



FORSVARET
Forsvarets høgskole

Seleksjon til Kontroll og Varsling

Seleksjonssystemets prediktive validitet

Per Sigurd Dvergsdal

Masteroppgave
Forsvarets høgskole
Høst 2016

Forord

Å skrive en Masteroppgave er en utfordrende og lærerik prosess. Perioden jeg har jobbet med oppgaven har vært en tid med nedturer og oppturer. Noen dager føles det meste i oppgaven veldig riktig og bra, andre dager føles oppgaven ikke fullt så bra. Etter hvert som dagene går, er det likevel interessant å legge merke til, hvordan oppgaven til slutt blir et endelig produkt. Det kreves en viss struktur og målrettet arbeid for å komme i mål.

Studiet har gitt meg mulighet til å studere seleksjonsprosessen inn til Kontroll og Varslingssystemet i Luftforsvaret. Selv om oppgaven kan sies å ha et smalt fokus, har undersøkelsen utvidet horisonten min for hva en seleksjon handler om. Skal en seleksjon gjennomføres grundig, er det mye forberedelser både i forkant og etterkant som skal tenkes på. Jeg har også fått ett mer ydmykt forhold til forskning. Det ligger mye hardt arbeid bak de som legger ut forskningsartikler.

I forbindelse med Masteroppgaven ønsker jeg først og fremst å rette en stor takk til Monica Martinussen ved Universitet i Tromsø. Uten henne hadde jeg aldri kommet i mål, på den måten jeg har gjort. Hun sitter inne med utrolig mye kunnskap innenfor statistikk, og ikke minst innenfor Luftfartspsykologi. I tillegg må jeg takke Sjefpsykolog i Forsvaret, Ole Christian Lang-Ree for å hjelpe meg med datainnhenting. Jeg vil takke Luftforsvarets Seleksjonssenter med Per-Espen Ødegaard og hans kollegaer for å hente inn seleksjonsdata. Jeg vil også takke Grethe Sjøli ved Luftforsvarets Kontroll og Varslingsskole, for å hjelpe til med innhenting av skolerresultat. Til slutt må jeg også takke min samboer, Beate. Uten deg, hadde jeg ikke klart å holde meg så strukturert, som jeg faktisk har vært.

Per Sigurd Dvergsdal

Forsvarets Høgskole 28. november 2016

Sammendrag

Formålet med denne studien var å se på seleksjonstestenes prediktive validitet ved Luftforsvarets Seleksjonssenter (LSES) opp mot utdanningsresultat ved Luftforsvarets Kontroll og Varslingsskole (LKVS). Data ble samlet inn fra LKVS og LSES i perioden fra 2002 til 2016.

Prediktorene brukt i undersøkelsen var testene fra LSES innenfor Generell Teori (GT), Spatiale Evner (SE), Informasjonsprosessering (IP), Intervjuprognose, og Allment Evnenivå (AE) fra sesjon. Kriteriene ble valgt ut fra fagene innenfor den grunnleggende utdanningen og Kontrollutdanningen. Innsamlede data fra LKVS, ble også brukt til å undersøke om den grunnleggende utdanningen kunne predikere noe for den spesialisierende Kontrollutdanningen.

Innsamlede variabler ble først brukt i korrelasjonsanalyser for å se på sammenhenger mellom prediktorer og kriterier. Deretter ble det gjennomført en multipl hierarkisk regresjonsanalyse for å se på forklart varians for de ulike kriteriene, og i hvilken grad prediktorene kan predikere fremtidige resultat.

For den grunnleggende utdanningen viste spesielt AE og GT testene prediktiv validitet for oppnådde prestasjoner. I tillegg viste det en sterk sammenheng mellom prediktoren Engelsk og oppnådde prestasjoner for alle kriterier. Intervjuprognosen viste også noe prediktiv validitet for denne utdannelsen. For Kontrollutdanningen var det få sammenhenger mellom seleksjonstestene og oppnådde prestasjoner. En mulig forklaring på dette var størrelsen på utvalget ($N = 59-92$). Resultatene viste en sammenheng mellom prediktorer og oppnådd resultat i teoretisk fag, men ikke for de praktiske fagene utført i simulator. Det var noen flere sammenhenger mellom den grunnleggende utdanningen og Kontrollutdanningen. Her viste spesielt hovedkarakteren for den grunnleggende utdanningen prediktiv validitet for det ene praktiske faget, *Air Combat Training*.

Undersøkelsen viser at deler av seleksjonstestene viser god prediktiv validitet for den grunnleggende utdanningen. Seleksjonstestene viser lav validitet for prediksjon av utdanningsresultat for Kontrollutdanningen. For å gjøre seleksjonsprosessen enda bedre, anbefales det å gjennomføre en jobbanalyse for funksjoner innenfor både Kontroll og Varsling.

Summary

The purpose of this study was to evaluate the test battery currently used for selection of Air Surveillance and Control students at the Control and Reporting School in Norway. The sample used in this study, were students graduated at the school from 2002 to 2016.

Selection is currently based on tests encompassing General Mental ability (GMA), Spatial Ability, Information processing and an Interview. Criteria were collected from the basic Control and Reporting education, and the Fighter Control education. In addition, it was examined if the basic education could predict education results at the later Fighter Control education.

This study demonstrated that especially GMA had predictive validity for the basic education. In addition, Interview ratings showed some predictive validity for some of the criteria used. Finally, the test measuring English was a good predictor of all criteria used for this education. Correlations between selection tests and the Fighter Control education, showed more or less no predictive validity. One possible explanation to the lack of findings is the small samples sizes, especially for the Fighter Control education ($N = 59-92$). Significant correlations between GMA and theoretical academic results were observed, but there were no significant correlations between the selection tests and simulation scores. The correlation and regression analysis showed some predictive validity between the basic education and the Fighter Control Education. Especially the overall grade for basic education showed good predictive validity for Air Combat Training.

Overall, this study demonstrates that the test battery shows good predictive validity for the basic education. The test battery shows low validity to predict Fighter Control education results. The only significant correlation was between GMA and theoretical academic performance.

To improve the selection process, it is recommended to perform a Job Analysis for both Surveillance functions and Fighter Control functions in order to find new predictors that may be used in the selection process, and also in order to develop valid and relevant criteria for validation studies.

Innholdsfortegnelse

1 Innledning	1
1.1 HVA HANDLER STUDIEN OM?	1
1.2 HVORFOR ER DETTE INTERESSANT Å FORSKE PÅ?	1
1.3 KONTROLL OG VARSLING	2
1.4 SELEKSJON TIL K&V OG UTDANNING PÅ LKVS	3
1.5 PROBLEMSTILLINGER	4
1.6 FORVENTNINGER TIL RESULTATENE	5
2 Forskning på seleksjon	6
2.1 JOBBANALYSE	6
2.2 VALIDERINGSSTUDIER	7
2.3 META-ANALYSE	9
2.4 TRE FEILKILDER	10
2.5 FORSKNINGSFUNN	11
2.5.1 Et historisk tilbakeblikk på seleksjon	11
2.5.2 Generell intelligens som prediktor	14
2.5.3 Spatiale evner og Informasjonsprosessering	16
2.5.4 Intervju	17
2.5.5 Femfaktormodellen (FFM)	19
3 Seleksjon og utdanningen ved LKVS	20
3.1 SELEKSJON OG UTDANNINGSMODELL	20
3.1.1 Seleksjonsprosessen i praksis	20
3.1.2 Utdanningsmodell før 2007	21
3.1.3 Utdanningsmodell etter 2007	22
3.1.4 Utdanning fra krigsskole etter 2006	23
4 Metode	25
4.1 DELTAKERE	25
4.2 PREDIKTORER	25
4.3 KRITERIER	28
4.4 VURDERING AV FRAFALL	32
4.5 PROSEDYRE	33
4.5.1 Statistiske analyser	33
5 Resultat	34
5.1 KORRELASJONSANALYSE	34
5.2 MULTIPPEL HIERARKISK REGRESJONSANALYSE	39
6 Diskusjon	44
6.1 DRØFTING PROBLEMSTILLING 1	44
6.2 DRØFTING PROBLEMSTILLING 2	49
6.3 DRØFTING PROBLEMSTILLING 3	55
6.4 SVAKHETER OG STYRKER VED STUDIEN	58
7 Konklusjon	61
7.1 VEIEN VIDERE	62
7.2 MULIG VIDERE FORSKNING	64
8 Litteraturliste	65

1 Innledning

Forsvaret er på mange måter kjent for sine tester og seleksjonsmetoder. For de som skal i førstegangstjeneste, er sesjon første møte med Forsvaret. Er ønsket å ta en lederutdanning i Forsvaret, må kandidater gjennomføre opptaksprøver gjennom Forsvarets opptak og Seleksjon (FOS) med en varighet på to uker (Forsvaret, 2016). En typisk overskrift i media kan være «Her velger psykologer ut lederspirer til Forsvaret» (Stranden, 2015). Når kandidatene først har kommet inn i Forsvaret er mulighetene mange, ved valg av fagretninger. Avhengig av fagretning, krever noen av disse også seleksjon. Luftforsvaret gjennomfører seleksjon for flere fagretninger. Den mest kjente er sannsynligvis pilotutdanningen. For at pilotene skal fungere optimalt i luftrommet, kreves dette støtte av en annen fagretning. I det sivile blir disse ofte omtalt som flygeledere. I forsvaret kalles de Fighter Controller (FC), eller jagerflykontrollører.

1.1 Hva handler studien om?

FC består av et større fagmiljø som heter Kontroll og Varsling (K&V). Utdanningen til K&V foregår på Luftforsvarets Kontroll og Varslingsskole (LKVS), og formålet med denne undersøkelsen er å evaluere seleksjonen av kandidater til LKVS. Det har tidligere vært gjennomført en studie, der formålet var å kartlegge behovet for en systematisk seleksjon av personell inn til K&V (Olsen, 2002). Denne studien ble gjennomført omtrent på samme tid som det ble besluttet å innføre bruk av seleksjonstester (Olsen, 2002). Av denne grunn var det ikke nok kvantitative data til å kunne gjøre en analyse av seleksjonstestenes prediktive validitet. Det har med andre ord ikke blitt gjennomført undersøkelser nasjonalt, av sammenhengen mellom seleksjonssystemet og senere prestasjoner i form av utdanning innenfor K&V. I denne studien er utdanningsresultatet fra alle elever ved LKVS fra 2002 frem til 2016 tatt med. Ved å se på dette utvalget, vil målet være å se på om de seleksjonstestene som har blitt gjennomført ved Luftforsvarets seleksjonssenter (LSES), har en prediktiv validitet og sikrer gode kandidater til utdanningen.

1.2 Hvorfor er dette interessant å forske på?

Fra 2002 og frem til i dag har årskullene variert i størrelse ved utdanningen på LKVS. Basert på datasettet innhentet til denne undersøkelsen, er det totalt 218 personer som har gjennomført utdanningen fra 2002 og frem til 2016. Dette gir et gjennomsnitt på i underkant av 17

personer per år. Dette betyr at det er relativt få som tar denne utdanningen årlig. I tillegg er det to fagretninger innenfor K&V bransjen. Disse to retningene er henholdsvis Kontroll og Varsling. Bakgrunnen for at LSES startet med å selekttere K&V personell, var for å sikre at de som valgte Kontrollretningen hadde de riktige evnene for å kunne utøve yrket sitt. Siden alle som tar utdanningen ikke gjennomfører Kontrollutdanning blir det et lavere antall til denne utdanningen. Foreløpig er det ikke noe forskning i Norge, som sier noe om sammenhengen mellom seleksjonstestene som blir utført ved LSES og de endelige utdanningsresultatene ved LKVS. Det blir brukt betydelig med ressurser for å få personellet utdannet og det er derfor viktig å sikre at de beste kandidatene velges ut. Et relativt lite antall elever per år, krever fortsatt en viss mengde med instruktører som må utdanne studentene. Den største utfordringen er likevel turnover til andre områder i Forsvaret, eller at de slutter i Forsvaret. Nyttien ved en god seleksjon kan være avgjørende for K&V bransjen. Kombinasjonen av høye utdanningskostnader i lag med at viktigheten av jobben, krever en seleksjon som er effektiv. Bedre kandidater vil sannsynligvis gi bedre prestasjoner under utdanning, samt mindre frafall. Andre mulige gunstige konsekvenser kan være bedre prestasjoner i jobben i etterkant og kanskje mindre turnover.

1.3 Kontroll og Varsling

K&V har ansvaret for luftkontroll og varslingsoperasjoner. De har ansvar for kontinuerlig luftvarsling, luftbildeproduksjon og kontroll med taktiske luftstridsmidler (Forsvarets Høgskole, 2014, s. 117). K&V som bransje utøver sin operative tjeneste først og fremst ved ARS¹, eller ved ARE/TACS², og er delt inn i tre hovedfunksjonsområder. Dette er Kontroll, Varsling og Ledelse. Selv om ledelse er et hovedfunksjonsområde innen K&V er bransjen delt inn i to fagsøyler, henholdsvis Kontroll og Varsling. Fagsøylen Kontroll har ansvaret for kontroll av kampfly³, hvor de kommuniserer med kampflypiloter og leder eller støtter disse avhengig av type oppdrag. Fagsøylen Varsling er ansvarlig for å produsere et gjenkjent luftbilde. Dette innebærer døgkontinuerlig overvåking av norsk luftrom. Ved å skape denne oversikten klarer de å identifisere alt av lufttrafikk, om det er uønsket eller ønsket (Forsvaret, 2015). Selv om dette er to forskjellige fagretninger, kreves det en kontinuerlig dynamisk

¹ Air Control Centre-Recognized air picture Production Centre-Sensor Fusion Post

² ARS Element/Tactical Air Control System

³ Også omtalt som jagerfly

samhandling mellom Kontroll og Varsling for å sikre det overordnede målet, som er å sikre overvåking og kontroll av det norske luftrom.

De som blir utdannet innen fagsøylen Kontroll får fagbenevnelsen Fighter Controller (FC), også nevnt som Jagerflykontrolloffiser eller Kontrollør. Hovedoppgaven til en FC er taktisk ledelse av kampfly, slik at tildelte mål kan identifiseres, avvises eller engasjeres (Inspektorat, 2016). Det var først i år 2000 at en systematisk seleksjon av fremtidige jagerflykontrolloffiserer ble innført. Grunnen til denne innføringen var at det var mange innenfor fagmiljøet som mente at en FC burde ha mye av de samme egenskapene som krevdes for en flyger, siden de jobbet under lignende forhold. I tillegg lignet arbeidsoppgavene for en FC mye på arbeidsoppgavene til en sivil flygeleder sine arbeidsoppgaver (Olsen, 2002). *United States Air Force Research Laboratory* fra 2001, gjennomførte en kognitiv jobbanalyse om jagerflykontrolloffiserer om bord på en AWACS plattform⁴ (Fahey, Rowe, Dunlap, & deBoom, 2001). I denne analysen blir det spesifikt nevnt at en FC kan sammenlignes med en *Air Traffic Controller*⁵ (ATC), med noen viktige forskjeller. Kommersielle fly skyter sjelden på hverandre, og en sivil flygeleder trenger aldri å overvåke et luftbårent signal på skjermen for å finne ut av hvilken hensikt dette flyet har. ATC har heller ikke behov for å vurdere engasjements kriterier, med tanke på om det må skytes ned eller ikke. Et annet viktig forhold mellom ATC og FC, er at FC i stor grad er opptatt av å rettlede fly mot hverandre. For eksempel å iverksette en avskjæring av et fly eller en luft til luft tankning. Sivile flygeledere skal prøve å holde fly fra hverandre (Fahey et al., 2001). De krevende arbeidsoppgavene viser noe av viktigheten av å selektere riktig personell inn som FC.

1.4 Seleksjon til K&V og utdanning på LKVS

Etter seleksjon av jagerflykontrolloffiserer startet i år 2000 har utdanningsmønsteret forandret seg. Frem til og med 2006 var FC utdanningen i stor grad forbeholdt de som gjennomførte Luftkrigsskolen (LKSK). Med noen få unntak, vil dette si at det var kun de som gjennomførte LKSK og FC utdanning som gjennomførte seleksjon ved LSES. I 2006 ble det besluttet at denne utdanningen kunne gjennomføres på befalsskolenivå. I 2007 ble derfor første kullet på befalsskolenivå gjennomført med kontrollutdanning, med den konsekvens at fra dette året ble de aller fleste som søkte på K&V testet gjennom LSES.

⁴ AWACS er et mobilt luftsystem som står for Airborne Warning and Control System og utfører i stor grad de samme funksjonene som K&V bransjen gjør i Norge.

⁵ ATC er betegnelsen for en sivil flygeleder.

I utdannelsen på LKVS får de først en grunnleggende utdanning, og deretter kan de spesialisere seg innen Varsling eller Kontroll. De som gjennomførte utdanningen ved LKSK og Mågerø frem til 2006, fikk både en utdanning innen fagsøylen Kontroll og fagsøylen Varsling. Dette for at offiseren skulle få en bredde innenfor begge områder og på sikt, når de kunne tiltre ledelsesfunksjoner, hadde kunnskap fra begge fagsøyler. Etter 2006 ble utdanningsstrukturen noe annerledes. Til tross for endringene i 2006, har det alltid vært en grunnleggende utdanning i bunn, som har gitt elevene en forståelse for begge fagsøylene. Denne utdanningen heter Grunnleggende Kontroll og Varslingsutdanning (GKV). Av denne grunn er det mulig å se på resultater fra og med 2007 og frem til i dag på resultatene de har oppnådd ved GKV utdanningen. De som har gjennomført Kontrollutdanning⁶ fra 2007 og frem til i dag har også gjennomført GKV utdanning. Dette åpner for muligheten til å se på om det er sammenheng mellom GKV utdanningsresultat og påfølgende utdanningsresultat innenfor Kontrollutdanningen. De som gjennomførte kontroll utdanning før 2007, vil ikke ha resultater fra GKV, men de vil ha utdanningsresultat innenfor Varsling, siden denne utdanningen var en kombinert utdanning innenfor både Kontroll og Varsling.

Det har ikke vært noen forskning eller valideringsstudier innenfor Varsling spesielt. Innenfor Kontroll faget har det derimot vært mer forskning. Da spesielt innenfor flygelederyrket. Innenfor dette yrkesområdet og generell forskning innenfor Luftfartspsykologi, viser at seleksjonstester har en prediktiv validitet for både jobbprestasjoner og utdanningsprestasjoner (Broach & Manning, 1997; Martinussen, Jenssen, & Jøner, 2000; Pecena et al., 2013).

1.5 Problemstillinger

Det er to områder som peker seg ut med tanke på seleksjon til K&V. Den første vil være å se på sammenhengen mellom seleksjonstestene og resultatene på LKVS, fortrinnsvis for GKV og Kontrollutdanningen. Det andre området er om det er sammenheng mellom hva elevene oppnår av Varsling og GKV resultat og resultat på kontrollutdanning. Med dette som utgangspunkt vil denne oppgaven undersøke følgende tre problemstillinger:

Problemstilling 1:

Undersøke seleksjonstestenes prediktive validitet opp mot utdanningsresultat på GKV

⁶ Kontroll utdanning vil også forekomme med ordlyden FC utdanning. Dette er det samme.

