inn på prestasjonsdata etter kvart som tida går og kandidatane utviklar seg vidare.

3. *N* har vore tilstrekkeleg stor slik at prediktiv validet for dagleg flyging, både på T-37 og T-38 vart avdekka.

4. USAF vurderte ein strykprosent rundt 10 % i høve sjølve UPT/SUPT-flygarutdanninga som akseptabel (Carretta, 2000). På 90-talet låg strykprosenten knytt til flygarprestasjon for USAF stabilt på 22 % over fleire år (Carretta, 1997). For Luftforsvaret, som for tida utdannar rundt seks nye flygar ved ENJJPT i året, ville ein tilsvarande strykprosent truleg vera uakseptabel. Ein styrke med studien er at ein har fått inn offisielle data frå hovudkvarteret til AETC i USA som det er god grunn til å lita på. Med ein komplett oversikt over alle norske flyelevar som starta på ENJJPT-utdanninga i denne perioden, der både dei som fullførte og dei som vart sendt heim er oppført (inkludert årsaka til elimineringa), har studien avdekket at strykprosenten gjennom desse 15 åra var 1 %.

5. Ein styrke med prediktorane var at det vart nytta standardiserte testar. Standardiserte testar er laga slik at same prosedyre, apparat/utstyr og skåring vert nytta uavhengig av tid og stad (Cronbach, 1970). Dessutan er AE-målet sin reliabilitet og validitet godt dokumentert (Sundet et al., 2004). Ein styrke med teorikriteriet er at avviklinga av teorieksamen ved ENJJPT (fleirvalgsmetoden) var lagt opp etter ein standardisert mal. Syvertsen (1960) trekkjer det å nytta skriftege standardiserte prøvar fram som ein faktor som kan bidra til styrka kriteriereliasitet.

6. Utdanninga ved ENJJPT var regulert og kontrollert av HQ AETC. Kontrollen føregjekk på ein detaljert måte, og avvik vart rapportert til hovudkvarteret sentralt. Kvar einaste flymanøver som skulle vurderast på T-37 og T-38 var nedskriven med nøyaktige skildringar (Maneuver Item File - MIF) på kva som var godkjent og ikkje godkjent (jfr. karakterane U, F, G og E). Utdanninga vart kontrollert jamnleg av eigne inspeksjonsteam utsendt frå HQ AETC. Dei var på utkikk etter avvik mellom anna i høve standardiseringa mellom dei ulike flyskvadronar og underavdelingane (*flights*). Den interne standardiseringsinnsatsen på elev-evalueringsavdelinga (*Checksection*), Instruktørutdanningsavdelinga (*PIT*) og Instruktørevalueringsavdelinga (*Stan EVAL*) vart òg kontrollert.

Flyskvadronane ved ENJJPT gjennomførte i tillegg eigne rutinemessige standardiseringsmøter/samlingar mellom inspeksjonane utført av AETC HQ. Denne standardiseringsinnsatsen bør ein rekna med var eit forsøk på å bidra til høg reliabilitet ved ENJJPT prestasjonskriteria, både for å sikra at dei rette individane vart utdanna til flygarar, men òg at utdanninga føregjekk på ein rettviss og juridisk haldbar måte. AETC omtalar desse kriteria som objektive (U. S. Air Force, 2003), men om det verkeleg var tilfelle har eg ikkje evidens til å seia noko om. For at ein prøve skal vera objektiv lyt kvar observatør (eller dommar) som ser prestasjonen koma til identisk resultat i vurderingsrapporten (Cronbach, 1970). Den svake korrelasjonen ( $r = .19$ ) mellom daglege flygarprestasjonar og eksamensflyging på T-38 som

ikkje er avgrensa av RR (då alle 175 norske kom gjennom programmet), kan tyda på at ein her kan ha vore utsatt for faktorar som har påverka eksamensresultatet. For T-37 var korrelasjonen større ( $r = .37$ ). Carretta og Ree (1995) kunne visa til korrelasjonar mellom *Daily Flying Average* og *Check Flight Average* for UPT/SUPT utdanning med T-37 og T-38 på høvesvis  $r = .41$  og  $r = .64$  (observert). Forskjellen mellom amerikanske og norske korrelasjonar er størst i høve flyging med T-38 og såleis skil ENJJPT seg frå dei andre UPT/SUPT-skulane. På den andre sida kan standardiseringsinnsatsen, trass spørsmål knytt til kor påliteleg vurderingane av elevane sine prestasjonar til ein kvar tid har vore, likevel seiast å ha vore forsøkt gjennomført på ein systematisk måte gjennom detaljerte retningslinjer og kontrollrutinar. Denne innsatsen vert difor likevel trekt fram som ein mogleg styrke ved kriteria som er valt.