Problemstilling 2:

Undersøke seleksjonstestenes prediktive validitet opp mot utdanningsresultat på Kontroll

Problemstilling 3:

I hvilken grad kan resultat fra Varsling og GKV utdanning predikere utdanningsresultat på Kontroll?

1.6 Forventninger til resultatene

Tidligere forskning innenfor Luftfartpsykologi tilsier det skal være en sammenheng mellom kognitive seleksjonstester og utdanningsresultat. Spesielt innenfor generell intelligens (*g*) (Schmidt & Hunter, 1998), men også andre spesifikke tester som spatiale evner⁷ og innenfor informasjonsprosessering (Martinussen et al., 2000). Til problemstilling nummer 1, bør det være en viss sammenheng, spesielt innenfor tester for generell intelligens. Kanskje i noen mindre grad for de spesifikke evnetestene. Noe forskningslitteratur mener generell intelligens predikerer arbeidsprestasjoner på tvers av yrker (Schmidt & Hunter, 1998). De spesifikke evnetestene viser til bedre prediktive validitet for spesifikke yrker, hvor det er mer behov for disse evnene (Damitz & Eissfeldt, 2004). Siden GKV er en generell utdanning, uten spesifikk utdanning innenfor Kontroll, kan det være slik at de spesifikke evnetestene ikke viser prediktiv validitet for GKV utdanningen. Til problemstilling nummer 2, bør det være en sammenheng mellom seleksjonstestene og utdanningsresultat. Både i forhold til generell intelligens og spesifikke evnetester. For problemstilling nummer 3, forventes det noen grad av sammenheng. Antakelsen er at de som presterer bra innenfor GKV også gjør det bra innenfor Kontroll, og på denne måten kan tidligere atferd predikere senere atferd (Huffcutt, Roth, & McDaniel, 1996; Martinussen, 1996).

⁷ Spatiale evner: «those abilities that enable a person to locate an object in space, mentally rearrange objects, recognize shapes, and so on.» (Boer, 1991, s. 104)

2 Forskning på seleksjon

Hensikten med dette kapittelet, er å gi et overblikk over hvordan forskning på seleksjon foregår og har utviklet seg. Videre vil det bli presentert noen metodiske utfordringer som er forbundet med slike valideringsstudier.

2.1 Jobbanalyse

«*To match the right person with the right kind of job*» (Suresh, Ramachandran, & Srivastava, 2012). Det kan gjerne sies at dagens arbeidsplasser har blitt mer turbulente og uforutsigbare. Økende konkurranse, nye markeder, innovasjon, utvikling og globalisering har påtvunget firmaer og organisasjoner til å ha et enda større fokus på seleksjon. I lys av dette har jobbanalyser blitt mer vanlig. Eissfeldt og Heintz (2002) mener at en fullverdig jobbanalyse må være gjennomført, for å oppnå et optimalt seleksjonssystem.

Det første steget i en seleksjonsprosess er å utføre en jobbanalyse. Målet med en jobbanalyse er å identifisere ansvar, oppgaver, kognitive krav, kunnskap, ferdigheter, evner og eventuelt andre krav som stilles. En jobbanalyse er således en systematisk prosedyre for å samle, dokumentere og analysere informasjon om innhold, kontekst og betingelser om jobben. Den demonstrerer at det er en klar sammenheng mellom jobboppgaven, og kompetanse for å utføre oppgaven (Suresh et al., 2012). Det er flere måter å gjennomføre en jobbanalyse på, og det endelige resultatet bør lede til en utvikling av tester og kriterier som er målbare (Carretta & Ree, 2000). I en jobbanalyse er det fortrinnsvis to områder som blir kartlagt, henholdsvis en beskrivelse av jobben og en personspesifikasjon. Beskrivelse av jobben, vil være en kartlegging av ansvar og oppgaver som jobben innebærer (Martinussen & Hunter, 2008, s. 110). I K&V vil dette komme frem i en stillingsbeskrivelse. En enda mer detaljert beskrivelse av ansvar og oppgaver vil komme frem i funksjonsbeskrivelsen til hver enkelt posisjon. En funksjonsbeskrivelse vil inneholde en forholdsvis detaljert oppsummering av gjøremål, og oppgaver som vedkommende skal kunne håndtere i posisjon. Det andre området som er personspesifikasjon, omfatter hvilke egenskaper en person bør ha for å kunne utføre jobben best mulig. Dette vil derfor være en spesifisering av hvilke fysiske og psykiske egenskaper vedkommende bør ha. Innenfor K&V bransjen i Norge, er det ikke gjennomført noen dokumentert personspesifikasjon av de forskjellige funksjonene. Av denne grunn har ikke K&V bransjen i Norge en klar jobbanalyse av de forskjellige arbeidsposisjonene ved en ARS. Selv om K&V bransjen ikke har dette, er det mulig å se til studier og undersøkelser innenfor

ATC. Det skal også sies at større studier har vist at evnetester kan predikere jobprestasjoner uavhengig av yrke, noe som også kan tilsi detaljerte jobbanalyser nødvendigvis ikke er det viktigste (Schmidt & Hunter, 1998; Schmidt, Oh, & Shaffer, 2016).

En jobbanalyse kan gjennomføres på forskjellige måter, avhengig av hva målet er med jobbanalysen. I forbindelse med luftfart, er en av de mest brukte jobbanalysemodellene Fleishmans metode (*Job Analysis Survey*) (Goeters, Maschke, & Eissfeldt, 2004, s. 99). Denne metoden tar utgangspunkt i å spørre arbeidstakerne i å vurdere sine egne jobbkraav. Siden et problem er at de aller fleste arbeidstakere ikke har utdannelsen til å kunne beskrive sin erfaring i sofistikerte termer, utviklet Fleishman en syvdelt skala med enkle definisjoner innenfor de psykologiske komponentene. Totalt er det syv forskjellige områder som ble vurdert, med en rekke undergrupper i hvert område. Områdene var kognitive evner, fysiske evner, sensoriske evner, kunnskap og ferdigheter, og til slutt ferdigheter og sosiale evner (Goeters et al., 2004, s. 99-100). Denne metoden ble brukt i en studie av tyske flygeledere (Eissfeldt & Heintz, 2002). I tillegg til den originale skalaen Fleishman metoden benyttet seg av, ble det også tilført skalaer som beskrev mer personlige egenskaper. Dette var egenskaper som samarbeidsevne, kommunikasjonsferdigheter og evne til å håndtere stress. I denne undersøkelsen kom det frem at det var spesielt kognitive evner som hurtig persepsjon⁸, visualisering⁹, selektiv oppmerksomhet¹⁰ og tidsdeling¹¹ som fikk høyest rangering. I tillegg var det svært få av de kognitive evnene som fikk lav skåre (Eissfeldt & Heintz, 2002).

2.2 Valideringsstudier

Utvelgelsen av både flygere og flygeledere har utviklet seg over tid. For over 100 år siden kastet brødrene Wilbur og Orville Wright krone og mynt, på hvem som skulle være den første til å ta flyturen med deres selvbygde flymaskin (Martinussen, 2005, s. 291). I dag har utvelgelsesmetodene utviklet seg til å ta i bruk moderne systemer og databaserte tester. Kanskje spesielt innenfor pilot og flygelederyrket, har det blitt lagt ned mye ressurser, for å finne de rette seleksjonstestene til å selektene riktige kandidater til yrket. For pilotyrket har det

⁸ *Speed of closure*: «the ability to organize different pieces of information quickly into a meaningful pattern» (Eissfeldt & Heintz, 2002, s. 17)

⁹ *Visualization*: «related to the imagination of spatial relationships or movements» (Eissfeldt & Heintz, 2002, s. 17)

¹⁰ *Selective Attention*: «the ability to concentrate on a task over time without being disturbed» (Eissfeldt & Heintz, 2002, s. 17)

¹¹ *Time sharing*: «the ability to shift efficiently between different tasks» (Eissfeldt & Heintz, 2002, s. 17)

vært begrunnet med de høye kostnadene til utdanning, og for å bedre sikkerhet og reduksjon av antallet ulykker. Dette gjelder også for flygelederyrket, hvor det i et høyteknologisk miljø med stadig høyere flytrafikk gjør oppgavene mer krevende. Konsekvensene av feil kan også være fatale.

For å selektere personell til slike yrker er det vanlig å bruke psykologiske tester. Disse testene betegnes som prediktorer. For å kunne være sikker på at disse prediktorene skal kunne predikere for fremtidige jobbprestasjoner, er det vanlig å gjennomføre valideringsstudier. Valideringsstudier undersøker hvor godt et testresultat er korrelert med senere prestasjoner i jobben eller under utdanning. Indikatorene på senere jobbprestasjonen blir kalt kriterier. Jobbprestasjoner kan være både skoleprestasjoner eller andre mål på jobbutførelse. En psykologisk test for å vurdere om vedkommende er egnet for å være FC vil derfor være en prediktor. Ønsket med testen er å kunne predikere hvem som vil prestere godt i jobben. Det er derfor viktig at testene som blir brukt til seleksjon, kan dokumentere at de predikerer framtidige prestasjoner (Martinussen & Hunter, 2008, s. 114-115).

Valget av prediktorer bør ikke bare være basert på antagelser, men noe som er relevant for fremtidige jobbprestasjoner. Prediktorene kan for eksempel bli identifisert gjennom jobbanalyser. Ved seleksjon er det etter hvert blitt mange forskjellige typer metoder som blir brukt. Dette kan være alt i fra intervju, til arbeidsprøver, til forskjellige typer psykologiske tester. Psykologiske tester innebærer evnetester og personlighetstester og er mye brukt innenfor luftfart, spesielt for piloter og flygeledere. Videre kan det benyttes skolekarakterer og arbeidserfaring (Martinussen & Hunter, 2008, s. 115-116). Nevnte metoder har forskjellig ulik prediktiv validitet. Carretta og Ree (2000) beskriver prediktiv validitet som den fundamentale test og seleksjonsutfordringen. Dette fordi at i den grad en test kan være til nytte, bestemmes av hvor sterk sammenhengen er opp mot resultatene på et relevant kriterium. Selv om det er brukt mye anstrengelser for å utvikle seleksjonsmetoder basert på teorier, hjelper det ikke uten at man er i stand til å dokumentere dette empirisk. Teori uten belegg er i beste fall verdiløst. Med andre ord så bør det foreligge evidens for at en test kan brukes som prediktor. Dette kan en finne gjennom tidligere valideringsstudier

Lokale valideringsstudier undersøker sammenhengen mellom et testresultat og et mål på en arbeidsprestasjon (kriteriet). Arbeidsprestasjonen kan være vurdert gjennom en

simulatorøvelse eller vurdering gjort av en instruktør. Et testresultat kan også være en kombinasjon av flere tester hvor en kan lage en samlet skåre på resultatene, før de blir korrelert med kriteriet (Martinussen & Hunter, 2008, s. 118). I denne sammenheng er det ikke bare prediktor som er viktig, men også at valget av kriterium er riktig. Et kriterium bør ha god validitet og reliabilitet. For å sikre validiteten eller gyldigheten, kan dette gjøres ved å se på tidligere jobbanalyser. Disse analysene vil ha vurdert og identifisert hvilke kriterier som kjennetegner en god arbeidsprestasjon. Reliabiliteten kan også være utfordrende. Innenfor K&V bransjen enten om det er gjennom utdanning eller på ARS, så blir kandidat evaluert av instruktører som måler arbeidsprestasjonen til kandidaten. Innenfor utdanning blir det brukt to til tre instruktører. For å ivareta reliabiliteten vurderer instruktørene selvstendig kandidatens prestasjon. I etterkant sammenligner instruktørene sine observasjoner og kommer frem til en vurdering av kandidat. Slike personvurderinger innebærer en del fallgruver. Dette kan være dersom kandidaten gjør det bra på ett område, med følgekonsekvensen at det faller positivt ut på de andre områdene, selv om kandidaten nødvendigvis ikke gjør det så bra. I andre tilfeller kan det være vanskelig å bruke hele skalaen, som kan føre til lite variasjon på de ulike punktene i vurderingen. Det blir vanskelig å måle variasjon dersom prestasjonene vurderes like bra. Det er derfor viktig at instruktører blir trent på dette, og spesifiserer de gode og dårlige resultatene på de ulike områdene (Martinussen & Hunter, 2008, s. 116-117).

2.3 Meta-analyse

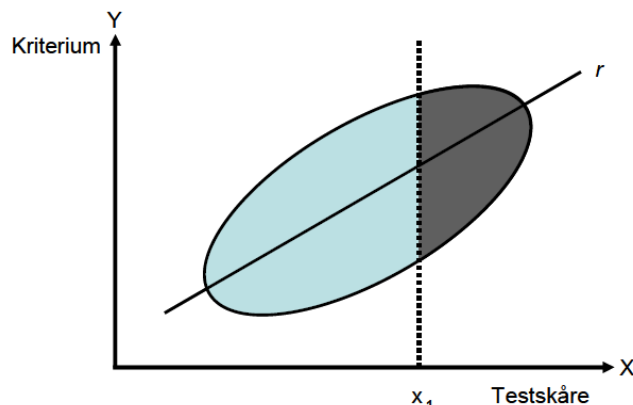
Innenfor valideringsstudier skilles det mellom lokale valideringsstudier og meta-analyser. En svakhet med lokale valideringsstudier er at utvalget (N) kan bli for lite og dermed blir den statistiske styrken for svak. Den statistiske styrken er basert på hvor høy sannsynligheten er for å oppdage en signifikant effekt (Carretta & Ree, 2000). Ved å få tilgang til større utvalg (N) vil den statistiske styrken være bedre. Sannsynligheten for at den sanne korrelasjonen mellom prediktor og kriterium blir avdekket vil dermed være større. Som et resultat av dette vil den prediktive validiteten bli bedre, og den kan i større grad generaliseres (Martinussen & Hunter, 2008, s. 118). En metode for å få bedre statistisk styrke er å gjennomføre en meta-analyse. En meta-analyse er en kombinasjon av tidligere undersøkelser innenfor samme felt. Denne teknikken kom på 1970 tallet, og var en teknikk for å integrere resultat fra flere studier inn i en studie (Jones, 1991). Det finnes for øvrig flere teknikker for å utføre en meta-analyse (Martinussen & Hunter, 2008, s. 118). For å gjennomføre en meta-analyse er det behov for en felles måleenhet. Blant flere standardiserte indekser, er Pearsons r en mye brukt statistisk

indeks på samvariasjon, både i valideringsstudier og meta-analyser. Pearsons r brukes til å forklare styrken på sammenhengen mellom prediktor og kriterium (Ringdal, 2013, s. 304). Vanligvis bruker valideringsstudier korrelasjonsresultat for hver enkelt prediktor, disse kan fint brukes i en meta-analyse. Det kan også forekomme valideringsstudier der det kun er brukt multippel regresjonsanalyse. Siden denne type analyse er brukt for å måle flere prediktorer som er samlet for å predikere et kriterium, kan ikke disse studiene brukes i en meta-analyse. Heller ikke de bivarierte korrelasjonene rapportert i en slik analyse kan brukes (Martinussen & Hunter, 2008, s. 118).

2.4 Tre feilkilder

Ved bruk av både valideringsstudier og meta-analyser er det spesielt tre statistiske feilkilder som må tas hensyn til. Dette er manglende redusert spredning i testskårene, manglende reliabilitet i variablene og bruk av todelt kriterium. Todelt i denne sammenheng vil eksempelvis være et bestått, ikke bestått kriterium (Martinussen & Hunter, 2008, s. 117).

Redusert spredning, også kalt *Restriction of Range* (RR), er en feilkilde som forekommer når utvalget er begrenset, på grunn av tidligere seleksjon kandidatene har vært gjennom. En streng seleksjon vil føre til et mindre utvalg til å inkludere i studiet. Variansen eller spredningen av det variablene måler blir dermed mindre (Martinussen & Hunter, 2008, s. 119). Ved inntak til K&V vil dette være aktuelt, da de som er selektert inn til utdanningen allerede er selektert gjennom tester, intervju og tidligere karakterer. Dette betyr at den beste halvdel allerede er plukket ut, og det er disse det er mulig å hente ut kriterie data på (Carretta & Ree, 2000). Ved å se på Figur 1, kommer dette tydeligere frem. Figuren illustrerer sammenhengen mellom testskåre og kriterium. Hele utvalget blir representert gjennom den ovale sirkelen. Det skraverete feltet til høyre for X_1 representerer selektert utvalg. Beregnet korrelasjon for det selekterte utvalget blir følgelig svakere, siden ikke hele utvalget er tatt med. Jo strengere seleksjon er, desto lavere blir korrelasjonen. Det finnes metoder for å korrigere for RR, hvor Schmidt og Hunter (2014) sin metode er mest brukt i forbindelse med valideringsstudier. Med denne metoden er det mulig å korrigere for RR i tillegg til manglende reliabilitet i kriteriet (Martinussen & Hunter, 2008, s. 119-120).



Figur 1: Illustrasjon av «Restriction of Range» (Martinussen & Hunter, 2008).

Reliabilitet i variablene er en annen feilkilde som må tas hensyn til. I denne sammenheng må det skilles mellom uavhengige og avhengige variabler. Avhengige variabler vil være kriteriene. Uavhengige variabler vil være testresultatene (prediktorene). I denne oppgaven vil det være de faglige resultatene som vil være de avhengige variablene. Jo lavere reliabiliteten til testene og kriteriet er, desto svakere blir den observerte korrelasjonen. Det er vanlig å korrigere for manglende reliabilitet i kriteriet i valideringsstudier, da man er interessert å anslå den prediktive validiteten til testene, med de feil og mangler som de måtte ha (Martinussen & Hunter, 2008, s. 120-121).

Den tredje og siste feilkilden er bruken av kriterium som er todelt, også kalt dikotomt kriterium (Carretta & Ree, 2000). Innenfor det operative miljøet på ARS i K&V systemet er dette mye brukt. Denne type kriterium kan skape en korrelasjon som vil være lavere enn den burde, eller det man ville ha oppnådd om prestasjonene hadde blitt målt kontinuerlig. Eksempelvis med en karakterskala fra A til F. Dette har lenge vært et kjent problem og det finnes statistiske korreksjonsmetoder for dette, som kan estimere korrelasjonen dersom kriteriet ikke hadde vært dikotomt (Carretta & Ree, 2000). I denne oppgaven vil ikke dette bli et problem da LKVS sitt utdanningssystem krever karakterskala fra A til F.

2.5 Forskningsfunn

2.5.1 Et historisk tilbakeblikk på seleksjon

Forskning innenfor seleksjon av jagerflykntrolløffiserer er relativt begrenset. Forskning innenfor seleksjon av piloter både militært og sivilt, i tillegg til seleksjon av sivile flygeledere

har det vært mer av (Martinussen, 2005). Selv om det ikke eksisterer så mye forskning innenfor yrket som FC eller jagerflykontrolløffiser, startet de første undersøkelsene under andre verdenskrig med denne kategorien. Kontrollører (FC) ble mer og mer viktige på grunn av nye taktikker og strategier innenfor bruk av luftrom og kommando og kontroll. Dette krevde mer støtte fra ikke-flyvende personell. Derfor ble det allerede i 1941 opprettet et undersøkelsesprogram for å kunne identifisere «kontrollør ferdigheter». Her ble det utarbeidet fire tester som målte persepsjonshastighet og nøyaktighet, spatiale evner, deduktive slutninger og induktive slutninger. I 1950-årene startet *Civil Aviation Administration* (CAA), forløperen til *Federal Aviation Administration*¹² (FAA), ett felles program med det amerikanske flyvåpenet. Dette for å identifisere tester som kunne predikere gode resultater for de som skulle utdanne seg til flygeledere (Hättig, 1991). Fra 37 psykometriske tester var det matematiske slutninger, data- og numeriske ferdigheter, abstrakte slutninger og persepsjonshastighet som demonstrerte høyest prediktiv validitet. Senere studier på 60 og 70-tallet bekreftet tidligere funn for hva som var de beste prediktorene for ATC prestasjoner. 1960-tallet var også karakterisert med en kraftig økning i luftfartsindustrien, og raskere og tyngre fly. Dette, i tillegg til utvikling innen radar og data teknologi, ledet til nye ATC reguleringer. Kombinasjonen av reguleringer og teknologi førte til en økning i mental arbeidsbelastning for arbeidsprestasjoner. Forskjellige studier dokumenterte da at kontrollører eldre enn 36 år viste markant dårligere test og jobb prestasjoner (Hättig, 1991). På 70-tallet startet FAA utviklingen av en simulator som kunne måle simultankapasitet. Denne testen var en film som presenterte lufttrafikk i sanntid på en skjerm. Senere ble det utviklet, som et resultat av denne simulatoren, en papir- og blyant utgave. Denne testen ble hetende «*Multiplex Controller Aptitude Test*» (Hättig, 1991). På grunn av høye kostnader i seleksjonsprogrammet, ble det på 90-tallet utviklet et databasert testbatteri som skulle erstatte det tidligere seleksjonsprogrammet. Dette testbatteriet var designet for å teste spatiale slutninger, korttidshukommelse, oppdage bevegelser og oppfattelse av bevegelsesmønstre (Broach & Manning, 1997).

I Europa ble seleksjon av flygeledere et stort tema på slutten av 1970-tallet. Fram til da hadde studenter og kadetter blitt selektert gjennom en standard seleksjon. Denne seleksjonen var ikke spesifikt knyttet opp mot egenskaper ATC personell hadde behov for. I 1979 startet

¹² Det amerikanske luftfartstilsynet.

derfor EUROCONTROL¹³ en undersøkelse av sine medlemsland, i et forsøk på å få en oversikt over hvilke metoder de brukte. De fleste landene målte da resonneringsevne, spatial persepsjon, verbale evner og hukommelse. I 1984 kunne de fleste EUROCONTROL medlemmer rapportere om etablering av nye dedikerte seleksjonsprosedyrer basert på psykometriske¹⁴ standarder. *Royal Air Force* (RAF) kunne på samme tid ta i bruk det første komplette databaserte seleksjons-testbatteri for flygeledere. I 1986 startet Tyskland, *German Armed Forces* (GAF), et simulator basert testsystem basert på en papir- og blyant test utarbeidd av *Deutsche Forschungs- und Versuchsanstalt für Luft- und Raumfahrt* (Hättig, 1991). I 2003 lanserte EUROCONTROL et databasert testbatteri for utvelgelse av flygeledere. Testbatteriet erstattet dermed papir- og blyanttestene i de aller fleste land (Martinussen, 2005, s. 294). Testbatteriet som EUROCONTROL utviklet ble hetende *First European Air Traffic Controller Selection Test* (FEAST). Dette var en testpakke som inneholdt en serie med forskjellige type tester. Utviklingen av testpakken ble gjort med bakgrunn i flere jobbanalyser. Resultatet av dette ble en modell hvor de fikk kartlagt hvilke egenskaper som var viktig for en flygeleder (Rathje, Golany, & Eissfeldt, 2004). Totalt ble det listet opp 26 egenskaper som var viktig. Disse ble videre delt inn i tre hovedgrupper. Dette var persepsjon, resonnering og beslutningstaking. FEAST har for øvrig utviklet seg en del fra den første versjonen.

LSES sitt testbatteri er spesielt utviklet for seleksjon av piloter. I denne pakken er det plukket ut spesifikke tester til K&V. Når LSES gjennomfører sin seleksjon til K&V er det Generell teoriforståelse (GT), Spasiale evner (SE) og Informasjonsprosessering (IP) som blir testet. I tillegg har de også i den senere tid gjennomført intervjuer av kandidatene. Utover dette brukes det også Alminnelig evnenivå (AE) kartlagt ved sesjon. Dette betyr at det som FEAST har i sin pakke, blir i stor grad ivarettatt i LSES sine seleksjonstester. Av denne grunn, skulle det være grunn til å tro at kanskje disse seleksjonstestene, burde fungere bra som prediktorer til seleksjon for å bli FC.

Etter nå å ha redegjort både for et historisk tilbakeblikk på seleksjon og noen generelle funn, vil oppgaven ha et mer eksplisitt fokus på forskning innenfor de prediktorene LSES nytter seg av, ved seleksjon til K&V bransjen. LSES bruker AE, GT, SE og IP i tillegg til

¹³ «*European Organisation for the Safety of Air Navigation*» (EUROCONTROL, 2016). En organisasjon med totalt 41 medlemsland.

¹⁴ Psykomterisk: «*is the statistical study of psychological tests*» (Passer & Smith, 2011, s. 337).

intervjuprognose for å selekere personell til K&V. I den videre redegjørelsen vil dette bli brukt som et utgangspunkt for evner som viser prediktiv validitet.

2.5.2 Generell intelligens som prediktor

Både Alminnelig evnenivå (AE) og Generell teoriforståelse (GT) er mål på det vi kan kalle generell intelligens (Schmidt & Hunter, 1998; Skoglund, Martinussen, & Lang-Ree, 2015). Innenfor dette området er det blitt gjennomført mange studier. Det blir derfor redegjort for både generelle forskningsfunn om intelligens og om intelligens innenfor luftfart.

Intelligens blir ofte omtalt som *General Mental Ability* (GMA) (Schmidt et al., 2016), i andre tilfeller kan det også bli omtalt som *General Cognitive Ability* (GCA) (Schmidt, 2002). GMA eller intelligens (*g*) blir regnet for å være evnen til å lære. I denne sammenheng er ikke *g* referert til det genetiske potensiale, men til en utviklet GMA. Til tross for at det er en erkjennelse at *g* er sterkt påvirket av gener, reflekterer også GMA mer en bare genetikk. Det er også andre evner enn GMA som er relevante for prestasjoner og læring i mange jobber. Dette kan være psykomotoriske evner, sosiale evner og fysiske evner. Imidlertid varierer sistnevnte type evner mer i prediktiv validitet mellom ulike jobber, og har generelt lavere validitet (Schmidt, 2002).

Innenfor den psykometriske tradisjonen, har forskere lenge prøvd å finne svaret på hvilke kognitive evner som finnes og hvordan de er organisert, gjennom bruken av faktoranalyse (Passer & Smith, 2011, s. 337). Med denne metoden får de studert sammenhengen mellom de forskjellige spesifikke evnetestene, og kan kartlegge de faktorene som kan best forklare mønsteret av korrelasjoner mellom ulike evnetester. Metoden ble først brukt av Spearman (1924), hvor han mente at samvariasjonen mellom de ulike spesifikke evnetestene skyldtes en generell intelligensfaktor (*g*-faktor). Senere har det blitt vanlig å dele intelligens i en hierarkisk modell på tre nivå (Passer & Smith, 2011, s. 340). Modellen kalles «*The three-stratum theory of cognitive abilities*» (Carroll, 2003) og bygger på blant annet *Spearman's* teori. Det nederste nivået består av spesifikke kognitive ferdigheter. Det midterste nivået er gruppefaktorer, og det øverste nivået består av en generell intelligensfaktor (*g*) (Passer & Smith, 2011, s. 340).

Hunter og Hunter (1984) gjennomførte en meta-analyse på over 400 studier. Der kom de frem til at validiteten for GMA økte jo mer krevende yrket var. Høy kompleksitet i jobben viste en

gjennomsnittlig prediktiv validitet på $r = .58$, medium kompleksitet $r = .51$ og lav kompleksitet $r = .38$. Også andre studier viser til at GMA har høy prediktiv validitet ved seleksjon (Schmidt, 2002; Schmidt & Hunter, 1998).

En undersøkelse av Ree og Earles (1992) stilte seg spørsmålet om hvordan arbeidsgivere kan ansette de beste kandidatene. Dette baserte de blant annet på en undersøkelse av piloter og navigatører med et utvalg på $N = 5000$. Formålet med undersøkelsen var å se på «*performance on the job*». Resultatene de kom opp med viste en høy prediktiv validitet, og intelligens (g) alene kunne i stor grad kunne predikere jobbprestasjoner. De mente videre at ved å bruke intelligenstester, kunne en arbeidstaker forvente at kandidaten ville gjøre en god jobb uavhengig av hvilken type jobb det var.

Damitz og Eissfeldt (2004) tok utgangspunkt i påstanden om at g alene kunne predikere jobbprestasjoner, og at spesifikke evner kan predikere veldig lite utover g (Ree & Earles, 1992). I tillegg ble det også henvist til Olea og Ree (1994) som utførte en studie på 4000 piloter med papir og blyanttester, hvor de fant at g var den beste prediktoren for alle kriteriene som ble brukt. Damitz og Eissfeldt (2004) gjennomførte deretter en studie ved det tyske luftfartssenteret, *German Aerospace Center (DLR)*, hvor de undersøkte g på 193 kandidater. Til denne undersøkelsen valgte de spesifikke prediktorer, for å forsikre seg om at undersøkelsen kunne være direkte sammenlignbar med Olea og Ree (1994) sin undersøkelse. Prediktorene for å måle g ble dermed kun valgt ut fra DLR sitt papir-og-blyant testbatteri, og ingen arbeidsprøve tester¹⁵. De brukte deretter åtte forskjellige typer kriterier, fire innenfor teoretiske evner og fire innenfor praktiske evner. Resultatet var varierende, med signifikante korrelasjoner på tre av åtte kriterier. Etter de utførte regresjonsanalyser i etterkant, fikk de til resultat at for noen av kriteriene var g viktig, men for andre kriterier var den ikke viktig. Konklusjonen av denne studien, var at g alene ikke kunne predikere kandidatenes utdanningsprestasjon. Vet at de tilføyde spesifikke evnetester innenfor de andre områdene spatiale evner og informasjonsprosessering, fikk de merkbart høyere forklart varians i regresjonsanalysen.

¹⁵ *Work sample test* er basert på premisset om at prediktor måler for fremtidig atferd observert under lignende situasjon (Martinussen & Hunter, 2008, s. 115).

2.5.3 Spatiale evner og Informasjonsprosessering

Spatiale evner (SE) er en primær evne av menneskelig intelligens, som er klart separert fra verbal intelligens og generell resonneringsevne. Det eksisterer forskjellige definisjoner av SE, men generelt forklart er det de evner som gjør til at en person kan lokalisere et objekt i et rom, mentalt flytte om på objekter og gjenkjenne former. En annen måte å si det på, kan være at det er en mental manipulering av former (Boer, 1991). Informasjonsprosessering (IP) er evnene til å kunne ta innover seg informasjon, ved hjelp av sensoriske evner, som gjennom hørsel eller bilder. Gjennom denne behandlingen blir det tolket, med den konsekvens at informasjon kan gå tapt eller overført videre til lagring, eller handling (Kallus, Barbarino, & Van Damme, 1997). Blant annet arbeidsminne er en vital del av IP, hvor dette blir definert som «flaskehalsen» for informasjonsprosessering. Arbeidsminnet er et kort av begrenset kapasitet og kan behandle begrenset med informasjon. Dette spiller derfor i sammenheng med langtidsmminnet (Kallus et al., 1997).

Goeters et al. (2004, s. 110) viser til at kognitive evner er meget sentralt for flygelederyrket. De definerer kognitive evner som den dynamiske prosessen av å innhente data, prosessere og distribuere informasjon innenfor sentrale aspekt av lufttrafikk kontroll. Denne type informasjon kommer fra flere kilder, som visuelt, lyd og gestikulering. Og denne type informasjon må gjøres på samme tid (simultant). Det er utførelse av flere oppgaver på samme tid med høye krav på informasjonslagring og fortløpende justeringer av plan på grunn av ny innkommende informasjon. Typisk for lufttrafikk kontroll er tidskritiske oppgaver og kompleksiteten i oppgaven. I Goeters et al. (2004) sin jobbanalyse ble det brukt totalt 28 kontrollører, hvorav 16 var under trening og 12 av dem var erfarne ferdig utdannede kontrollører. I denne undersøkelsen som ble gjennomført ved *Deutsche Flugsicherung GmbH* (DFS) kom de frem til at *Time Sharing* var den av de kognitive evnene som viste seg som den viktigste evnen. Deretter var det *Selective Attention* og *Problem sensitivity*.

I en annen studie fra Tyskland ble det gjennomført en relativt omfattende undersøkelse med 430 kandidater, hvor de undersøkte sammenhengen mellom seleksjon av kandidater og de endelige utdanningsresultatene som kandidatene oppnådde (Pecena et al., 2013). I denne undersøkelsen kom det frem at de beste prediktorene var oppmerksomhet og multi-oppgave løsning.

I 2000 ble det gjennomført en undersøkelse for å se på alle valideringsstudier av tester som hadde blitt brukt til å velge ut flygeledere (Martinussen et al., 2000). Målet med undersøkelsen, var å se på om de kunne estimere den prediktive validiteten av de forskjellige prediktorene som hadde blitt brukt opp gjennom tidene. Dette resulterte i 35 forskjellige undersøkelser fra 25 forskjellige artikler. De forskjellige artiklene hadde blitt utført mellom 1952 og 1999, og majoriteten av studiene hadde blitt gjennomført i USA. Utvalget i denne meta-analysen varierte mellom 374 og 11255 personer for de ulike kategoriene av prediktorer. Resultatene de kom frem til var et underestimat av den sanne prediktive validiteten, siden de ikke fant informasjon som gjorde det mulig å korrigere for redusert spredning og reliabilitet i kriteriene. De beste prediktorene var tester som målte spatiale evner (.27) og *work sample* (.30). *Work sample* er en test som er laget slik at testen skal være mest mulig likt et reelt jobb scenario. I tillegg viste *composite score* (.30) høy prediktiv validitet. Siden *composite score* er en samlet indeks, som viser en kombinasjon av flere tester, er det kanskje ikke overraskende at dette viste en bedre prediktiv validitet. Kunnskap om yrket (.18) og simultankapasitet (.19) var de testene som hadde lavest korrelasjon.

2.5.4 Intervju

Intervju er jevnt over mye brukt, både i forbindelse med jobbansettelse og i gjennomføringen av seleksjon. Ansettelses intervju kan enten være strukturerte eller ustrukturerte intervju (Huffcutt et al., 1996). Ustrukturerte intervju har ikke noe bestemt format eller ett sett med forberedte spørsmål. Det innehar heller ikke en bestemt prosedyre. Svarene på spørsmål er vanligvis ikke skrevet ned og skåret etter en fast mal, men den totale evalueringen av kandidaten er basert på et oppsummert inntrykk og vurdering. Et strukturert intervju er forskjellig på mange punkter. I tillegg er spørsmålene normalt basert på en jobbanalyse og intervjueren kan ikke avvike fra planlagte spørsmål (Schmidt et al., 2016). En studie av McDaniel, Whetzel, Schmidt, og Maurer (1994) viste til, at i tillegg til ustrukturerte intervju og strukturerte intervju, spilte det en rolle om det var et panelintervju eller utført individuelt av en person. Siden panelintervju hadde høyere administrative kostnader enn individuelle intervju, måtte panelintervju ha en høyere prediktiv validitet, for at det skulle være kosteffektivt. I en meta-studie til Huffcutt, Culbertson, og Weyhrauch (2013) bruker de også en tredje variant av strukturerte intervju, som de kaller semistrukturert intervju.

En analyse gjort av McDaniel et al. (1994) så på sammenhengen mellom intervjuprognoser og jobbutførelse. Resultatene viste at for individuelle strukturerte intervju var $r = .25$ og individuelle ustrukturerte intervju var $r = .18$. For panelintervju viste strukturerte intervju $r = .20$ og ustrukturerte intervju $r = .18$. Denne analysen viste altså at individuelle intervju hadde høyere validitet enn intervjupanel.

I en oppdatert analyse av Schmidt et al. (2016), basert på den opprinnelige studien av Schmidt og Hunter (1998), viste den gjennomsnittlige validiteten for både strukturerte og ustrukturerte intervju at den fikk den samme validitetskoeffisienten på $r = .58$. Det samme datagrunnlaget hadde blitt brukt tidligere fra Schmidt og Hunter (1998), her viste resultatet for ustrukturerte intervju $r = .38$ og for strukturerte intervju $r = .51$. Bakgrunnen for at det ble forskjellige resultat var bruken av en mer oppdatert metode for korreksjon av RR. Den nye og mer presise metoden for korrigerings av RR viste altså relativt store forandringer i validiteten for både strukturerte og ustrukturerte intervju.

I studien til Pecena et al. (2013) av 430 ATC studenter i Tyskland, ble det gjennomført semistrukturerte intervju der resultatet ble delt opp i tre forskjellige prediktorer. Henholdsvis generell motivasjon, yrkesrettet motivasjon og sosial kompetanse. Intervjupanelet bestod for øvrig av to psykologer innenfor luftfart og to ATC kontrollører. Generell motivasjon viste høy prediktiv validitet for eksamensprestasjoner innenfor både teori (-.41) og simulator (-.16), i tillegg til totalt utdanningsresultat (-.17). For yrkesrettet motivasjon fikk de relativt høy prediktiv validitet, innenfor både teori (-.19) og utdanningsresultat (-.11), men likevel lavere enn for generell motivasjon. For simulator ble det ikke angitt noen prediktiv validitet siden korrelasjonen ikke var signifikant. Dette indikerer at generell motivasjon predikerer bedre enn yrkesrettet motivasjon, spesielt innenfor simulator trening. Det totale utdanningsresultatet ble predikert av både generell og yrkesrettet motivasjon. Totalt sett gir likevel undersøkelsen et entydig svar på at generell motivasjon predikerer bedre enn yrkesrettet motivasjon. Til slutt viste også studien at sosial kompetanse hadde en prediktiv validitet innenfor simulator (-.16) og ett kriterium som ble satt av instruktørene (.14). Dette omhandlet evnen til å håndtere press, planlegging og koordinering. Spesielt sistnevnte er interessant da samarbeid også er viktig innenfor dette yrket.