## 7 Konklusjon

Føremålet med studien var å undersøkje i kva grad opptaksprøvene til LFS hadde prediktiv validitet for seinare flygarprestasjonar. Resultata oppnådd på desse prøvane vart korrelerte med utdanningsresultata for dei som gjennomførte jagarflygarutdanninga ved ENJJPT mellom 1989 – 2004 ( $N = 175$ ). Hovudprediktorane var testgruppene generell intelligens, teknisk innsikt og spatiale evner, psykomotoriske evner og simultankapasitet samt overføring/retningsevne. Testresultata innan desse evneområda var avgjerande for om ein kandidat skulle takast opp som flyaspirant ved LFS eller ikkje, og dei vart difor nytta som prediktorar i studien. I tillegg vart Alminnelig evnenivå (AE) frå sesjonstestane nytta som prediktor.

Prestasjonskriteria var gjennomsnittleg resultat frå teorieksamen, dagleg flyging og eksamensflyging på T-37 og T-38 oppnådd gjennom den 13 månader lange utdanninga. Prestasjonskriteria var avgjerande for rangering av den einskilde flyelev, og alle måtte prestera over minstemål knytt til desse kriteria for å fullføra utdanninga.

Innanfor flygarseleksjonsdomenet har ein opp gjennom tida vore på konstant søk etter å finna empirisk evidens for at det fins måtar å testa flygarkandidatar på som gjev tilfredstillande prediktiv validitet til at eit seleksjonssystem vil vera betre i bruk framfor eit anna. Trass i at ein ikkje finn sikre bevis for at slike samanhengar eksisterar er det funne evidens som har gjeve god grunn til å halda fast ved systemet og gjera vedtak om personellseleksjon basert på det. For norske tilhøve er det i denne studien vist at grupper av testar, og nokre individuelle testar, kan visa betre prediktiv validitet enn andre i høve flygarutdanning ved ENJJPT. Testgruppene knytta til evner innan teknisk innsikt, romforståelse, psykomotoriske evner/simultankapasitet og retningsevner viste prediktiv validitet for flygarprestasjon med T-37 og T-38 i høve dagleg instruksjonsflyging. AE viste prediktiv validitet for teoriresultat og dagleg flyging med T-38. Ingen prediktorar hadde samanheng med eksamensflyginga ved ENJJPT. Dette er eit avvik i høve anna forskning på UPT/SUPT-flygarutdanning, og fleire perspektiv er diskutert for å kunna tilby moglege forklaringar på dette resultatet.

Studiar av nyare dato som har søkt å akkumulera forskingsresultat har hatt vanskar med å finne eintydige konklusjonar på kva kunnskapar, evner, ferdigheitar med meir ein flygarkandiat eigentleg lyt ha for å lukkast. I arbeidet med dette har forskaren støtt på utfordringar både i høve definisjonar og teori knytt til kva evner flygaren treng. Ein har likevel funne at nokre evner og ferdigheiter er nærare knytt til det å lukkast med militær flygarutdanning enn andre (Damos, 2011; Howse & Damos, 2011) nemleg oppfattingsevne og spatiale evner.



I 2013 har ikkje innhaldet i militær flygarseleksjon endra seg vesentleg frå det som er skildra i denne studien. Kognitive evner som har vore sett på som særskild viktige for å lukkast i høve utdanning til flygaryrket er framleis spatiale evner, evne til logisk tenking og resonnering samt god tal- og ordforståing (Veldhuis, Matton, & Vautier, 2012). Paullin, Katz, Bruskiwicz, Houston, og Damos (2006) trekkjer fram at forskingsresultata til no har vist klare teikn til at generelle kognitive evner, eller generell intelligens (*g*), er ein viktig prediktor for flygarprestasjonar. Samstundes meiner dei det er grunn til å hevda at nokre tilleggsmål òg har betydning: gode psykomotoriske evner, oppfattningsevne, kortidsminne, interesse for flyging, flyerfaring og visse personlegheitstrekk. Fråværet av prediktiv validitet for *g* (Gr.1) i denne studien bør ikkje føra til ei svekking av denne evna i høve nasjonal flygarseleksjon. Ulike årsaker til at Gr.1 som prediktor står fram i kontrast med AE er diskutert. Ein har slutta at kandidatane har skåra høgt på desse testane og at *g* i noko grad har verka positivt inn på skåren oppnådd på dei andre testgruppene.