2.5.5 Femfaktormodellen (FFM)

Ved gjennomføring av seleksjon til K&V har intervjupanelet brukt FFM som et supplement. Dette for å få et bedre grunnlag når prognose på intervjuet settes, og for å danne seg et raskt bilde av kandidaten, og mulige utfordringer som bør utforskes mer under intervjuet. Siden intervjupanelet bruker denne modellen er det behov for kort å forklare hva FFM innebærer.

Den mest anvendte modellen for personlighet innenfor seleksjon heter FFM, eller «the Big Five». FFM er en modell hvor fem overordnede trekk for personlighet anvendes (Gjestrud, Engetrøen, Stamsø, & Martinsen, 2009). Ved hjelp av fem generelle og sammensatte personlighetsfaktorer skal disse forklare den systematiske variansen i personlighetsvurderingene.

Det finnes flere forskjellige systemer internasjonalt, hvor Forsvaret sin variant er bygget på en norsk utviklet test utviklet av Harald Engvik (Engvik, 1994). Den heter 5PF mil, og er normert med en militærpopulasjon som referanse (Engvik, 1994). Betegnelsene på de fem faktorene basert på den norske testen 5PFmil er Dominans (*Extraversion*), Varme (*Agreeableness*), Kontroll (*Conscientiousness*), Følelser (*Emotional stability*) og Åpenhet (*Intellect eller Openness to new experiences*) (Engvik & Føllesdal, 2005).

Selv om Forsvaret opererer med sin egen versjon av FFM, er det fortsatt i stor grad likheter til tilsvarende engelskspråklige tester. Schroeder, Broach, og Young (1993) gjennomførte en studie med «Big Five» for å se på om hva som kjennetegner ATC personell i forhold til den generelle befolkning. Populasjonen fra befolkning var varierende med alt i fra universitetsutdannede til folk som kun hadde grunnskole. Aldersspennet var mellom 18 og 32 år. Det totale antallet var på nesten 10000 personer. Av ATC personell var det i overkant av 1000 personer. Resultatene viste at det var en forskjell mellom befolkningen generelt og ATC personell. Kontrollørene viste seg å være mer utadvendt, og hadde et høyere behov for søken etter spenning. De utrykte mer positive følelser og var mer pliktoppfyllende enn normen.

3 Seleksjon og utdanningen ved LKVS

I dette kapittelet presenteres hvordan seleksjon inn til K&V blir gjennomført, med vekt på den praktiske gjennomføringen. For å få en oversikt for hvilke personellgrupper som er med i utvalget til undersøkelsen, er det behov for å forklare utdanningssystemet til LKVS. Dette vil også bidra til å skape forståelse for hvorfor noen av kriteriene som blir brukt i undersøkelsen vil ha en varierende N .

3.1 Seleksjon og utdanningsmodell

Seleksjon gjennom LSES inn til K&V har i stor grad vært den samme i hele perioden fra 2002 og frem til i dag, med noen mindre endringer. Det presenteres derfor en generell modell av seleksjonsprosessen. Utdanningssystemet har derimot vært gjennom flere større endringer, og blir presentert med flere modeller.

3.1.1 Seleksjonsprosessen i praksis



Figur 2. Seleksjonsprosessen

Figur 2 viser en beskrivelse av hvordan seleksjonsprosessen blir gjennomført. Uavhengig av om personell blir selektert inn gjennom befalsskole eller krigsskole, må de på ett eller annet tidspunkt gjennom denne seleksjonsprosessen.

Fase 0. Opptakskravene til å søke på utdanning og dermed oppmøte til seleksjon ved LSES består av flere muligheter. Etersom utdanningssystemet til K&V har variert, har også opptakskravene variert. Dette vil bli nærmere forklart i avsnittene for utdanningsmodeller.

Fase 1 består av papir og databaserte tester. Dette testbatteriet er i utgangspunktet utviklet for seleksjon av piloter til Luftforsvaret. Av denne grunn blir ikke alle testene i testbatteriet gjennomført. Testene som blir brukt til denne seleksjonen er testgruppene Generell Teori (GT), Spasiale Evner (SE) og Informasjonsprosessering (IP). GT testene som måler generelt evnenivå, består av Ravens Matriser, Tallrekker, Ordforståelse, Matematikk og Engelsk. SE

og IP testene består av Planes, Instruments, Hands, Attention, Determinasjonsgerât, Digit recall, Numbers og Vigilance. Alle nevnte tester vil bli forklart i kapittelet for metode.

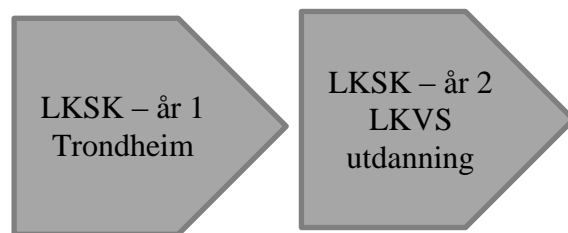
Fase 2, er et semistrukturert intervjupanel med varighet på 30-60 minutt, og fungerer som et supplement til testene. Panelet består av en psykolog, en fra Luftforsvarets seleksjonssenter (LSES) og en representant fra fagmiljøet i K&V. Hensikten er å skaffe seg et inntrykk av kandidaten. Intervjuet vil kunne avdekke motivasjon, holdninger, atferd, modenhet, selvinnsikt og kommunikasjons egenskaper. I tillegg blir 5PFmil brukt for å vurdere personlighetstrekk. Intervjuet blir betraktet som viktig, spesielt for de som gjennomfører befalsutdanning. De som gjennomfører første del av befalsutdanningen kan velge mellom flere utdanningsretninger, herunder K&V. I enkelte åranganger har hele befalsskolekull blitt tatt inn til seleksjon, uavhengig om de ønsker å starte på K&V utdanning eller ikke. Dette kan bety at kandidater som skårer høyt på evnetestene blir tatt inn på utdanningen, selv om de i utgangspunktet ikke ønsker denne utdanningen. Dette må intervjupanelet kartlegge.

Fase 3. Gjennomføring av legesjekk hos Flymedisinsk Institutt (FMI). Her må det oppnås en sertifisering som gir de en gyldig autorisasjon i henhold til gjeldende medisinske krav for luftpersonell. Dersom de ikke består disse kravene kan de ikke bli FC, men de kan fortsatt utdanne seg innen Varsling. I noen situasjoner kan det være slik at intervjuet blir gjennomført før FMI har gjennomført sine undersøkelser. Dersom kandidater ikke består legesjekken hos FMI og intervjuet er gjennomført, er det en pragmatisk tilnærming, hvor kandidat kan få tilbud om utdanning innen Varsling, dersom vedkommende er interessert.

Fase 4. LSES vurderer i samråd med fagmiljøet i K&V hvem som tas opp til utdanning.

3.1.2 Utdanningsmodell før 2007

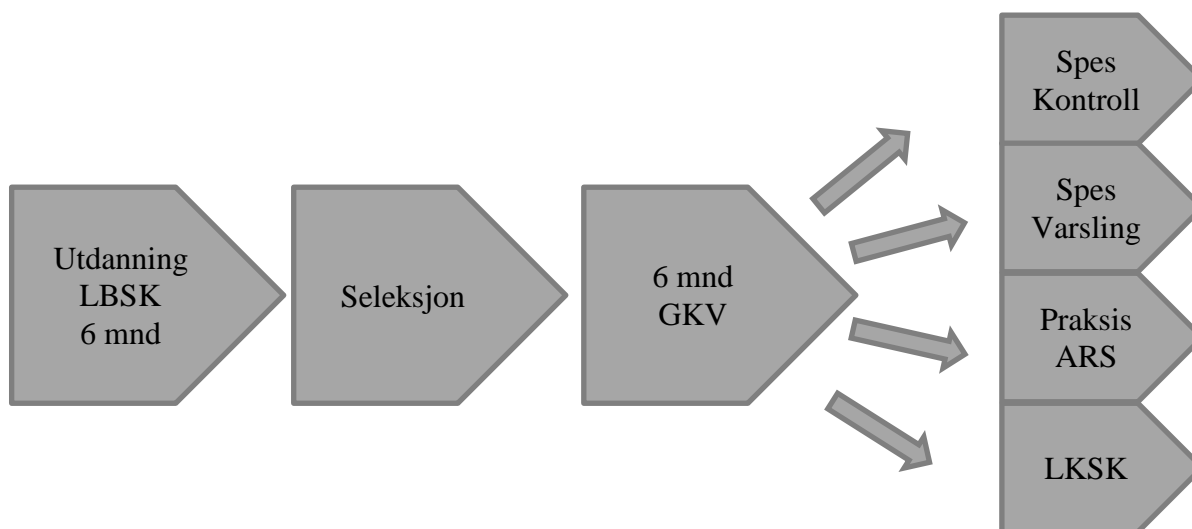
Før 2007 var K&V utdanningen for FC en integrert del av utdanningen gjennom Luftkrigsskolen (LKSK). Det første året ble gjennomført ved LKSK i Trondheim, og det andre året ble gjennomført ved LKVS på Mågerø (figur 3). Alle som gjennomførte utdanningen var testet i forkant av LSES før de startet utdanning på LKSK. Utdanning på LKVS omfattet begge fagfelt, henholdsvis Kontroll og Varsling. Selv om normalløpet for å bli utdannet til FC var gjennom LKSK, var det også mulig for personell som ikke hadde utdanning gjennom LKSK å ta FC utdanningen ved LKVS.



Figur 3. Gammel utdanningsordning

3.1.3 Utdanningsmodell etter 2007

Fra 2007 ble utdanningsnivået for å bli FC justert fra Grunnleggende offisersutdanning¹⁶ (GOU) til grunnleggende befalsutdanning (GBU). GBU er et utdanningsløp med varighet på to år. Hvor det første året, består av et grunnleggende befalskurs (GBK) og en Grunnleggende Fagutdanning (GFU). Det andre året av GBU er praksisperioden. GBK blir gjennomført ved Luftforsvarets Befalsskole Kjevik (LBSK), mens GFU gjennomføres ved Luftforsvarets fagskoler (Utdanningsinspektorat, 2010). I løpet av tiden befalelevne gjennomfører sin GBK utdanning, må elevene søke på hvilken fagutdanning de ønsker å gjennomføre. LKVS sin GFU heter Grunnleggende Kontroll og Varslings (GKV) utdanning, med en varighet på et halvt år.



Figur 4. Normalløpet

¹⁶ For å oppnå GOU status, er normalen å gjennomføre LKSK 3-årig bachelor utdanning. Et annet alternativ kan være det som kalles kvalifiseringskurs.

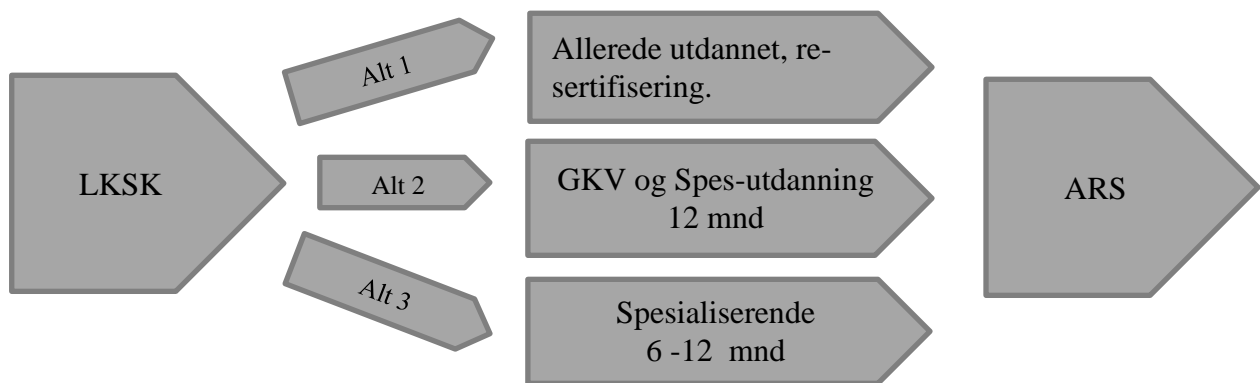
Illustrert i figur 4, ble normalløpet inntak på befalsskole med påfølgende GBK og deretter seleksjon og GKV utdanning på LKVS. Noen år har hele GBK kullet gjennomført opptakstestene for inntak til GKV utdanning. I slike tilfeller har det gjennom intervjuet blitt kartlagt hvem som ønsker, og ikke ønsker utdanning til K&V, av de som klarer testene. GKV modulen er en felles grunnleggende utdanning som er obligatorisk for alt operativt personell innen K&V. Modulen leder frem til den første autorisasjonen, som er et krav for videre trening mot andre autorisasjoner innenfor K&V (Luftforsvaret, 2016). Ved å gjennomføre GKV modulen åpner dette opp for flere alternativer (figur 4). I løpet av GKV modulen vil elevene vurdere hvilke alternativ de ønsker. Dersom GBU løpet skal følges, er alternativet praksis på ARS. Dersom de ønsker spesialisierende utdanning kan de velge spesialistutdanning innenfor Kontroll eller Varsling. Disse to retningene har en varighet på inntil to semester. De som velger kontroll blir da spesialiserte som jagerflykntrolloffiserer. Det siste alternativet er gjennomgående utdanning til LKSK for å få Bachelor.

Selv om GKV modulen er direkte bundet gjennom GFU, kan også andre kandidater enn elever som gjennomfører GBU ta denne utdanningen. Avhengig av behovet til K&V systemet vil det derfor bli lyst ut kunngjøring hvor andre personellkategorier kan søke. Opptakskravet til GKV er GBU eller utdanning som utskrevet befal (UB). I tillegg må samtlige kandidater selekteres gjennom LSES. Det har imidlertid vært noe variasjon i hva som skal være opptakskravet. Dette har resultert i at også personell med fullført førstegangstjeneste har kunnet ta GKV modul, ved å skrive under kontrakt for plikttjeneste. Ikke alle disse har gjennomgått seleksjon ved LSES. Av denne grunn vil det være noe manglende informasjon ved innhenting av data fra LSES sitt testbatteri for noen kandidater.

3.1.4 Utdanning fra krigsskole etter 2006

Med utdanningsløpet nevnt i figur 4, skal i utgangspunktet alle som starter utdanning på LKSK (GOU) ha minimum GKV utdanningen fra før av. Illustrert i figur 5, alternativ 1, vil dette si at når de kommer tilbake til K&V trenger de bare å oppdatere seg, og bli sertifisert på nytt for å utøve jobben sin. I noen tilfeller vil det være at andre kandidater som ikke har denne utdanningen fra før av også kan ta denne utdanningen. Dette kan være kandidater fra andre grener i Luftforsvaret, eller fra Hær og Sjøforsvaret. De vil følge alternativ 2 i figur 5. I tillegg må de gjennom opptaksprøvene ved LSES.

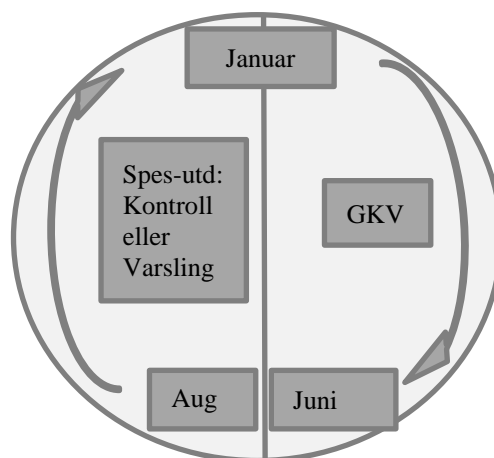
For de som har GKV fra før, men ikke spesialisierende utdanning, følger alternativ 3, i figur 5.



Figur 5. Utdanningsløp fra LKSK.

Unntaksvis har det hendt at kandidater som ikke består testene ved LSES og ikke får tilbud om Kontrollutdanning, har fått tilbud om spesialisierende Varslingsutdanning.

Siden kadetter ved LKSK uteksamineres på våren, starter de opp utdanningen ved LKVS på høsten. Siden GKV utdanningen starter i januar, som en konsekvens av befalsutdanningen, må da kadettene starte opp med sin Kontrollutdanning uten å ha GKV. Siden GKV utdanningen og Kontrollutdanningen er relativt forskjellig, lar dette seg gjøre. Utdanningssyklus ved LKVS er eksemplifisert i figur 6. Det må spesifiseres at spesialistutdanningen vil gå inn i GKV utdanningen. For de som følger alternativ 2 løpet, i figur 5, innebærer dette en kombinerende av spesialistutdanning med GKV utdanning fra januar til juni.



Figur 6. Utdanningssyklus

4 Metode

4.1 Deltakere

Et stort antall av elevene ved LKVS i perioden 2002 og frem til og med 2016 har blitt undersøkt i denne studien. Den spesifikke perioden ble valgt grunnet LKVS flyttet fra Kongsvinger til Mågerø, og startet opp med utdanning ved Mågerø sommeren 2002. Det var derfor naturlig å starte med det første kullet fra LKVS på Mågerø. Tilgangen til data var også enklere å finne fra og med 2002. Totalt ble det samlet inn data fra 218 personer, av disse var det $N = 212$ som kunne benyttes fra LKVS. Kjønnfordelingen for inntak til GKV utdanningen er 13 % kvinner. For de som har gjennomført FC utdanning er det under 5 % kvinner. Alderen på de som har gjennomført utdanningen er mellom 19 og 26 år, og kun unntaksvis er det eldre personell som har gjennomført utdanningen. Til sammen 6 personer hadde sluttet i løpet av utdanningen og disse ble ekskludert fra studien på grunn av ufullstendige resultater. Befalsskolekullene fra 2002 til 2006 ble ikke tatt med, da disse hadde for store variasjoner i utdanning til at det kunne vurderes i sammen med de øvrige kandidatene. I tillegg var de heller ikke innom LSES for seleksjon, siden de fulgte gammel utdanningsordning. Kull fra spesialistutdanningen fra Varsling ble ikke tatt med i problemstillingen, siden dette er en relativt ny utdanning, med svært få elever i hvert kull. Vurdering av andre grunner til frafall blir presentert og vurdert senere i kapitlet.

4.2 Prediktorer

Prediktorer benyttet til denne oppgaven er basert på testene til LSES. Prediktorene er fordelt inn i følgende hovedgrupper: Generell teoriforståelse (GT), Spasiale evner (SE), Informasjonsprosessering (IP) og Intervjuprognose. Testene som måler samme type av evner er slått sammen i nevnte hovedgrupper, basert på tidligere faktoranalyser av testbatteriet (Martinussen & Torjussen, 2004). Innenfor hver hovedgruppe er det spesifikke evnetester som også vil gjelde som prediktorer i denne oppgaven. Dette for å se på om noen enkelttester predikerer bedre enn andre. Oppgaven vil også ta hensyn til allment evnenivå (AE) fra sesjon. Denne prediktoren blir brukt både i forbindelse med seleksjon inn til allmenn verneplikt og befalsskole (Forsvarssjefen, 2008).

For alle prediktorene, foruten Intervjuprognosen, blir det regnet ut resultat fra 1-9, hvor 9 er høyest (stanine skala)(Sundet, Barlaug, & Torjussen, 2004). Domenesnittet har en *cut-off* på

4.0. For de enkelte evnetestene er det 3.0. Dersom kandidaten på en av evnetestene oppnår under 3.0, aksepteres dette ikke. AE uttrykkes også i form av en stanine skala, men er ikke en del av testbatteriet, og følger således ikke disse *cut-off* reglene. Kravet for AE til inntak på befalsskole er for øvrig 5.0 (Forsvarssjefen, 2008).

Prediktorer:

Allment Evnenivå (AE). Testingen utføres på sesjon og består av tre tester; Regneproblemer, Figurregler og Ordlikhet (Skoglund et al., 2015). Regneproblemer måler matematiske ferdigheter og resonneringsevne. Figurregler kartlegger abstrakt resonneringsevne og har likheter med Ravens matriser. Ordlikhet måler ordforståelse (Skoglund et al., 2015). Alle testene har tidsbegrensning. En studie av Sundet, Tambs, Magnus, og Berg (1988) gjennomførte en test-retest for å undersøke reliabiliteten til testene. Resultatene viste henholdsvis .84, .90 og .27 for regneproblemer, figurregler og ordlikhet.

Generell teoriforståelse (GT). Dette hoveddomenet består av flere papir- og blyant tester. Gjennomsnittet av disse testene utgjør skåren for GT.

Raven: Dette er en Raven sin *Advanced Progressive Matrices* (Raven, 1941), og er en velkjent intelligens test hvilket måler induktive slutninger. Testen måler ikke-verbale aspekter ved generell intelligens. Hver testoppgave inneholder en matrise med figurer, hvor en av figurene mangler. Under matrisen er det flere deler som kan passe inn i matrisen, hvor bare en del vil være den riktige. I hver enkel deltest blir kandidaten spurt om å identifisere det manglende element som gjør mønsteret komplett (Martinussen & Torjussen, 1998).

Tallrekker er en test som inneholder en logisk rekke med tall, hvor oppgaven er å fylle in de to siste tallene i serien (Martinussen & Torjussen, 1998).

Ordforståelse er en verbal test hvor kandidaten blir presentert med grupper av tre ord og blir spurt om å finne det fjerde ordet som har en relasjon til alle nevnte ord (Martinussen & Torjussen, 1998).

Matematisk forståelse og engelsk kunnskaper: Dette er to forskjellige tester som måler de generelle kunnskapene innenfor disse to områdene (Martinussen & Torjussen, 1998).

Spatiale evner (SE). Dette hoveddomenet består av flere tester. Gjennomsnittet av disse testene utgjør skåren for SE.

Planes (640). Dette er en spatial evnetest som innebærer visualisering og mental rotering. Kandidaten må tenke seg at vedkommende sitter i et flysete. Et bilde av et fly blir vist i ett av hjørnene på skjermen. Kandidatens oppgave blir da å avgjøre så raskt som mulig hvordan spaken må beveges slik at flyet flyr i riktig retning. (Martinussen, Torjussen, Storsve, & Hjerkin, 2004).

Instrument Comprehension (INSB-650). Testen måler generell og spatial resonneringsevne. Dette innebærer å vise en evne til å visualisere hvordan flyet er i det tredimensjonale rom. Kandidaten blir presentert seks forskjellige typer instrument i et fly som blir grundig forklart. Kandidaten må således identifisere den korrekte orienteringen av flyet fra et gitt bilde. (Martinussen et al., 2004).

Hands (660). Testen gir kandidaten informasjon på øret som må brukes som søkekriterium. Kandidaten må deretter søke etter visuelle objekt på skjermen, og bestemme seg for hvor mange av de presenterte objektene som stemmer overens med mottatt melding. Oppgaven er derfor primært en test av arbeidsminne, samt hvor raskt noen kan bevege seg fra verbalt til visuelt materiale og foreta en riktig avgjørelse (Martinussen et al., 2004).

Informasjonsprosessering (IP). Dette hoveddomenet består av flere tester. Gjennomsnittet av disse testene utgjør skåren for IP.

Attention (670). Denne testen krever hurtig forståelse av opp-ned og høyre-venstre dimensjonen. En 4x4 matrise med forskjellige farger og tall i hver kvadrant er presentert med en muntlig beskjed fortalt gjennom hodetelefon. Ved hjelp av beskjeden skal kandidaten identifisere hvilken kvadrant som skal fokuseres på i følge bestemte regler. (Martinussen et al., 2004).

Digit recall (680). Dette er en korttidshukommelse test. Kandidaten blir presentert et sett av nummer med forskjellig størrelse. Med en gang nummeret forsvinner skal kandidaten skrive inn nummeret (Martinussen et al., 2004).

Determinasjonsgerat (DTG)-690. Måler reaktiv stress toleranse, oppmerksomhet og reaksjonshastighet. Dette gjøres ved hurtig forandring og kontinuerlig optisk og akustisk stimuli. Kandidaten blir presentert en skjerm, hvor forskjellige farger plutselig dukker opp. I tillegg kommer det lyd gjennom hodetelefon. Ettersom farger dukker opp på skjermen og lyd gjennom hodetelefon, må kandidat klikke korrekt på tastatur og pedaler i henhold til gitte instruksjoner gitt i forkant av test (Martinussen et al., 2004).

Numbers (700). Denne testen måler persepsjonshastighet og er knyttet til rask, nøyaktig og effektiv bruk av visuell persepsjon, i tillegg til arbeidsminne. Et stort antall tall av forskjellig størrelse er spredt over hele skjermen og presentert fire ganger. Oppgaven er, innenfor begrenset tid, å finne så mange tall i en rekkefølge som mulig. Eksempelvis fra 21 til 24 (Martinussen et al., 2004).

Vigilance (720). Testen måler oppmerksomhet og kandidaten er presentert med en 9x9 matrise. Hver celle i matrisen kan bli identifisert av nummer langs venstre side og toppsiden av matrisen. Kandidaten utfører fortløpende spesifiserte oppgaver, hvor noen er rutine oppgaver, mens andre er av høyere prioritet (Martinussen et al., 2004).

Intervjuprognozen. Prognosen opererer med en skala fra 1-6. 1 er ikke anbefalt. 2 er anbefaling under tvil. 3 og høyere er graderinger av anbefalt.

Til slutt kommer det en prediktor som er spesifikt laget for problemstilling nummer 3. Problemstillingen er om resultat fra Varsling og GKV kan predikere utdanningsresultat på Kontroll. Siden det eksisterer utdanningsresultat innenfor Varsling tilbake til 2002, er det mulig å kombinere disse resultatene med Varslingsresultat fra GKV utdanningen. Denne prediktoren vil få navnet: Total Varsling Praktisk (TotalVP).

4.3 Kriterier

Kriterier er valgt ut både fra GKV og Kontrollutdanningen fra LKVS. Fra GKV er det plukket ut teoretiske og ett praktisk fag. I tillegg ble hovedkarakter fra hele GKV utdanningen valgt ut. Fra Kontrollutdanningen ble det valgt tre praktiske og ett teoretisk fag. Dette for å se på sammenhenger innenfor både de teoretiske og praktiske fagene opp mot prediktorene. Det ble lagt vekt på de praktiske fagene, da det er de fagene som ligner mest på hva de skal jobbe med etter endt utdanning. LKVS opererer med fagplaner innenfor alle utdanningsområder.

Beskrivelsene av kriteriene er derfor plukket ut fra disse fagplanene. Utdanningen fra 2002 og frem til 2016 har utviklet seg. Dette er naturlig både på grunn av prosedyrer og direktiv som forandrer seg. I tillegg har det blitt mer normalt med bruk av simulator under utdanningen. Fagene i utdanningen har også endret seg noe. Noen fag har blitt splittet opp, og noen fag har blitt slått sammen. Det har også dukket opp nye fag underveis. For å finne mest mulig korrekte kriterier har det vært behov for å studere hvert enkelt fag, for å se på hva som er sammenlignbart fra år til år. I tillegg har det også vært behov for å slå sammen noen fag, for å samle de under samme kriterium. For mange av kriteriene var det ikke mulig å slå sammen fagene fra LKSK i perioden fra 2002 til 2006, sammen med utdanningen etter 2006. For denne undersøkelsen er det derfor bare ett kriterium for Kontrollutdanningen, som har hele utvalget fra 2002 til 2016. Dette forklares nærmere etter alle kriterier er forklart.

Det er spesifikt beskrevet hvilket kriterium som inneholder kombinert resultat fra før og etter 2006. Alle kriterier opererer med eksamenskarakterer fra A-F. Disse karakterene er gjort om fra bokstavkarakterer til tallkarakterer fra 1 til 6, hvor 6 er best. Navnet på kriteriet sier hvilken utdanning det er tatt fra. Dersom GKV står først, indikerer dette at resultater kun er hentet fra GKV utdanningen fra 2007 til 2016. Dersom kriteriet starter med Kontroll, indikerer dette at resultater er hentet fra Kontroll utdanningen fra 2007 til 2016. Dersom kriteriet ikke starter med verken GKV eller Kontroll, innebærer dette at resultatene er hentet fra hele perioden 2002 til 2016.

Kriterier

GKV, Operative prosedyrer og bestemmelser (GOPB): Dette er et teoretisk fag som har operert med forskjellige betegnelser i perioden fra 2002 til 2016. I noen årganger har faget blitt delt opp i flere fag, med eksamensresultat i hvert fag. Andre årganger har fagene blitt slått sammen til ett stort fag, med felles eksamensresultat. Alle fag ble derfor gjennomgått for å se på faginnhold. Deretter ble de fag som inneholdt i stor grad samme pensum, slått sammen til kriteriet OPB. Dette kriteriet inneholder prosedyrer og bestemmelser innenfor både nasjonale og NATO publikasjoner. Kriteriet gjelder for alle kull fra og med 2007.

GKV, Systemforståelse (GSFT). Faget gir en forståelse for de systemene K&V bransjen bruker i det operative miljøet. Dette er et fag som gir eleven et teoretisk grunnlag for hvordan systemene fungerer praktisk. Dette faget har også operert med forskjellige navn og

forskjellige fagkombinasjoner under seg. Det har også i dette tilfellet vært behov for å slå sammen noen fag. Eksamen blir gjennomført som en ordinær skoleprøve. Kriteriet gjelder for alle kull fra og med 2007.

GKV, Varsling Praktisk (GVP). Dette faget er først og fremst et praktisk fag, som omhandler praktisk forståelse og bruk av systemer innenfor Varsling. Eksamen er en praktisk evaluering med bruk av simulator. Kriteriet gjelder for alle kull fra og med 2007.

GKV Hovedkarakter (GKVH). Dette kriteriet er snittresultat for hva kandidatene oppnådde gjennom GKV utdanningen. GKV utdanningen består av flere fag enn hva som er nevnt i de tre foregående kriteriene. Dette kan være spesifikke fag innenfor kommunikasjonssystemer, radarteori, metrologi og lufttrafikkteneste. Kriteriet gjelder for alle kull fra og med 2007.

Kontroll, Military Flying Regulations (KMFR). Dette er et fag som handler om hvilke bestemmelser og krav som stilles til personell innenfor K&V, med fokus på Kontroll. Faget gir elever kunnskap om hvilke bestemmelser som gjelder for sivil og militær lufttrafikk. Eksamen blir gjennomført som en ordinær skoleprøve. Kriteriet gjelder ikke for LKSK kullene (2002-2006), da fagkombinasjon og innhold i utdanningen for denne perioden, gjorde det vanskelig å kombinere med dette kriteriet.

Kontroll, Air Combat training (KACT). Faget utvikler elevenes kunnskaper og ferdigheter innen luftstrid i begrensede scenarioer. I praksis vil dette si at FC må gi fortløpende oppdateringer til sine piloter om hvor motstridende fly opererer. Det må også fortløpende tas hensyn til engasjementsregler for når egne piloter kan engasjere motstridende fly. Når visse kriterier oppfylles, kan FC gi ordre om engasjement. I tillegg er det mye annen informasjon i fra andre aktører som også må tas hensyn til. Dette kan være oppdateringer fra ATC og kontinuerlige endrede forhold i hva som skjer i luftrommet. Dette å være seg endrede begrensninger til luftrom, eller annen lufttrafikk i og utenfor tildelt luftrom. Faget er både et teoretisk og praktisk fag med hovedvekt på den praktiske utførelsen. Faget kombinerer bruken av teoretisk kunnskap innbakt i praktisk utførelse. Den praktiske delen utføres hovedsakelig simulert, men også utføres under «live» trening avhengig av tilgjengelighet på oppdrag. Eksamen gjennomføres med en praktisk evaluering av kandidat i simulator. På grunn av nye simulatormuligheter, ble dette faget utviklet og gjennomført fullt ut ved LKVS med

tilhørende eksamen for første gang i 2007. Faget legger opp til mye egentrening på kveldstid. LKSK kullene (2002-2006) er derfor ikke med i dette kriteriet.

Kontroll, Air Policing 1(KAP1). *Air Policing* er ivaretagelsen av NATO sitt luftrom i fredstid og skal bidra til NATO sitt kollektive forsvar. *Air Policing* innebærer å ha *Quick Reaction Alert* (QRA) jagerfly på kontinuerlig beredskap 24/7 (Command, 2016). Disse to jagerflyene skal respondere på fly som opererer uidentifisert eller på en uvanlig måte. Gjennom overvåking av luftrommet vil slike typer fly eller objekt bli oppdaget. Oppdraget til FC er dermed å lede jagerflyene til objektet og avskjære og identifisere det. Faget KAP1 skal gi elever kunnskaper og ferdigheter innenfor jagerflykontroll, med hovedvekt på avskjæringsteknikker. Faget er både et teoretisk og praktisk fag med hovedvekt på den praktiske utførelsen. Den praktiske utførelsen består hovedsakelig av simulator trening. Avhengig av progresjon vil også elev få prøve seg med «live» trening. Eksamen gjennomføres med en praktisk evaluering av kandidat. Faget legger opp til mye egentrening på kveldstid og fritid. Dette er den første av to eksamener. Eksamen blir gjennomført ved halvveis løp i utdanningen. Det var ikke mulig å innhente data fra LKSK (2002-2006) kullene for første eksamen.

Kontroll, Air Policing 2 (KAP2). Andre og siste eksamen i AP faget. Innholdet i faget er det samme, men med strengere krav til gjennomføring av eksamen. Data er hentet fra 2007 til 2016.

Air Policing før og etter 2006 (AP). *Air Policing* eksamen for LKSK kullene fra 2002 til 2006, er så å si helt lik siste eksamen til nåværende Kontrollutdanning for *Air Policing*. I dette kriteriet er derfor AP eksamen for LKSK kullene kombinert med siste eksamen i *Air Policing* (KAP2) for Kontrollutdanningen etter 2006. Dette kriteriet har dermed et større utvalg.

Etter gjennomgang av alle fag fra 2002 til 2016 var det bare ett kriterium som kunne brukes for hele perioden. Flere forhold forklarer denne beslutningen. Utdanningen frem til 2006 hadde bare ett praktisk fag, *Air Policing*, hvor det ble gjennomført eksamen. De andre teoretiske fagene i denne utdanningen var såpass forskjellig fra fagene etter 2006, at det ble vanskelig å kombinere med det teoretiske faget med kriteriet KMFR. Det ble vurdert om det kunne velges kriterium kun for utdanningsperioden fra 2002-2006. Utvalget fra denne perioden var så lavt, at det ikke var hensiktsmessig.

4.4 Vurdering av frafall

En kvantitativ undersøkelse var nødvendig for denne undersøkelsen for å få et representativt bilde av populasjonen. Det er imidlertid noen av prediktorene og kriteriene som opererer med lavere N , enn andre kriterier. Av det totale antallet ($N = 218$) innhentet ved LKVS, var det $N = 212$ observasjoner som kunne benyttes. LKSK kullene (2002-2006) har en $N = 32$. Dette betyr at for kriteriene hvor LKSK kullene ikke er tatt med i kriteriet, er total $N = 180$. Kullene som har gjennomført kontrollutdanning har et langt mindre utvalg ($N = 92$). For de kriteriene hvor kun kontrollresultat fra og med 2007 er tatt med, er utvalget $N = 60$.

Kriterier som i utgangspunktet skal ha $N = 180$, er GOPB, GSFT, GVP og GKVH. GOPB ($N = 172$) mangler 8 observasjoner. Etter 2006 har det vært 8 elever fra LKSK som har begynt utdanning ved LKVS som ikke har gjennomført dette faget, da det har vært dekket i pensum av LKSK. GSFT ($N = 178$), mangler 2 observasjoner. Dette kan i utgangspunktet ikke forklares da alle skal ha gjennomført dette. En plausibel forklaring kan være at kandidater ikke har gjennomført eksamen, og av ukjent årsak ikke tatt eksamen i etterkant. GVP ($N = 179$) og GKVH ($N = 179$) mangler en observasjon, årsak ukjent. Kriteriene for KMFR ($N = 60$), KACT ($N = 60$), KAP1 ($N = 60$), og AP ($N = 92$) mangler ingen observasjoner. KAP2 ($N = 59$), mangler en observasjon, årsak ukjent.

For prediktorene varierer tallet på observasjoner fra $N = 132$ til 210. Hovedgrunnen til dette er på grunn av opptakskravet til GKV utdanning. Siden opptakskravet til GKV utdanning ikke alltid har krevd seleksjon gjennom LSES, har dette medført at noen ikke har gjennomført testene ved LSES. I tillegg er data som skal hentes inn mer enn 10 år tilbake i tid vist seg å være utfordrende å finne frem i. De er lokalisert på diverse Excel ark og permer som kan være vanskelig å finne. Dette betyr at for de seneste årene har det vært enklere for LSES å innhente data enn mer enn 10 år tilbake i tid. Det er relativt tilfeldig hva de har funnet data på innenfor de ulike evnetestene. Intervjuprognosen har lavest antall observasjoner med $N = 116$. Dette forklares med at ulikt personell har lagret disse dataene på forskjellig vis. Nyere data om intervjuprognosene har vært enklere å oppdrive enn eldre intervjuprognoser. Det er liten grunn til å tro at datagrunnlaget har systemiske frafall for prediktorene. Av den informasjonen som er innhentet virker det tilfeldig av hva de har klart å oppdrive.

4.5 Prosedyre

Prosjektet er godkjent av Norsk samfunnsvitenskapelig datatjeneste (NSD). I samarbeid med LKVS, ble alle utdanningsresultat som var mulig å oppdrive samlet inn. Eksamenskarakterer var en blanding av tall og bokstavkarakterer. Alt ble derfor gjort om til en felles karakterstruktur. Alle seleksjonsresultat ble hentet inn fra LSES, som oppbevarer seleksjonsdata. Siden innhenting av data gikk helt tilbake til 2002, var det noe krevende å finne frem alle data. Noe data var således ikke mulig å oppdrive. Alle resultat ble deretter sendt inn til sjefspsykologen for Forsvaret som koblet testene og skoleresultatene. Datafilen ble deretter anonymisert, før analysene ble gjennomført.