Seleksjonsprosessen som bidrog til at 175 av 177 kandidatar graduerte ved ENJJPT i perioden 1989 – 2004 ser ut til å ha vore ein suksess i forhold til utdanning av jagarflygarar. Studien har vist at opptaksprøvene hadde prediktiv validitet for elevane sine daglege flygarprestasjonar, både for T-37 og T-38, men for T-38 spesielt. Det var ved dagleg flyging med instruktør at kandidatane lærte å fly jetfly. Kunnskap og ferdigheita opparbeidde i løpet av dagleg flyging har vore ein viktig forutsetnad for å kunna bestå *checkrides*. Trass i at dagleg flyging korrelerer svakt med *checkrides* er det like fullt ein signifikant samanheng mellom desse flygarprestasjonane. Den manglande prediktive validiteten opptaksprøvene hadde i høve *checkride*-prestasjonar kan tyda på at måten desse praktiske flyeksamenane vert gjennomført på, saman med andre forhold, har påverkar resultatet meir enn det evnene til dei norske flyelevane har gjort. Faktorar som motivasjon, viljen til å prestera på eksamen, standardiseringsutfordringar og andre ytre faktorar som vær og uvante situasjonar i lufta, samt ein mogleg halo-effekt er diskutert som moglege årsaker. Ei mogleg svekka kriteriereliasitet for *checkride*-kriteriet har ikkje påverka kva flytype dei norske elevane skulle få tilbod om etter fullført utdanning.

Alt i alt viser resultatata av studien at Forsvaret hadde ein suksessrate på rundt 99 % i høve seleksjon og utdanning av nye jagarflygarar gjennom ein periode på 15 år. Dette resultatet kjem ikkje som ei fylgje av skår oppnådd på testane i gruppe 1 – 4 åleine. I tillegg til desse måla vart kvar og ein av dei 175 kandidatane utsatt for tappetest og intervju med psykolog, DMT-test, opptaksråd/uttakskommisjon og sist men ikkje minst, eit seleksjonsflygingsprogram ved LFS.

Prestasjonar på seleksjonstestane som fell utanom Gr.1 – 4 er ikkje ein del av studien, men desse resultatane kan òg ha vore faktorar som kan verka inn utdanningsresultatet. Spesifikke resultat oppnådd ved LFS, anna enn at alle 175 hadde bestått, var heller ikkje ein del av denne studien. Det høyrer likevel med til historia at dei vurderingane som instruktørane ved LFS gjorde undervegs, mellom anna vurderingar og vedtak som leia fram til klassifisering av den einskilde flyelev i kategoriane helikopter-, multi- og jagarflyutdanning, òg kan ha bidrege til ENJJPT-resultatet.

### 7.1 Praktisk verknad

Studien har avdekkja signifikante samanhengar mellom ulike evner og flygarprestasjonar når det gjeld dagleg flyging, samstundes som det er avdekkja fråver av samheng mot prestasjon under flyeksamen. Prestasjon under evalueringspress er viktig for flygaren fordi han/ho støtt skal evaluerast for å oppretthalda gyldige sertifikat. Desse sertifikatane med tilhørande rettigheit er ferskvare, då det er tale om avgrensa gyldigheitsperiodar. Prestasjon på *checkrides* er ein viktig faktor i vurderinga av den samla dugleiken til flygaren. Ynskjer ein å undersøkje *checkride*-prestasjonar for norske elevar vidare, vil det vera mogeleg. Det oppleves likevel ikkje som eit reelt problem, då me uansett ser ut til å få flygarane gjennom utdanninga. Viss *checkride*-resultata av ei eller anna årsak skulle tillegkast meir vekt i framtida, til dømes i høve utvelging av kandidatar som skal gå direkte frå ENJJPT til F-35, eller av andre grunnar, vil meir kunnskap om kva som påverkar *checkride*-resultatet for norske elevar vera viktig å henta inn meir informasjon om.

Forsking på flygarseleksjon lyt fortsetja då innhaldet i jobben som militær jagarflygar endrar seg over tid. Dei evnesetta som verka godt for flygarar i Luftforsvaret på til dømes F-104 og F-16 kan skilja seg frå evnebehovet for F-35 i framtida. Her tenkjer eg ikkje på evnene til det å fly på ein trygg og god måte, men meir på perspektivet som ser på flygaren som operatør på ein våpenplattform. Teknologisk utvikling har ført fram meir sofistikerte og komplekse våpensystem som har lagt større krav på flygaren (NATO, 1996). I dette perspektivet har flygaren eit potensiale til å gjera meir enn det ein kunne med eldre flytypar. Kva rolle F-35 skal ha i eit nymoderande nettverksbasert Forsvar veit me trass alt ikkje mykje om om enno. For å få meir kunnskap aspekta knytt til seleksjon for F-35, kan det vera hensiktsmessig å starta opp ein jobbanalyse for å finna ut kva oppgåver ein F-35 flygar lyt meistra på dei ulike oppdrag.

Flygarseleksjonen i Forsvaret fortset å basera seg på dei evnene som kjem fram i denne studien. Ein nyttar difor testar som er meint å måla desse på ein reliabel og valid måte. Dei manuelle apparatustestane vart etterkvart erstatta av databaserte testar (Torjussen og Hansen,

1999). Den prediktive validiteten for datatestane kan òg vera interessant å undersøkje nærare i høve ENJJPT-prestasjonar. Saman med denne studien kan ein då få høve til å samanlikna prediktiv validitet for dagens testbatteri mot *1980-batteriet*. Avhengig av kva ein slik studie hadde avdekka kunne det oppstå ein fornya diskusjon omkring flygarseleksjon og testbatteria sin samansetninga frametter for optimal effektivitet i seleksjonsarbeidet i Forsvaret.

Då det er påvist prediktiv validitet for testgruppe 2 – 4 og AE for ENJJPT flygarprestasjon vil det vera av interesse å vurdere å inkludera desse måla i høve LFS sine anbefalingar knytt til klassifisering av aspirantar som består LFS. Dette føreset ei antaking om at det å fly *fast jet* krev meir av desse evnene enn anna flygarutdanninga. Det vert hevda at krava som vert satt til jagarflygaren skil seg frå krav satt til fleirmotor- eller tankflyflygarar og at det difor er viktig å finna fram til dei kandidatane som eignar seg best til dei ulike flysystema (Strickland & Duke, 1996). Viss det er slik at ENJJPT med utdanning på jetfly er den mest krevjande utdanninga i høve flygarprestasjon, kan ei slik tilnærming auka sannsynlegheita for at kandidatane oppnår gode resultat på dagleg flyging.

Denne studien har bidrege med evidens som gjer sitt til at det framleis er god grunn til å meina at seleksjonssystemet på 80- og 90-talet har god prediktiv validitet for flygarprestasjonar generelt, men no for ENJJPT-prestasjonar i tillegg. Den suksessfulle gjennomføringsratio er spesielt trekt fram. Informasjon om kandidatane sine evneprofilar gjev høve til betre matching med jobben dei skal utføra, betre kvalitet på flyelevane, mindre sløsing med ressursar og ikkje minst betre kampkraft (Weeks et al., 1996). Det kan difor vera fornuftig å ta små og forsiktige steg viss Forsvaret skulle ynskja å endra seleksjonssystemet for flygarar. Ei eventuell endring lyt i så fall validerast, og evidens for prediktiv validitet lyt leggjast fram før eit nytt system bør nyttast som avgjerdsgrunnlag i ein framtidig seleksjonsprosess (Linn, Baker, & Dunbar, 1991).

## 8 Referansar

- Bailey, M., & Woodhead, R. (1996). I: NATO AGARD-CP-588, *Selection and training advances in aviation* (s. 8-1-9). Neuilly-Sur-Seine: Advisory Group for Aerospace Research & Development.
- Blower, D. J., & Dolgin, D. L. (1990). An evaluation of performance-based tests designed to predict success in primary flight training. *Proceedings of the Human Factors Society 34<sup>th</sup> Annual Meeting, 13*, 949-953.
- Burke, E. F. (1993). Pilot selection in NATO: An overview. I: R. S. Jensen & D. Neumeister (Red.), *Proceedings of the Seventh International Symposium on Aviation Psychology* (s. 373-378). Columbus: The Ohio State University.
- Carretta, T. R. (1987). *Spatial ability as a predictor of flight training performance* (AFHRL Tech. Paper No. 86-70). Brooks Air Force Base, TX: Manpower and Personnel Division, Air Force Human Research Laboratory.
- Carretta, T. R. (1989). USAF pilot selection and classification systems. *Aviation, Space, and Environmental Medicine, 60*, 46-49.
- Carretta, T. R. (1992a). Recent developments in U.S. Air Force pilot candidate selection and classification. *Aviation, Space, and Environmental Medicine, 63*, 1112-1114.
- Carretta, T. R. (1992b). Understanding the relations between selection factors and pilot training performance: Does the criterion make a difference? *International Journal of Aviation Psychology, 2*, 95-105.
- Carretta, T. R. (1997). Group differences on USAF pilot selection tests. *International Journal of Selection & Assessment, 5*, 115-127.
- Carretta, T. R. (2000). U. S. Air Force pilot selection and training methods. *Aviation, Space, and Environmental Medicine, 71*, 950-956.
- Carretta, T. R. (2005). *Development and validation of the test of Basic Aviation Skills (TBAS)* (AFHRL Tech. Report No. 2005-0172). Wright Patterson Air Force Base, OH: Human Effectiveness Directorate.
- Carretta, T. R. (2011). Pilot candidate selection method: Still an effective predictor of US Air Force pilot training performance. *Aviation Psychology and Applied Human Factors, 1*, 3-8.