4.5.1 Statistiske analyser

Samtlige statistiske analyser er utført ved hjelp av *Statistical Package for the Social Sciences* (SPSS, versjon 24). Programmet ble brukt for å beregne korrelasjoner mellom seleksjonstestene og utdanningsresultatene på LKVS. Pearsons r er valgt som korrelasjonsmål (Ringdal, 2013, s. 304). Pearsons r kan gi svar på samvariasjonen og styrken mellom to variabler. I tillegg kan den si noe om retningen på samvariasjonen (Jacobsen, 2015, s. 335). Verdiene for Pearsons r varierer mellom -1 og $+1$. Ifølge Jacobsen (2015) finnes det ikke noe perfekt svar på hva som er en sterk korrelasjon i samfunnsvitenskapen, og foreslår følgende retningslinjer basert på Cohen (1988); svak korrelasjon $r = .10$, middels korrelasjon $r = .30$, sterk korrelasjon $r = .50$. Ved bruk av korrelasjonsanalyse er det ikke nok bare å se på styrken mellom variablene, også signifikansen må vurderes. Signifikans sier noe om hvor troverdige resultatene er, og sannsynligheten for signifikante resultater øker med utvalget (Jacobsen, 2015, s. 335). I denne undersøkelsen er signifikans-nivået satt til 5 %.

For å se på sammenhengen mellom flere uavhengige variabler (prediktor) og avhengig variabel (kriterium) er det behov for å bruke multippel hierarkisk regresjonsanalyse. Den multiple korrelasjonskoeffisienten (R^2) angir forklart varians for de ulike trinnene i analysen. I tillegg kan den forklare den totale variansen for prediktorene (Ringdal, 2013, s. 415-449). I analyser med bruk av regresjon ble det brukt *pairwise missing*.

5 Resultat

5.1 Korrelasjonsanalyse

Tabellene gir oversikt over deskriptiv statistikk og korrelasjoner mellom seleksjonstestene og kriteriene. For problemstilling 1, vil korrelasjonsanalysen i tabell 1 og 2 være gjeldende. Det ble utformet to tabeller, og kriteriene for begge tabeller vil være de samme. Tabell 1, inneholder alle enkelttestene som LSES gjennomfører i tillegg til AE og Intervjuprognosen. Tabell 2, inneholder domenegruppene GT, SE og IP. I tillegg ble AE tatt med, da det blir enklere å se om det er forskjeller spesielt mellom AE og GT. For problemstilling 2, vil korrelasjonsanalysen i tabell 3 og 4 være gjeldende, det ble også her utformet to tabeller. Prediktorene er de samme som for tabell 1 og 2, mens andre kriterier er nyttet. For problemstilling 3, vil korrelasjonsanalysen i tabell 5 være gjeldende. Prediktorer er tatt fra GKV utdanningen. I tillegg er det en prediktor, som kombinerer Varslingsresultat fra 2002 til 2016. Kriteriene er fra Kontrollutdanningen.

Tabell 1: Gjennomsnitt, standardavvik og bivariate korrelasjoner mellom prediktorer og kriterier innhentet i første del av utdanningen ($N = 132 - 179$)

Variabler	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
Prediktorer																					
1. AE	6.0	1.3	-																		
2. Raven	5.4	1.8	.46**	-																	
3. Tall	4.9	1.7	.53**	.56**	-																
4. Regning	5.2	1.8	.47**	.44**	.50**	-															
5. Engelsk	5.9	2.0	.38**	.26**	.18*	.08	-														
6. Ordforståelse	5.3	1.7	.35**	.14	.25**	.13	.53**	-													
7. Planes	4.4	2.1	.36**	.43**	.14	.29**	.21*	.06	-												
8. Instrument C	4.6	2.2	.31**	.38**	.24**	.27**	.21*	.07	.48**	-											
9. Hands	5.2	2.0	.30**	.32**	.11	.21*	.23**	.15	.43**	.29**	-										
10. Attention	5.2	2.2	.43**	.43**	.31**	.33**	.29**	.21*	.40**	.31**	.50**	-									
11. DTG	4.9	1.8	-.04	-.02	-.12	-.06	.01	.08	.08	.08	.14	.19*	-								
12. Digit Recall	4.8	1.8	.21*	.21*	.21*	.18*	.14	.11	.20*	.30**	.30**	.27**	.00	-							
13. Numbers	5.5	1.8	-.01	.18*	.15	.06	-.09	-.07	-.11	-.02	-.06	.07	.14	.01	-						
14. Vigilance	5.0	1.8	.20*	.17*	.07	.13	.10	.04	.23**	.27**	.23**	.28**	.27**	.22**	-.01	-					
15. Intervjuprog. ^a	4.1	1.0	.37**	.26**	.29*	.35**	.23*	.24*	.26*	.25*	.13	.30**	.07	.06	.10	.03	-				
Kriterier																					
16. GOPB	4.4	1.0	.21**	.20*	.21*	.25**	.25**	.13	.03	.17	.00	.05	.01	.02	.09	-.01	.23*	-			
17. GSFT	4.6	1.0	.23**	.23**	.15	.10	.26**	.12	-.05	.00	.02	.10	.02	.04	.06	.01	.10	.59**	-		
18. GVP	4.7	1.0	.29**	.20**	.16*	.11	.27**	.09	.04	.13	.14	.15	-.06	-.12	.03	.01	.21*	.44**	.40**	-	
19. GKVH	4.5	0.8	.22**	.27**	.20*	.18*	.33**	.10	.02	.10	.02	.14	-.03	-.04	.04	.10	.18	.75**	.70**	.64**	-

Note. ^a $N = 116$. GOPB = GKV Operative Prosedyrer og bestemmelser; GSFT = GKV Systemforståelse; GVP = GKV Varsling Praktisk; GKVH = GKV Hovedkarakter. * $p < .05$, ** $p < .01$, *** $p < .001$ (to-halet).

I tabell 1 er det AE, Raven og Engelsk som oppnår høyest signifikante korrelasjoner med kriteriene. I tillegg korrelerer Tallserier og Regning signifikant med enkelte kriterium.

Tabell 1 viser at AE korrelerer signifikant med alle kriteriene GOPB, GSFT, GVP og GKVH fra $r = .21^{**}$ til $r = .29^{**}$. Av de spesifikke evnetestene korrelerer Raven og Engelsk signifikant med alle kriterier fra $r = .20^*$ til $r = .33^{**}$. Tall korrelerer med tre av kriteriene GOPB, GVP og GKVH fra $r = .16^*$ til $r = .21^*$. Regning korrelerer to av kriteriene GOPB og GKVH med henholdsvis $r = .25^{**}$ og $r = .18^*$. I tillegg korrelerer Intervjuprognosen signifikant med både GOPB faget og GVP med $r = .23^*$ og $r = .21^*$. Det er ingen signifikante korrelasjoner mellom testene innenfor SE og IP og kriteriene i tabell 1.

Tabell 2: Gjennomsnitt, standardavvik og bivariate korrelasjoner for grupper av prediktorer og kriterier innhentet under første del av utdanningen ($N = 139-179$).

Variabler	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
Prediktorer										
1. AE	6.0	1.3	-							
2. Generell Teori	5.3	1.2	.65**	-						
3. Spasiale evner	4.7	1.6	.41**	.43**	-					
4. Info.Pross.	5.1	1.1	.31**	.36**	.45**	-				
Kriterier										
5. GOPB	4.4	1.0	.21**	.31**	.09	.06	-			
6. GSFT	4.6	1.0	.23**	.26**	-.01	.08	.59**	-		
7. GVP	4.7	1.0	.29**	.25**	.14	.02	.44**	.40**	-	
8. GKVH	4.6	0.8	.22**	.32**	.05	.09	.75**	.70**	.64**	-

Note: GOPB = GKV Operative Prosedyrer og bestemmelser; GSFT = GKV Systemforståelse; GVP = GKV Varsling Praktisk; GKVH = GKV Hovedkarakter. $*p < .05$, $**p < .01$, $***p < .001$ (to-halet).

I tabell 2 korrelerer Både AE og GT signifikant med alle kriterier. AE sine korrelasjoner er allerede forklart i tabell 1. GT korrelerer signifikant fra $r = .25^{**}$ til $r = .32^{**}$.

Tabell 3: Gjennomsnitt, standardavvik og bivariate korrelasjoner mellom prediktorer og kriterier til kontrollutdanningen ($N = 59$ til 210).

Variabler	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
Prediktorer ^a																						
1. AE	6	1.3	-																			
2. Raven	5.4	1.8	.46**	-																		
3. Tall	4.9	1.7	.53**	.56**	-																	
4. Regning	5.2	1.8	.47**	.44**	.50**	-																
5. Engelsk	5.9	2.0	.38**	.26**	.18*	.08	-															
6. Ordforståelse	5.3	1.7	.35**	.14	.25**	.13	.53**	-														
7. Planes	4.4	2.1	.36**	.43**	.14	.29**	.21*	.06	-													
8. Instrument C	4.6	2.2	.31**	.38**	.24**	.27**	.21*	.07	.48**	-												
9. Hands	5.2	2.0	.30**	.32**	.11	.21*	.23**	.15	.43**	.29**	-											
10. Attention	5.2	2.2	.43**	.43**	.31**	.33**	.29**	.21*	.40**	.31**	.50**	-										
11. DTG	4.9	1.8	-.04	-.02	-.12	-.06	.01	.08	.08	.08	.14	.19*	-									
12. Digit Recall	4.8	1.8	.21*	.21*	.21*	.18*	.14	.11	.20*	.30**	.30**	.27**	.00	-								
13. Numbers	5.5	1.8	-.01	.18*	.15	.06	-.09	-.07	-.11	-.02	-.06	.07	.14	.01	-							
14. Vigilance	5.0	1.8	.20*	.17*	.07	.13	.10	.04	.23**	.27**	.23**	.28**	.27**	.22**	-.01	-						
15. Intervjuprog.	4.1	1.0	.37**	.26**	.29**	.35**	.23*	.24*	.26	.25*	.13	.30**	.07	.06	.10	.03	-					
Kriterier ^b																						
16. KMFR	4.4	0.9	.22	.10	.39**	.62**	.31*	.39**	-.09	.03	.03	.21	-.11	.27	-.11	.08	.11	-				
17. KACT	4.4	1.2	.00	-.17	.04	.19	.20	.04	-.11	-.04	.04	.15	.06	.12	-.02	.20	-.29	.38**	-			
18. KAP1	4.3	1.3	.29*	.23	.20	.03	.19	.12	.13	.22	-.01	.04	-.21	.03	.03	.12	.12	-.09	-.03	-		
19. KAP2	4.5	1.1	.05	-.01	.06	.08	-.24	-.16	.02	-.09	-.06	-.15	.14	.00	-.11	.13	-.09	.02	-.09	.18	-	
20. AP	4.5	1.0	-.04	-.04	-.01	.00	-.20	-.11	.02	-.09	-.06	-.15	.14	.00	-.11	.13	-.03	.02	-.09	.18	1.00**	-

Note: ^a $N = 139$ - 210 . ^b $N = 59$ - 92 . KMFR = Kontroll Military Flying Regulations; KACT = Kontroll Air Combat Training; KAP1 = Kontroll Air Policing eksamen 1; KAP2 = Kontroll Air Policing eksamen 2; AP = Air Policing fra 2002 til 2016. * $p < .05$, ** $p < .01$, *** $p < .001$ (to-halet).

For kriterier i tabell 3, er det kun AP ($N = 92$) som har hele utvalget fra 2002 til 2016. Resterende kriterier har $N = 59$ - 60 .

I tabell 3 korrelerer AE, Tall, Regning, Engelsk og Ordforståelse signifikant med enkelte kriterier. KMFR ble predikert av Tallserier, Regning, Engelsk og Ordforståelse, fra $r = .31^*$ til $r = .62^{**}$. Tilsvarende ble KAP1 predikert av AE ($r = .29^*$). Det eneste kriteriet med innhentet data fra 2002 til 2016 er AP kriteriet. Selv om dette kriteriet har større utvalg får det ingen signifikante korrelasjoner.

Tabell 4: Gjennomsnitt, standardavvik og bivariate korrelasjoner for grupper av prediktorer og kriterier innhentet for variablene til Kontrollutdanningen ($N = 59-210$).

Variabler	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9
Prediktorer ^a											
1. AE	6.0	1.3	-								
2. Generell Teori	5.3	1.2	.65**	-							
3. Spasiale evner	4.7	1.6	.41**	.43**	-						
4. Info.Pross.	5.1	1.1	.31**	.36**	.45**	-					
Kriterier ^b											
5. KMFR	4.4	.9	.22	.52*	-.04	.14	-				
6. KACT	4.4	1.2	.00	.07	-.06	.20	.38**	-			
7. KAP1	4.3	1.3	.29*	.23	.14	.02	-.09	.03	-		
8. KAP2	4.4	1.1	.05	-.08	-.08	.01	.02	-.09	.18	-	
9. AP	4.5	1.0	-.04	-.12	-.08	.01	.02	-.09	.18	1.00**	-

Note: ^a $N = 139-210$. ^b $N = 59-92$. KMFR = Kontroll Military Flying Regulations; KACT = Kontroll Air Combat Training; KAP1 = Kontroll Air Policing eksamen 1; KAP2 = Kontroll Air Policing eksamen 2; AP = Air Policing fra 2002 til 2016. * $p < .05$, ** $p < .01$, *** $p < .001$ (to-halet).

I tabell 4 er det to signifikante korrelasjoner mellom prediktorer og kriterier. AE korrelerer signifikant med KAP1 med $r = .29^*$, og GT korrelerer signifikant med KMFR med $r = .52^*$.

Tabell 5: Gjennomsnitt, standardavvik og bivariate korrelasjoner for grupper av prediktorer fra GKV og Varslingsutdanning for variablene til Kontrollutdanningen ($N=59-210$).

Variabler	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8	9	10
Prediktorer												
1. GOPB	4.6	.8	-									
2. GSFT	4.4	1.0	.59**	-								
3. GVP	4.7	1.0	.44**	.40**	-							
4. Total VP	4.7	1.0	.44**	.40**	1.00**	-						
5. GKVH	4.6	1.0	.75**	.70**	.64**	.64**	-					
Kriterier												
6. KMFR	4.4	.9	.33**	.49**	.05	.05	.12	-				
7. KACT	4.4	1.2	.10	.35**	.09	.09	.35**	.38**	-			
8. KAP1	4.3	1.3	.15	.19	.18	.18	.27*	-.09	-.03	-		
9. KAP2	4.4	1.1	.29*	.19	.15	.15	.17	.02	-.09	.18	-	
10. AP	4.5	1.0	.29*	.19	.15	.12	.17	.02	-.09	.18	1.00**	-

Note: KMFR = Kontroll Military Flying Regulations; KACT = Kontroll Air Combat Training; KAP1 = Kontroll Air Policing eksamen 1; KAP2 = Kontroll Air Policing eksamen 2; AP = Air Policing fra 2002 til 2016. * $p < .05$, ** $p < .01$, *** $p < .001$ (to-halet).

I tabell 5 presenteres korrelasjoner mellom oppnådde prestasjoner i GKV utdanningen og Kontrollutdanningen. I tillegg er prediktoren Total Varsling Praktisk (TVP) tatt med. Som tidligere nevnt i metode kapittelet, er TVP en kombinasjon av oppnådde Varslingsresultat fra 2002 til 2016. Siden det er kun ett kriterium (AP) for Kontrollutdanningen, hvor hele utvalget er tatt med fra 2002 til 2016, er det korrelasjonen mellom TVP og AP som er interessant å se på. Korrelasjon mellom TVP og AP var ikke signifikant.

GKV Hovedkarakter (GKVH) korrelerer signifikant med KACT og KAP1 med $r = .35^{**}$ og $r = .27^*$. GKV Operative Prosedyrer og Bestemmelser (GOPB) korrelerer signifikant med

KMFR, KAP2 og AP fra $r = .29^*$ til $r = .33^{**}$. GKV Systemforståelse (GSFT) korrelerer signifikant med KMFR og KACT med $r = .49^{**}$ og $r = .35^{**}$.

5.2 Multippel hierarkisk regresjonsanalyse

For å se på i hvilken grad de forskjellige prediktorene samlet sett kan predikere utdanningsresultat, ble det gjennomført hierarkisk multippel regresjonsanalyser. Selv om mange av prediktorene i korrelasjonsanalysen ikke viste noen signifikante verdier, blir de tatt med i regresjonsanalysen. I tabell 6 og 7, er prediktorene delt opp stegvis i de forskjellige gruppene av tester. AE ble plassert inn på samme trinn som GT, siden det er i samme gruppe av evner som blir testet. SE og IP ble satt inn i to forskjellige trinn. Intervjuprognose ble satt inn til slutt for seg. I tabell 8, ble oppnådde resultat i GKV brukt som uavhengige variabler, for å se om de kan predikere noe for oppnådde resultat i Kontroll.

I den første regresjonsanalysen er det prediksjon av oppnådde resultat i henholdsvis OPB, Systemforståelse, Varsling og GKV som er undersøkt.

Tabell 6. Hierarkisk multipl regressjonsanalyse for oppnådde prestasjoner i GKV.

Prediktor	GOPB		GSFT		GVP		GKVH	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Steg 1:	.10**		.07*		.09**		.11**	
AE		.01		.14		.22		.04
GT		.29*		.24		.09		.32*
Steg 2:	.00		.02		.00		.01	
S.E		-.05		-.18		.03		-.12
Steg 3:	.00		.00		.01		.00	
I.P		-.05		.04		-.12		.00
Steg 4:	.01		.00		.01		.00	
Intervjuprognose		.12		-.01		.10		.06
Total R^2	.11*		.10		.11*		.12*	
N	172		178		179		179	

Note: Alle koeffisienter er tatt fra siste steg i regressjonsanalysen. GOPB = GKV Operative Prosedyrer og bestemmelser; GSFT = GKV Systemforståelse; GVP = GKV Varsling Praktisk; GKVH = GKV Hovedkarakter. * $p < .05$, ** $p < .01$, *** $p < .001$.

Tabell 6 viser at steg 1 med AE og GT, predikerte alle kriterier signifikant med forklart varians fra 7 % til 11 %. Modellen som helhet er signifikant for både GOPB, GVP og GKVH med 11 % og 12 % forklart varians. Når man inspiserer de individuelle prediktorene er det to signifikante betaverdier for GT, på henholdsvis $\beta = 0.29$ og $\beta = 0.32$ i forhold til prediksjon av GOPB og GKVH.

I den andre regresjonsanalysen, tabell 7, er det prediksjon av fagene innenfor Kontroll som er undersøkt. Kriteriet AP er ekskludert fra regresjonsanalysen. Det var ikke grunn til å ta med dette kriteriet da det viste lavere korrelasjoner enn KAP2.