- 
- Carretta, T. R., & Ree, M. J. (1994). Pilot candidate selection method: Sources of validity. *International Journal of Aviation Psychology, 4*, 103-117.
- Carretta, T. R., & Ree, M. J. (1995). Air Force Officer Qualifying Test validity for predicting pilot training performance. *Journal of Business and Psychology, 9*, 379-388.
- Carretta, T. R., & Ree, M. J. (2003). Pilot selection methods. I: P. S. Tsang & M. A. Vidulich (Red.), *Principles and practice of aviation psychology* (s. 357-396). Mahwah, NJ: Lawrence Erlbaum Associates.
- Carretta, T. R., Rodgers, M. N., & Hansen, I. (1993). *The identification of ability requirements and selection instruments for fighter pilot training*. (Interim Tech. Pap. AL/HR-TP-1993-0016). Brooks Air Force Base, TX: Armstrong Laboratory.
- Cobe, A. W. (1996). Fighter pilot – a moving constant. I: NATO AGARD-CP-588, *Selection and training advances in aviation* (s. K 1-8). Neuilly-Sur-Seine: Advisory Group for Aerospace Research & Development.
- Cook, M., & Ward, G. (1996). Understanding the requirement: A review of common problems in training, selection and design. I: NATO AGARD-CP-588, *Selection and training advances in aviation* (s. 21-1-19). Neuilly-Sur-Seine: Advisory Group for Aerospace Research & Development.
- Cronbach, L. J. (1970). *Essentials of psychological testing*. New York: Harper & Row.
- Damos, D. L. (1993). Using meta-analysis to compare the predictive validity of single- and multi-task measures to flight performance. *Human Factors, 35*, 615-628.
- Damos, D. L. (1996). Pilot Selection Batteries: Shortcomings and Perspectives. *International Journal of Aviation Psychology, 6*, 199-209.
- Damos, D. L. (2011). *KSAOs for military pilot selection: A review of the literature*. (Technical Report No. AFCAPS-FR-2011-0003). Randolph Air Force Base, TX: Air Force Personnel Center.
- De Kock, F., & Schlechter, A. (2009). Fluid intelligence and spatial reasoning as predictors of pilot training performance in the South African Air Force (SAAF). *SA Journal of Industrial Psychology, 35*, 1-8.
- Duke, A. P., & Ree, M. J. (1996). Better candidates fly fewer training hours: Another time testing pays off. *International Journal of Selection and Assessment, 4*, 115-121.

- Dunbar, S. B., & Linn, R. L. (1991). Range restriction adjustments in the prediction of military job performance. I: A. K. Widgor & B. F. Green (Red.), *Performance assessment for the workplace* (Volume II). Washington, DC: National Academy Press.
- Fiske, D. W. (1947). Validation of naval aviation cadet selection tests against training criteria. *Journal of Applied Psychology*, 31, 601-614.
- Forsvaret (1989). *Psykologiske undersøkelser ved flygerskoleuttak*. (Brev datert 11. august 1988). Oslo Mil/Akershus: Forsvarets Psykologiske og Pedagogiske senter.
- Forsvaret (2013, februar 8). Kva skjer på sesjonsdagen? Lokalisert 7. mars 2013 på Verdsveven: <http://forsvaret.no/verneplikt/sesjon/sesjon-del-2/Sider/sesjonsdagen.aspx>.
- Griffin, G. R., & Koonce, J. M. (1996). Review of psychomotor skills in pilot selection research of the US military services. *International Journal of Aviation Psychology*, 6, 125-147.
- Hansen, I. (1987). Psykologiske metoder ved flygeruttak. Del A; Beskrivelse av testene. *Militærpsykologiske Meddelelser*, 15. Oslo: Forsvarets psykologiske og pedagogiske senter.
- Hansen, I. (2006). Bidrag til psykologitjenestens historie i Forsvaret fra 1946-2006. *Militærpsykologiske Meddelelser*, 25. Oslo: Forsvarets Sjefpsykolog, FIL og FSS.
- Hansen, I., Heim, J., Nordby, K., Storsve, O., og Værnes, R. (1989). Psykologiske metoder ved flygeruttak. Del B; Utprøving av nytt testbatteri. *Militærpsykologiske Meddelelser*, 16. Oslo: Forsvarets psykologiske og pedagogiske senter.
- Henmon, V. A. C. (1919). Air service tests of aptitude for flying. *Journal of Applied Psychology*, 3, 103-109.
- Holoviak, S. J., Yauch, D. W., & Ercoline, W. R. (1996). I: NATO AGARD-CP-588, *Selection and training advances in aviation* (s. 32-1-7). Neuilly-Sur-Seine: Advisory Group for Aerospace Research & Development.
- Howse, W. R., & Damos, D. L. (2011). *Historical scientific analysis of aviator selection*. (Technical Report No. AFCAPS-FR-2011-0008). Randolph Air Force Base, TX: Air Force Personnel Center.
- Hunter, D. R., & Burke, E. F. (1994). Predicting aircraft pilot-training success: A meta-analysis of published research. *International Journal of Aviation Psychology*, 4, 297-313.

- 
- Hunter, D. R., & Burke, E. F. (1995). *Handbook of pilot selection*. Aldershot: Ashgate Publishing.
- Kjenstadbakk, T. J. (2012). Seleksjon til befalsskolen. *Masteroppgåve*. Oslo: Forsvarets Høgskole.
- Manning, T. A. (2007). *Major changes in undergraduate pilot training 1939 – 2002*. Randolph Air Force Base, TX: HQ AETC.
- Martinussen, M. (1989). Seleksjon av flygere. *Hovedfagsoppgave*. Oslo: Universitetet i Oslo.
- Martinussen, M. (1997). Pilot selection: A validation and meta-analysis of tests used for predicting pilot performance. *Doktorgradsavhandling*. Tromsø: Universitetet i Tromsø.
- Martinussen, M., & Torjussen, T. (1998). Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. *International Journal of Aviation Psychology*, 8, 33-45.
- Martinussen, M. & Hunter, D. (2008). *Luftfartpsykologi*. Bergen: Fagbokforlaget.
- McClelland, D. C. (1972). Do I.Q. tests measure intelligence? *The Saturday Evening Post*, Summer, 12-14.
- McClelland, D. C. (1993). Intelligence is not the best predictor of job performance. *Current Directions in Psychological Science*, 2, 5-6.
- North Atlantic Treaty Organization (1996). *Selection and training advances in aviation*. (Papers presented at the Aerospace Medical Panel Symposium held in Prague, in May 1996. AGARD-CP-588). France: Advisory Group for Aerospace Research & Development.
- Linn, R. L., Baker, E. L., & Dunbar, S. B. (1991). Complex, performance-based assessment: Expectations and validation criteria. *Educational Researcher*, 20, 15-21.
- Lohman, D. F. (1997). The history of intelligence testing in context: The impact of personal, religious, and scientific beliefs on the development of theories and tests of human abilities. I: R. F. Dillon (Red.), *Handbook on testing* (s. 82-106). Westport: Greenwood Press.
- Olea, M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology*, 79, 845-851.
- Paullin, C., Katz, L., Bruskiwicz, K. T., Houston, J., & Damos, D. (2006). *Review of aviator selection*. (Interim Technical Report 1183). Arlington, VA: US Army Research Institute for Behavioral and Social Sciences.

- 
- Ree, M. J., & Carretta, T. R. (1997). What makes an aptitude test valid? I: R. F. Dillon (Red.), *Handbook on testing* (s. 65-81). Westport: Greenwood Press.
- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1, 86-89.
- Ree, M. J., & Earles, J. A. (1993). *g* is to psychology what carbon is to chemistry: A reply to Sternberg and Wagner, McClelland, and Calfee. *Current Directions in Psychological Science*, 2, 11-12.
- Ringdal, K. (2009). *Enhet og mangfold. Samfunnsvitenskaplig forskning og kvantitativ metode* (2. utg.). Bergen: Fagbokforlaget.
- Royal Canadian Air Force (2011, juni 7). Undergraduate pilot training system (UPTS). Lokalisert 28. februar 2013 på Verdsveven: <http://www.rcaf-arc.forces.gc.ca/cfts-fevs/page-eng.asp?id=1397>.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, 124, 262-274.
- Signori, E. I. (1949). The Arnprior Experiment: A study of WW II pilot selection procedures in the RCAF and RAF. *Canadian Journal of Psychology*, 3, 136-150.
- Standards for educational and psykologisk testing*. (1999). Washington DC: American Psychological Association.
- Strickland, W. J. & Duke, A. P. (1996). Technical evaluation report. I: NATO AGARD-CP-588, *Selection and training advances in aviation* (s. T 1-6). Neuilly-Sur-Seine: Advisory Group for Aerospace Research & Development.
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32, 349-362.
- Syversen, J. L. (1960). Psykologisk utvelgning av flygere II (Psychological selection of fighterpilots II). *Militærpsykologiske Meddelelser*, 5. Oslo: Forsvarets psykologiske avdeling.
- Tett, R. P., Jackson, D. N., & Rothstein, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology*, 44, 703-742.