Tabell 7. Hierarkisk multipl regressjonsanalyse for resultat oppnådd i kontroll

Prediktor	KMFR		KACT		KAP1		KAP2	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Steg 1:	.29**		.01		.09		.02	
AE		-.13		-.02		.25		.08
GT		.75***		.23		.08		-.19
Steg 2:	.07		.01		.00		.00	
S.E		-.31		-.16		.05		-.08
Steg 3:	.00		.06		.01		.01	
I.P.		.08		.28		-.11		.08
Steg 4:	.01		.12*		.00		.00	
Intervjuprognose		-.09		-.39*		.01		.03
Total R^2	.37**		.19		.10		.03	
N	60		60		60		59	

Note: Alle koeffisienter er tatt fra siste steg i regresjonsanalysen. KMFR = Kontroll Military Flying Regulations; KACT = Kontroll Air Combat Training; KAP1 = Kontroll Air Policing eksamen 1; KAP2 = Kontroll Air Policing eksamen 2. * $p < .05$, ** $p < .01$, *** $p < .001$.

Analysen i tabell 7 viser at steg 1 med AE og GT forklarte 29 % av variansen for *Flying regulations*. Kun én av de individuelle prediktorene (GT) var signifikante med en betaverdi på $\beta = .75$. Prediktorene forklarte til sammen 37 % av variansen i *Flying regulations*. Til slutt så har Intervjuprognozen en negativ regresjonskoeffisient på $\beta = -0.39$.

I den siste regresjonsanalysen, tabell 8, er det undersøkt om oppnådde resultat gjennom GKV utdanningen kan predikere resultat i kontrollfagene. I steg 1, ble seleksjonstestene satt inn. Kun GKV Hovedkarakter ble tatt med som prediktor for GKV utdanningsresultat i steg 2. Det ble ikke utført regresjonsanalyse for Varsling praktisk faget. Dette faget viste ingen signifikante korrelasjoner mot kontrollutdanningen i korrelasjonsanalysen.

Tabell 8. Hierarkisk multipl regressjonsanalyse for resultat oppnådd i kontroll.

Prediktor	KMFR		KACT		KAP1		KAP2	
	ΔR^2	β	ΔR^2	β	ΔR^2	β	ΔR^2	β
Steg 1: Seleksjonstester	.37**		.19		.10		.04	
AE		-.13		-.04		.24		.20
GT		.75***		.11		.01		-.23
S.E		-.32		-.11		.08		-.08
I.P		.08		.27		-.11		.07
Intervjuprognose		-.09		-.41		-.01		-.10
Steg 2: GKVH	.01	-.08	.13*	.38*	.04	.22	.04	.22
Total R^2	.38*		.32	.	.05		.04	
N	60		60		60		59	

Note: Alle koeffisienter er tatt fra siste steg i regresjonsanalysen. KMFR = Kontroll Military Flying Regulations; KACT = Kontroll Air Combat Training; KAP1 = Kontroll Air Policing eksamen 1; KAP2 = Kontroll Air Policing eksamen 2. * $p < .05$, ** $p < .01$, *** $p < .001$.

Analysen i tabell 8 viser for kriteriet KMFR, at den totale modellen er signifikant med 38 % forklart varians. Seleksjonstestene i steg 1 predikerer kriteriet KMFR signifikant med forklart varians på 37 %. Den individuelle prediktoren GT viser en signifikant betaverdi for KMFR med $\beta = 0.77$. GKVH predikerer kriteriet KACT signifikant med forklart varians på 13 %. Regresjonskoeffisienten for GKVH viser en betaverdi på $\beta = 0.38$.

Oppsummert så viser regresjonsanalysen i tabell 6, at det er først og fremst er AE og GT testene som kan forklare noe av variasjonen i de faglige prestasjonene ved den grunnleggende utdanningen. Tabell 7 viser at AE og GT testene kan forklare noe av variansen for det teoretiske kriteriet *Military Flying Regulations* (KMFR), innenfor kontrollutdanningen. Utover dette viser tabell 7, at seleksjonstestene ikke kan forklare prestasjoner oppnådd i utdanningen for de andre avhengige variablene. Tabell 8, viser at hovedkarakteren i den grunnleggende utdanningen (GKVH) kan forklare noe av variansen oppnådd i det praktiske faget *Air Combat Training* (KACT). Resterende avhengige variabler kan ikke forklares av den grunnleggende utdanningen.

6 Diskusjon

Diskusjonen vil ta utgangspunkt i resultatene som har blitt presentert, og de viktigste funnene. Problemstillingene vil bli diskutert hver for seg. Innenfor hver problemstilling vil korrelasjonsanalysen bli diskutert sammen med regresjonsanalysen. Til slutt vil styrker og svakheter ved studien bli diskutert.

6.1 Drøfting problemstilling 1

“Undersøke seleksjonstestenes prediktive validitet opp mot utdanningsresultat ved GKV.”

Det er først og fremst AE og GT testene som korrelerer signifikant med kriteriene i tabell 1 og 2. Regresjonsanalysen i tabell 6, forsterker ytterligere inntrykket av at steg 1 med både AE og GT, kan forklare mesteparten av den totale variansen for alle kriteriene. Dette er i tråd med flere tidlige studier som også dokumenterer nytten av generell intelligens som prediktor (Olea & Ree, 1994; Ree & Earles, 1992; Schmidt et al., 2016)

Både AE og GT testene skal måle generelt evnenivå (Martinussen & Torjussen, 1998; Skoglund et al., 2015). Ree og Earles (1992), mente at generell intelligens er den beste prediktoren av jobbprestasjoner. Studien de utførte, omhandlet blant annet å kunne predikere «*training success*». Undersøkelsen omfattet 78049 kandidater, innenfor luftfart og 89 forskjellige tekniske jobbkurs. Dataene var hentet ut fra *Armed Services Vocational Aptitude Battery*¹⁷ (ASVAB). De kom frem til at for en tredjedel av jobbene var intelligens (*g*) den eneste signifikante prediktoren. I tillegg var gjennomsnittskorrelasjonen med «*training success*» $r = .76$, etter korreksjon for *Range Restriction* (RR). Ved å legge til andre tester som omhandlet andre mer spesifikke evner, økte den multiple korrelasjonen med $r = .02$. Med andre ord, spesifikke evner forbedret prediksjon i liten grad. Som et direkte motsvar på denne undersøkelsen, kom McClelland (1993) med undersøkelsen *Intelligence is not the best predictor of job performance*. Her mente McClelland at intelligens overhodet ikke var den beste prediktoren. Derimot mente han at ikke-kognitive evner kunne forklare minst like mye, av hvorfor noen gjør en bedre jobb enn andre. Videre mente han også at de ikke hadde tatt innover seg andre variabler som hadde mye å si. Dette kunne være variabler som sex, rase, utdanningsnivå, sosial klasse og andre ikke-kognitive evner. Disse variablene kunne påvirke jobbprestasjoner vel så mye som intelligens. Brown, Le, og Schmidt (2006) sier i sin

¹⁷ Seleksjonsbatteriet til *US Military*

undersøkelse, at det er ingen grunn til å forvente, at skreddersydde tester vil være bedre enn en total vurdering av *General Mental Ability* (GMA). Denne undersøkelsen fremstår med mer tyngde, da denne studien er kritisk til hvordan blant annet Ree og Earles (1992) utførte sin studie. Brown et al. (2006) tok utgangspunkt i 10 forskjellige skoler innenfor den Amerikanske marinen. De sammenlignet deretter GMA med spesifikke evnetester, basert på *Final School Grade* (FSG). Resultatene de kom opp med, var at de spesifikke evnetestene ikke predikerte bedre enn GMA. De mente derfor at dette kastet et positivt lys over undersøkelsen til Ree og Earles (1992), hvor andre type evnetester som ikke målte GMA økte forklart varians med $r = .02$. Resultatene som Brown et al. (2006) kom frem til, viste i tillegg at GMA korrelerte med *Final School Grade* (FSG) fra $r = .34$ til $r = .58$ avhengig av skole de gikk på.

Med tanke på disse analysene, er det således ikke overraskende at AE og GT testene korrelerte signifikant med kriteriene innenfor Grunnleggende Kontroll og Varslingsutdanningen (GKV). Både AE og GT korrelerte med kriteriene fra $r = .21$ til $r = .32$. Dette kan virke som relativt lave korrelasjoner, men på samme tid er ikke disse variablene korrigert for *Range Restriction* (RR). Nevnte undersøkelser er heller ikke spesifikke undersøkelser for Luftfart. Undersøkelsen til Damitz og Eissfeldt (2004) er rettet mot flygeleder (ATC) seleksjon. Denne undersøkelsen viste at mål på generell intelligens (g) alene, ikke kunne predikere utdanningsprestasjoner. Spesifikke evnetester virket å spille en viktig rolle for seleksjon. Sagt på en annen måte, så har generell intelligens noe å si, men ikke alt. Det skal også tillegges at jobbutførelse mellom ATC og *Fighter Controller* (FC) i forhold til Varsling, kan sies å være veldig forskjellig. Det er flere likheter mellom ATC og FC enn ATC og Varsling. Slik sett er kanskje ikke Damitz og Eissfeldt (2004) sin undersøkelse direkte anvendbar for denne problemstillingen. For den grunnleggende utdanningen ved LKVS, virker det generelle evnenivået AE og GT testene måler, å være en god prediktor.

Tabell 1 og 2 i denne undersøkelsen, viser at andre spesifikke evnetester innenfor Spasiale Evner (SE) og Informasjonsprosessering (IP), i liten grad predikerte de valgte kriteriene. I utdanningen for Grunnleggende Kontroll og Varsling (GKV) er det fokus på teoretiske fag og ett praktisk fag innenfor Varsling. Varsling praktisk (GVP) er forskjellige i fra de andre kriteriene, i og med at dette er det eneste praktiske faget som blir gjennomført med bruk av simulator. Denne testen er en kombinasjon av spørsmål fra eksaminator, på samme tid som

kandidaten må gjennomføre oppgaver som dukker opp på simulatoren. Hovedmålene med GKV utdanningen, er å få studentene forberedt til å fungere i et operasjonsrom (ARS), for å bemanne posisjoner innenfor Varsling, eller at de skal ta videre spesialisierende utdanning på LKVS. Hvis det var et fag, hvor det kunne vært forventet at testene for IP og SE kunne fått signifikante korrelasjoner, var det mellom IP/SE og Varsling praktisk (GVP). Det påfølgende spørsmålet blir derfor hvorfor IP og SE ikke korrelerer med Varslingskriteriet. En mulig forklaring, er at de evnene som IP og SE måler, ikke er relevante for det aktuelle kriteriet. De to prediktorene kartlegger henholdsvis spatiale evner og evne til informasjonsprosessering eller *multi-tasking*. Goeters et al. (2004) viser til at kognitive evner er meget sentralt for flygelederyrket. Jobbanalysen viste at utførelsen av flere oppgaver på samme tid med høy presisjon, krever visse type egenskaper som IP og SE måler. Selv om også Varsling praktisk (GVP) er krevende, er det ikke sikkert at disse evnene er like viktige for dette kriteriet. Det er i alle fall det korrelasjonsanalysen viser. Ree og Earles (1992) har sammen med andre undersøkelser (Olea & Ree, 1994; Ree & Carretta, 1996), hevdet at spesifikke evner i liten grad kan predikere utover generell intelligens (g). I denne undersøkelsen for den grunnleggende utdanningen, kan det tyde på at dette stemmer.

Det har ikke vært noen forskning eller studier hva angår Varslingsfaget, så dette er således den første valideringsstudien som blir utført innenfor dette faget. På samme tid kan det også ligge andre forklaringer bak, hvorfor IP og SE ikke korrelerer signifikant med Varsling praktisk (GVP). Det første kan være reliabiliteten til gjennomføringen av eksamen for Varsling praktisk (GVP) er dårlig. Eksamensresultater er hentet fra 2007 og frem til 2016. Med varierende instruktører i perioden, og kanskje mindre endringer i innhold av faget, kan dette påvirke hvor konsistent og presist kriteriet er målt (Ringdal, 2013). Begrepsvaliditeten bør derimot være bra for dette kriteriet, da det bør måle relativt presist det man er interessert i (Martinussen & Hunter, 2008). Det andre kan være mangel på jobbanalyse. Jobbanalyse blir nevnt å være en viktig komponent for å få et seleksjonssystem til å fungere optimalt (Eissfeldt & Heintz, 2002; Martinussen & Hunter, 2008). Som tidligere nevnt i kapittelet for forskning på seleksjon, innehar K&V systemet i praksis jobbkravene for hver enkelt posisjon, men mangler personkravene. En analyse av personkravene ville fått frem hvilke evner en kandidat innenfor Varsling bør ha. Dermed ville det også kunne kommet bedre frem hvilke prediktorer som vil være korrekt å bruke i forbindelse med seleksjonen. På den andre siden er ikke seleksjonen LSES gjennomfører for Varslingsfunksjonene. Seleksjonen LSES gjennomfører,

er for kontrollutdanningen. Hadde det ikke vært for kontrollutdanningen ville ikke det vært gjennomført seleksjon for den grunnleggende utdanningen (GKV) ved LSES. Dersom seleksjon ikke hadde vært gjennomført ved LSES, ville det hovedsakelig kun vært den naturlige seleksjonen gjennom inntak til GBU.

Innenfor gruppedomenet GT, viser flere prediktorer signifikante korrelasjoner med kriteriene. Det som utpeker seg, er prediktoren Engelsk. Engelsk korrelerer signifikant med alle fire kriteriene i tabell 1. Undersøkelsen til Pecena et al. (2013) utført i Tyskland med 430 ATC kandidater under innledende grunnopplæring, viser at Engelsk er et fag som kan ha innvirkning på kandidatens prestasjoner. Av en større undersøkelse, målte de blant annet engelsk kompetanse hos studenten. Resultatet viste signifikant korrelasjon på $r = .20$, mellom engelsk kompetanse og teoretisk karakter for den innledende utdanningen. Den viste også signifikant korrelasjon for de som måtte ta eksamen på nytt ($r = -.09$), og totalresultatet ($r = .08$). Engelsk kompetanse ble også målt opp mot innledende trening på simulator, hvor det ikke ble funnet noen signifikant korrelasjon. Pecena et al. (2013) nevner også i sin drøfting at siden noen kurs blir holdt på engelsk, vil elever med bedre engelskkunnskaper ha mindre problemer med å holde følge med emnet. De med mindre bra engelskkunnskaper, må anstrenge seg ekstra for å forstå språket og holde følge med emnet. På grunn av ekstra engelskundervisning under den innledende fasen av utdanningen, vil språkkunnskap etter hvert bli underordnet og mindre viktig som en faktor for prestasjonsforskjeller. Derfor var den prediktive validiteten sterkest for teoretisk prestasjon i den innledende fasen ($r = .20$). For de som måtte ta eksamen på nytt ($r = -.09$) og totalresultat ($r = .08$), var resultatene svake. De svake validitetskoeffisientene, kan tolkes som en indikasjon på at et minimum av engelskkunnskaper er en forutsetning, for å starte trening og forebygge uheldig utfall i starten av den innledende treningen. I studien til Martinussen og Torjussen (1998) utførte de en validering og meta-analyse av det norske testbatteriet for seleksjon av piloter. Dette innebar gjennomføring av to faktoranalyser for å undersøke sammenhengen mellom testene i batteriet. I tillegg ble det utført en korrelasjonsundersøkelse for å evaluere den prediktive validiteten av testene mot to kriterier på pilot prestasjon. På de to kriteriene for pilot prestasjon viste Engelsk en korrelasjon fra $r = .17$ til $r = .26$.

Sammenlignet med studien til Pecena et al. (2013) og Martinussen og Torjussen (1998) viser prediktoren Engelsk i tabell 1, mer eller mindre høyere signifikante korrelasjoner med alle

kriteriene som er brukt i utdanningen. Både Operative Prosedyrer og Bestemmelser (GOPB) og System Forståelse (GSFT) er teoretiske fag med relativt mye engelsk terminologi. Selv om ikke fagene blir holdt på engelsk, kan det tyde på at Engelsk er viktig for å kunne prestere i faget. I motsetning til undersøkelsen til Pecena et al. (2013), hvor engelskkunnskaper ikke korrelerte signifikant med simulator, var engelsk også signifikant korrelert med det praktiske kriteriet Varsling praktisk (GVP). Om dette er på grunn av at LKVS ikke kjører engelskundervisning er vanskelig å si. Resultatene i tabell 1, viser at prediktoren Engelsk jevnt over har høyere korrelasjoner med kriteriene enn for de andre prediktorene. Engelsk er dermed den beste prediktoren for Grunnleggende Kontroll og Varslingsutdanningen.

Den siste prediktoren som gjenstår er Intervjuprognosen. Den korrelerer signifikant opp mot både GOPB og GVP med svakt til middels sterke korrelasjoner ifølge Cohens kriterier (1988). Undersøkelsen til Pecena et al. (2013) viste at intervju kan fungere bra som prediktor for utdanningsresultat. I den oppdaterte meta-analysen av Schmidt et al. (2016), viste at strukturerte og ustrukturerte intervju fikk den samme validitetskoeffisienten på $r = .58$. Det skal sies at denne studien er både korrigert for RR og manglende reliabilitet, noe som gjør korrelasjonene høyere, enn hva denne studien kan vise til.

Selv om Intervjuprognosen korrelerer bra med både GOPB og GVP, korrelerer den ikke signifikant med de to andre kriteriene. Det kunne kanskje vært forventet at hovedkarakteren til GKV (GKVH) utdanningen ville korrelert signifikant, siden to av de andre fagene korrelerer signifikant med hverandre. Regresjonsanalysen viser også, at Intervjuprognosen bidrar i liten grad til bedre prediksjon. Til tross for dette er det god grunn til å fortsette med Intervju som prediktor. Hadde utvalget vært større for Intervjuprognosen, ville det hatt bedre statistisk styrke for studien og dermed lettere kunne få avdekket sammenhenger mellom denne prediktoren og kriteriet. En annen utfordring for GKV utdannelsen er å formidle hva K&V innebærer. Selv om kandidatene får informasjon om hva det egentlig er de utdanner seg til, kan det være forskjell på det som blir sagt, i forhold til virkeligheten. Intervjuet er dermed meget viktig for å bestemme seg for om kandidaten virkelig ønsker denne fagretningen.

Oppsummert er det AE og GT testene som virker å være gode prediktorer for akademiske prestasjoner innenfor den Grunnleggende Kontroll og Varslingsutdanningen ved LKVS. De viser god prediktiv validitet for enkeltfag, samt for hovedkarakteren (GKVH). Av flere

evnetester innenfor gruppedomenet GT, peker Engelsk seg ut som en god prediktor for alle kriterier. Til slutt viser Intervjuprognozen indikasjoner på at den kan være en god prediktor til den grunnleggende utdanningen, selv om regresjonsanalysen viser den i liten grad gir bedre prediksjon.

6.2 Drøfting problemstilling 2

“Undersøke seleksjonstestenes prediktive validitet opp mot utdanningsresultat på Kontroll?”