- 
- Thorndike, E. L. (1920). A constant error in psychological ratings. *Journal of Applied Psychology*, 4, 25-29.
- Torjussen, T., & Hansen, I. (1999). Forsvaret, best i test. *Tidsskrift for Norsk Psykologiforening*, 36, 772-779.
- Trankell, A. (1959). The psychologist as an instrument of prediction. *Journal of Applied Psychology*, 43, 170-175.
- Turnbull, G. J. (1992). A review of military pilot selection. *Aviation, Space, and Environmental Medicine*, 63, 825-830.
- U. S. Air Force (2000). *Formal flying training administration and management* (AETC Instruction No. 36-2205 Interim Guidance). Randolph Air Force Base, TX: HQ Air Education and Training Command.
- U. S. Air Force (2001, oktober 31). Major changes in Undergraduate Pilot Training 1939-2002. Lokalisert 27. Februar 2013 på Verdsveven:  
<http://www.aetc.af.mil/shared/media/document/AFD-070130-081.pdf>
- U. S. Air Force (2002). *Flying Training - T-37/T-38 Euro-NATO Joint Jet Pilot Training* (AETC Syllabus P-V4A-N). Randolph Air Force Base, TX: HQ Air Education and Training Command.
- U. S. Air Force (2003). *Formal flying training administration and management* (AETC Instruction No. 36-2205). Randolph Air Force Base, TX: HQ Air Education and Training Command.
- U. S. Air Force (2008). *Plan of operation to the memorandum of understanding concerning participation in the Euro-NATO Joint Jet Pilot Training Program*. Washington DC: Department of the Air Force.
- U. S. Air Force (2009). *U. S. Air Force fact sheet – Air Education and Training Command*. Washington DC: Department of the Air Force.
- U. S. Air Force (2010a). *Formal flying training administration and management* (AETC Instruction No. 36-2205 Vol.1). Randolph Air Force Base, TX: HQ Air Education and Training Command.

- 
- U. S. Air Force (2010b). *Formal flying training administration and management -T-1, T-6, T-38. and UH/TH-1H* (AETC Instruction No. 36-2205 Vol. 4). Randolph Air Force Base, TX: HQ Air Education and Training Command.
- U. S. Air Force (2013a). Inside Sheppard Air Force Base. Lokalisert 27. februar 2013 på Verdsveven: <http://www.sheppard.af.mil/units/80thflyingtrainingwing/index.asp>.
- U. S. Air Force (2013b). Inside Sheppard Air Force Base. Lokalisert 27. februar 2013 på Verdsveven: <http://www.sheppard.af.mil/library/biographies/bio.asp?id=14698>.
- Veldhuis, M., Matton, N., & Vautier, S. (2012). Using item response theory to evaluate measurement precision of selection tests at the French pilot training. *International Journal of Aviation Psychology*, 22, 18-29.
- Want, R. L. (1962). The validity of tests in the selection of Air Force pilots. *Australian Journal of Psychology*, 14, 133-139.
- Weeks, J. L., Zelenski, W. E., & Carretta, T. R. (1996). Technical evaluation report. I: NATO AGARD-CP-588, *Selection and training advances in aviation* (s. 1-6-11). Neuilly-Sur-Seine: Advisory Group for Aerospace Research & Development.
- Wheeler, J. L., & Ree, M. J. (1997). The role of general and specific psychomotor tracking ability in validity. *International Journal of Selection and Assessment*, 5, 128-136.

**Vedlegg A:**Tabell 7 – Korrelasjonsanalyse for ulike *Daily Flying* flydisiplinar

Tabell 7 gjev ein oversikt over deskriptiv statistikk og korrelasjonar mellom signifikante prediktorar (som i tabell 2) nytta ved opptaket til LFS og underkriteria til *T-37-* og *T-38 Daily Flying*. Prediktoren Gr.1 Generelt evnenivå er teke ut av tabellen på grunn av manglande statistisk signifikans. I denne tabellen er kriteriet delt opp i flydisiplinar som vart nytta ved ENJJPT i tida 1989 - 2004. Desse underkriteria var *T-37-* og *T-38 Contact*, *T-37-* og *T-38 Instrument* og *T-37-* og *T-38 Formation*. Kriteriet *T-37-* og *T-38 Navigation* er teke ut av analysen, då sistnemnde hadde avgrensa verknad for rangeringa.

Tabell 7. Gjennomsnitt, standardavvik og bivariate korrelasjonar for prediktorar som korrelerte med *T-37-* og *T-38 Daily Flying* kriteria presentert i forhold til ulike *Daily flying* flydisiplinar (N = 147 - 175).

Variablar	M	SD	0	1	2	3	4	5	6	7	8
<b>Prediktorar</b>											
0. Almennelig evnenivå (AE)	7.5	1.04									
1. Tekn.Ins/Romopplev. (Gr.2)	6.9	1.01	.34***								
2. Sim.kapasitet (Gr.3)	6.9	0.92	.29***	.12							
3. Presisjon/Retn.evne (Gr.4)	6.1	1.05	.15	.27**	.28***						
<b>Daily Flying Kriterier</b>											
4. T-37 Contact	62.0	6.05	.11	.20*	.22**	.20*					
5. T-37 Formation	58.8	8.46	-.01	.16	.06	.14	.76***				
6. T-37 Instrument	66.4	8.14	.15	.19*	.29***	.24**	.87***	.75***			
7. T-38 Contact	56.6	9.91	.21**	.24**	.28***	.19*	.68***	.69***	.72***		
8. T-38 Formation	69.0	10.32	.17*	.24**	.26**	.20*	.69***	.72***	.75***	.91***	
9. T-38 Instrument	67.2	10.40	.22**	.26**	.26**	.19*	.68***	.69***	.75***	.91***	.91***

\*  $p < .05$ ; \*\*  $p < .01$ ; \*\*\*  $p < .001$  (to-hala)

I tabell 7 ser ein at prediktoren AE korrelerte signifikant med alle *T-38 Daily flying* underkriteria *Contact*, *Formation* og *Instrument* med høvesvis  $r = .21$ ,  $r = .17$  og  $r = .22$ .

Testgruppe 2 Teknisk innsikt/romopplevelse, Gr.3 Simultankapasitet og Gr.4 Presisjon/retningssevne korrelerte med fleire *T-37-* og *T-38 Daily Flying* underkriteria (flykategoriar). Gr.2 og 3 korrelerte noko høgare med T-38 i forhold til T-37, medan Gr.4 korrelerte jamnt både med T-37 og T-38. Desse korrelasjonane var statistisk signifikante med  $r$  i spennet  $r = .19$  til  $r = .29$ . Gr.1 er utelatt då prediktoren ikkje hadde samanheng med ENJJPT-kriterer (jfr. tabell 2). Gr.2 Simultankapasitet korrelerte noko høgare enn dei andre to prøveklassane både i forhold til T-37 og T-38. Gr.2 korrelerte høgast med kriteria *T-37 Instrument* og *T-38 Contact* ( $r = .29$  og  $r = .28$ ).

Tabell 7 viser vidare at ingen prediktorar korrelerte med kriteria *T-37 Formation*. Korrelasjonane her var svake og ikkje-signifikante. I forhold til *T-38 Formation* sub-kriteriet korrelerte alle prediktorane med  $r$  i spennet  $r = .17$  til  $r = .26$ .

Diskusjon av resultatata:

Her kjem det fram at Gr.2, 3 og 4 hadde prediktiv validitet for alle flydisiplinane unntatt *T-37 Formation* (*T-37 Navigation* berre med Gr.2 ( $r = .22$ ), men dette kriteriet vart utelete fordi prestasjonar innan denne disiplinen hadde liten innverknad på rangeringa ved ENJJPT).

Ei mogleg forklaring til avviket knytt til formasjon kan vera at det er ei fyrstegongsoppleving for flyelevane å fly i formasjon med ein medelev i det andre flyet. Prestasjonane til den einskilde kan ha vorte påverka av den andre eleven sine prestasjonar både i negativ og positiv retning. På denne måten kan moglege samanhengar mellom eleven sine Gr.2 og 3 evner, som har prediktiv validitet med dei andre flydisiplinane, verta vanskelegare å få fram i høve formasjonsflyging. Interkorrelasjonane mellom *T-37 Contact*, *Formation* og *Instrument* er svakare enn det som kjem fram innanfor T-38 disiplinane og er svakast i dei tilfella *Formation* er involvert. Dette kan styrka hypotesen om at T-37 formasjonsdisiplinen skil seg ut i høve den einskilde elev sine prestasjonar som nemnt ovanfor. Reliabiliteten til kriteriet kunne vorte betre viss kvar elev flaug mot ein instruktør. Det er ikkje hensiktsmessig då det ville krevja dubla ressursbruk.

AE hadde prediktiv validitet for *T-38 Contact*, *Formation* og *Instrument*. AE korrelerte òg med *T-38 Navigation* ( $r = .20$ ), altså like godt som med dei andre flydisiplinane. AE har med det prediktiv validitet for utdanning på *fast jet* i alle flydisiplinar. Dette er interessant, men det

kan vanskeleg forklarast i denne studien. Likevel kan ein slå fast at AE har relevans som seleksjonsprediktor for dagleg flyging med T-38 under jagarflygarutdanninga i alle flydisiplinar.

Interkorrelasjon mellom dei ulike flydisiplinane for T-38 var høg. Det betyr at flyelevane som gjorde det bra innan ein disiplin òg gjorde det bra innan dei andre disiplinane.