Det er først og fremst tre funn som er interessant med denne analysen. Det første er mangelen på signifikante korrelasjoner mellom prediktorene og kriteriene, hvor det blir brukt simulator. Dette er kriteriet *Air Combat Training* (KACT) og *Air Policing kriteriene* (KAP1, KAP2 og AP). Det andre er den signifikante korrelasjonen mellom AE og første eksamen i *Air Policing* (KAP1), men ikke signifikant med andre eksamen (KAP2). I tillegg er det ingen signifikante bivariate korrelasjoner mellom *Air Policing kriteriene* (KAP1 og KAP2). Noe som kanskje burde være forventet, siden det er to eksamener som bygger på hverandre. Det tredje er at GT korrelerer signifikant med det ene teoretiske kriteriet *Military Flying Regulations* (KMFR), men ikke med resterende kriterier. I regresjonsanalysen viser prediktoren GT at den kan forklare mye av variansen for resultat oppnådd for KMFR.

Bakgrunnen for at LSES i utgangspunktet startet med å selektene K&V personell, var med hensikt å kunne selektene kandidater med de riktige evnene, for å kunne prestere som FC. Det er derfor overraskende at så mange prediktorer ikke korrelerer med kriteriene, kanskje spesielt IP og SE testene. Det naturlige spørsmålet blir derfor, hvorfor dette ikke korrelerer? Dette kan det være flere årsaker til. Den første og mest åpenbare forklaringen kan være det lave tallet N , og dermed lav statistisk styrke i studien. Dette varierer fra 59 til 92 over en periode fra 2002 og frem til 2016. Selv om dette er mer eller mindre det totale utvalget i denne perioden, er det såpass lav N i tillegg til en relativt streng seleksjon (RR), at det er vanskelig å få signifikante korrelasjoner (Martinussen & Hunter, 2008). AP kriteriet ble tatt med for også å få med resultatene fra LKSK kullene fra 2002 til 2006 i tillegg til resterende resultat fra 2007 til 2016. Dette kriteriet har i utgangspunktet et utvalg på $N = 92$, og dermed høyere N enn de andre kriteriene. Innenfor prediktorene Spasiale Evner (SE) og Informasjonsprosessering (IP), viser de bivariate korrelasjonene liten eller ingen økning i N sammenlignet med kriteriet for andre eksamen i *Air Policing* (KAP2). Dette betyr at det er få eller ingen tester fra LKSK

kullene fra 2002 til 2006 hvor LSES har funnet tester for IP og SE. Bakgrunnen for dette er utfordringen med å finne data såpass langt bak i tid. Informasjon ble hentet inn fra Excel ark og permer, som gjorde det vanskelig å finne. Det er dermed nødvendigvis ikke en systematisk målefeil, sannsynligvis er den usystematisk.

En annen årsak til hvorfor SE og IP ikke korrelerer signifikant, kan være at prediktorene måler andre typer egenskaper enn hva kriteriene krever av egenskaper. Testbatteriet til LSES er først og fremst laget for seleksjon av piloter, og ikke for seleksjon av K&V personell, og inneholder flere tester enn hva K&V personell blir testet for. Til testing og seleksjon av K&V personell er det som tidligere nevnt GT, IP og SE testene som er blitt plukket ut. Kan det være slik at de testene som blir kjørt for å teste IP og SE ikke kartlegger de evnene som en Kontrollør trenger? Den relativt omfattende meta-studien av Martinussen et al. (2000) viser at Spasiale Evner var en av de to viktigste prediktorene for ATC prestasjoner, hvor den korrelerte med $r = .25$ innenfor utdanning. Dette viser viktigheten av SE innenfor flygelederyrket. Jobbanalyse undersøkelsen av Pecena et al. (2013) viser at IP evner er sentrale, men kanskje ikke avgjørende. Der ble det målt signifikante korrelasjoner med $r = .18$ mot *Attention* og $r = .13$ mot *Processing speed*. I studien til Eissfeldt og Heintz (2002) i Tyskland undersøkte de om det kunne være forskjell mellom krav til evner i de forskjellige arbeidsposisjonene for ATC yrket. Selv om de fikk bekreftet at krav til evner kunne variere for arbeidsposisjonene, kom de også frem at evnene innenfor det kognitive domenet var viktig for alle posisjoner, herunder spesielt IP og SE. Det er dermed grunn til å tro at IP og SE er relevant for å kunne utføre en god jobb også innenfor jobben som Kontrollør (FC). Selv om testene for IP og SE i testbatteriet til LSES er pilot orientert, tester de fortsatt IP og SE. Det er dermed ikke belegg for å si, at IP og SE ikke kan predikere for fremtidige Kontrollører.

Det kan også være slik at IP og SE ikke korrelerer signifikant, på grunn av mengdetreningen som blir lagt ned i forbindelse med forberedelser til eksamen. Elevene har muligheten til å trene på egenhånd i lag med andre medelever på simulator. For spesielt siste eksamen for *Air Policing* og eksamen for *Air Combat Training* kan dette ha en sammenheng med at SE og IP ikke korrelerer signifikant med kriteriene. Den kognitive belastningen har blitt mindre på grunn av mengdetrening (Camp, Paas, Rikers, & van Merriënboer, 2001; Kallus et al., 1997).

Ackerman, Kanfer, og Goff (1995) utførte en undersøkelse, for blant annet å se på integrering på flere perspektiver av faktorer, som kan bestemme individuelle forskjeller i kunnskapsbygging. Dette innbefattet blant annet evner, personlighet og yrkesinteresser. Kandidatene ($N = 93$) ble rekruttert ved Universitetet i Minnesota. Testen foregikk på en simulator som kalles *Terminal Radar Approach Controller*. På det første forsøket studentene fikk prøve seg i denne simulatoren, klarte studentene i gjennomsnitt å håndtere $M = 5.89$ fly av totalt 28 fly. I alt 12 % av studentene klarte ikke å håndtere noen fly. Etter 30 forsøk, viste studentene klar forbedring ($M = 19.49$) og 2 % av studentene klarte å håndtere alle flyene. Denne testen ble videre korrelert med blant annet Spasiale Evner ($r = .56$) og Persepsjonshastighet ($r = .63$). Det var dermed en klar sammenheng mellom prediktorer og prestasjon. På samme tid viser også denne testen at når studentene hadde forsøkt 30 ganger, så var gjennomsnittet på å håndtere fly for hver student over 19. En grunn til den høye korrelasjonen var at det ikke var noen RR. Studien viste også at gjennom tillært kunnskap får studenten progresjon. Avhengig av hva som er minstekravet til testen, så er kanskje håndtering av 19 fly bra nok. I overført betydning, så kan dette bety at avhengig av minstekrav og mengdetrening, så kan de aller fleste klare å gjennomføre utdanningsmålet. Alle som er selektert til Kontrollutdanningen har et relativt høyt evnenivå innenfor SE og IP ($M = 4.7$ og 5.3). Kanskje krever ikke utdanningen evner over dette nivået. Kandidatene som skårer over minstenivået har dermed gode evner nok innenfor SE og IP, og dermed blir det ikke av betydning, hvor høyt over dette nivået de skårer. Dette trenger ikke å være slik når de skal utøve yrket sitt på ARS, men innenfor simulator kan det være slik. I alle fall, hvis eksamensgjennomføringen i stor grad er forhåndsplanlagt med få overraskelser underveis for kandidaten. På samme tid, dersom denne argumentasjonen er gyldig, så kan dette også bety det samme for de andre prediktorene. Det er dermed sannsynligvis liten grunn til å tro at dette stemmer. En annen forklaring kan være at hvis alle får muligheten til å trene så mye de vil, med det resultatet at alle elever stort sett presterer bra, blir det lite variasjon igjen til kriteriet. Dette kan være med på å forklare hvorfor ikke prediktorene SE og IP ikke korrelerer med oppnådde prestasjoner i Kontrollutdanningen. Studien til Ackerman et al. (1995) viser også at mengdetrening hjelper for å oppnå en prestasjon.

Utover manglende sammenheng mellom SE og IP og kriteriene, ble det innledningsvis nevnt det er signifikant korrelasjon mellom AE og første eksamen i *Air Policing* (KAP1), men ikke andre eksamen. Et mulig svar, kan være kan være reliabilitet og validitet til disse to kriteriene.

For å se etter forklaringer, er det behov for å beskrive *Air Policing* kriteriet noe mer detaljert. Ved avskjæring av et uidentifisert objekt, eller fly, er målet å lede to jagerfly bak det uidentifiserte flyet. Det er spesifikke krav til hvor nært jagerflyene skal ledes til det uidentifiserte flyet. Både i distanse bak flyet, men også sideveis. Dette krever presise kalkuleringer fra FC for å gi de riktige kommandoene til piloten, både når det gjelder hastighet, høyde og retning. Underveis er også beregninger avgjørende i forhold til når Kontrollør skal gi de endelige instruksjoner til piloten, for å komme bak det uidentifiserte flyet. Dersom Kontrollør er et sekund for sent ute med sin instruksjon, kan hele avskjæringen være ødelagt. Under eksamensgjennomføring må eleven gjennomføre flere slike avskjæring i simulator. I simulator vil den som spiller pilot, kun følge retningslinjene til FC. Dette betyr at den minste feilkalkulering, kan føre til store konsekvenser for utførelsen av avskjæringen. Under eksaminasjon er det alltid to sensorer. Som regel er det en sensor fra LKVS og en sensor fra ARS, avhengig av tilgjengelighet. Den ene sensoren sitter bak eleven og tar fortløpende vurderinger. Den andre sensoren sitter i en annen posisjon for å ta seg av kontrollmålinger underveis. I tillegg er det en tredje sensor, som skal være en «sparringspartner» for de to andre sensorene etter endt eksaminasjon.

Ut i fra dette, skulle det være grunn til å tro at både reliabilitet og validitet er ivaretatt. Det er liten tvil om at begrepsvaliditeten skal være høy, da kriteriene gjenspeiler i høy grad hva kandidat faktisk skal utføre når de er ferdig med utdanningen. For kriteriereliasiteten kan det være flere forhold. Siden data for undersøkelsen er sekundærdata (Jacobsen, 2015, s. 140), kan det oppstått målefeil. Dersom det har oppstått målefeil, kan det ha foregått ved to anledninger. Ved nedtegning av og analyse av data kan dette oppstå. Dette er liten grunn til å tro, da dataene er hentet ut fra vitnemålene. Den andre anledningen det kan oppstå, er ved vurdering av totalprestasjon av kandidat under eksamensgjennomføring. Dette er svært liten grunn til å tro, siden gjennomføringen foregår svært standardisert, med bruk av skjema, i tillegg til minimum to sensorer.

Siden AE korrelerer signifikant med første eksamen (KAP1) og ikke andre eksamen i *Air Policing* (KAP2), kan det være andre årsaker enn det kriteriet måler. Regresjonsanalysen i tabell 7, viser i steg 1 at AE og GT kan forklare 29 % av variansen for faget *Military Flying Regulations* (KMFR). I tillegg viser tabell 4 høy korrelasjon mellom GT og KMFR i følge Cohens kriterier (1988). Selv om KMFR måler et teoretisk fag sier dette noe om at

kandidatene har prestert bra på det teoretiske men ikke det praktiske. En studie fra USA undersøkte om det amerikanske luftfartstilsynet sitt seleksjonssystem¹⁸(AT-SAT) kunne øke prediksjon til det amerikanske militærvesen sitt seleksjonssystem¹⁹(ASVAB) (Carretta & King, 2008). Totalt var det 448 kandidater som hadde vervet seg til ATC utdanning gjennom det amerikanske Forsvaret. Disse kandidatene gjennomførte begge seleksjonssystemene. ASVAB er et kognitivt testbatteri brukt til alle som verver seg, for å få en oversikt over kvalifikasjoner og evner. AT-SAT er et testbatteri som er spisset mot ATC yrket basert på jobbanalyser. Av flere kriterier de brukte, var endelig utdanningsresultat for teoretisk fag og praktisk fag to av dem. Resultatene viste at ASVAB hadde en god prediktiv validitet for spesielt teoretisk akademiske fag. For det praktiske faget viste en arbeidsprøve test fra AT-SAT batteriet seg som en av de mest lovende. Denne testen alene kunne predikere bedre enn ASVAB batteriet for praktisk utførelse. Denne undersøkelsen viser altså at ASVAB batteriet er ett bra testbatteri, spesielt for teoretiske fag, men kanskje ikke i samme grad for praktiske fag, og det var spesielt arbeidsprøvetesten som kunne være med på å øke den prediktive validiteten til ASVAB batteriet. I forhold til denne studien, kan dette være sammenlignbart. Det teoretiske kriteriet (KMFR) viser sterk sammenheng med prediktoren GT, men ikke med de andre kriteriene. Om det er også slik for seleksjonsprosessen for Kontrollutdanningen er vanskelig å si. Ut i fra funnene i analysen, og til tross for lav N , kan det være indikasjoner for at seleksjonen predikerer bedre for akademiske fag. Det kan være at en arbeidsprøvetest kunne ha gjort seleksjonen enda bedre.

Utover dette, kan det være andre forhold som kan påvirke elevenes prestasjoner. I praktisk utførelse med mye eksternt press fra opp til tre sensorer, i tillegg til selvpålagt press, så kan det gå utover prestasjonen til kandidaten. Det skal sies at alle kandidater har minimum vært gjennom opptak til befalsskole. Noen av kandidatene har også vært gjennom og gjennomført utdanning gjennom LKSK. Således burde kandidatene være forberedt på å kunne håndtere situasjoner under press. I tillegg så har de vært gjennom seleksjon ved LSES. En av hovedargumentene med seleksjon gjennom testbatteriet til LSES, er å finne de riktige kandidatene som kan håndtere denne type press.

¹⁸ *Federal Aviation Administration Air Traffic Selection and Training battery (FAA AT-SAT)*

¹⁹ *Armed Services Vocational Aptitude Battery (ASVAB)*

I tillegg til at AE kun korrelerer med første eksamen i *Air Policing*, korrelerer ikke første og andre eksamen med hverandre. Som tidligere nevnt, krever *Air Policing* faget presise kalkuleringer med nøyaktige kommandoer til pilot, spesielt i den siste fasen, for å komme bak det uidentifiserte flyet. Det kan med andre ord være en del tilfeldigheter og dagsform som spiller inn, når eksamen blir gjennomført. Dette kan forklare at første og andre eksamen ikke korrelerer signifikant. I tillegg er det såpass få sammenhenger mellom prediktorer og kriterier, at det skaper grunn til å tro på dette har noe å si. Dette blir mye antagelser, og for å få enda bedre belegg på denne antagelsen, må enn se på hvordan Kontrollører faktisk gjør det i praktisk virke på ARS. I det praktiske virke vil det komme flere faktorer inn. Dette kan være vær, vind, dårlig radioforbindelse med pilot og oppdukkende hendelser som krever mye mer av de kognitive evnene. Siden kriteriene er hentet ut under utdanning, kan det være at disse skiller seg en del fra faktiske jobbprestasjoner. Ved å ha sammenlignet prediktorene opp mot faktiske jobbsituasjoner ved ARS ville det sannsynligvis gitt et bedre bilde av hvordan det virkelig er. Kanskje resultatet ville blitt annerledes.

Ingen prediktorer korrelerer signifikant med ACT Sim (KACT). Det er vanskelig å forklare hvorfor det ikke er noen sammenheng med dette. *Air Combat Training* (ACT) er det som blir gjennomført oftest av en Kontrollør, og betegnet som det mest krevende en Kontrollør utfører. Det er derfor overraskende at dette ikke viser noen sammenheng. Det vil også her være av interesse, å se på hvordan kandidatene sine jobbprestasjoner hadde vært på ARS. I en del scenarioer i det praktiske virke, kan utførelsen av ACT bli meget komplisert. Håndtering av mye informasjon både fra interne og eksterne aktører, kan kreve stor bruk av kognitive evner for å håndtere situasjonen. Ved å se på hvordan Kontrollører presterer, når de har kommet ut på ARS, ville muligens ha gitt noe mer svar. Igjen så kan dette bunne ut i at kriteriet kanskje ikke måler reelt hva som kreves av de kognitive evnene. Dette vil kanskje ha kunnet kommet bedre frem ved å ha brukt kriterier fra ARS.

Korrelasjonsanalysen i tabell 3, viser signifikant korrelasjon mellom Intervjuprognozen og KACT. Regresjonsanalysen i tabell 7 viser at Intervjuprognozen kan forklare 12 % av variansen for KACT. Det som er overraskende er at regresjonskoeffisienten viser en negativ betaverdi på $\beta = -.39$. Intervjuprognose tar hensyn til mange forhold, herunder motivasjon. En forklaring kan være at kandidatene starter først ved den grunnleggende utdanningen før videreutdanning innenfor Kontroll. Det kan kanskje være slik at motivasjon forandrer seg i

løpet av den grunnleggende utdanningen. Noen blir mer motivert, andre blir mindre motivert. Noen kandidater finner ut at dette ikke er noe de vil satse på. Siden det er såpass lav N , i kombinasjon med svake funn, kan det også forklare det uventede funnet. Dette må undersøkes nærmere i en studie med større N og dermed blir ikke denne observasjonen diskutert videre.

Kanskje kan det være hensiktsmessig med en jobbanalyse for jagerflykontrolløffiserer. Selv om det er gjennomført mange jobbanalyser innenfor ATC yrket, er det også en del forskjeller fra ATC yrket, og å være FC. Det kommer også frem i tidligere nevnte jobbanalyse studie til Eissfeldt og Heintz (2002), hvor de undersøkte om det kunne stille forskjellige krav til evner i forskjellige arbeidsposisjoner for ATC yrket. De sjekket tre forskjellige arbeidsposisjoner og bekreftet det var forskjeller i evnekrav til de forskjellige arbeidsposisjonene. Selv om dette handlet hovedsakelig handlet om forskjeller innenfor sensoriske evner, viser dette at hvis det er forskjeller på arbeidsposisjoner innenfor et yrke, vil det i alle fall være forskjell mellom ATC og FC. En god jobbanalyse vil få kartlagt hva en FC virkelig trenger av evner for å utføre en god jobb. Dette kan også gjøre seleksjonssystemet enda bedre, og dermed større sannsynlighet for å få riktige kandidater inn til utdanningen. Kanskje dette i større grad også kan hjelpe til med turnover problematikken som K&V bransjen sliter med.

6.3 Drøfting problemstilling 3

«I hvilken grad kan resultat fra Varsling og GKV utdanning predikere utdanningsresultat på Kontroll?»

Korrelasjonsanalysen i tabell 5, har tatt utgangspunkt i antagelsen om at kanskje første del av utdanningen kan predikere for resultat i den videregående Kontrollutdanningen. I denne analysen ble Total Varsling Praktisk (TVP) tatt med som prediktor for å inkludere resultat fra LKSK kullene. Det eneste kriteriet denne prediktoren kunne fått et annet resultat i forhold til kriteriet Grunnleggende Varsling Praktisk (GVP), er prediktoren *Air Policing* (AP). Prediktoren AP inneholder også resultat fra LKSK kullene. Til tross for et høyere utvalg, viser den ingen signifikans.

Hovedkarakteren til GKV (GKVH) utdanningen viser signifikant korrelasjon med to av de praktiske kriteriene. Henholdsvis *Air Combat Training* (KACT) og første eksamen i *Air Policing* (KAP1). Ved å se på regresjonsanalysen i tabell 8, viser den at GKVH kan forklare 13 % av variansen for kriteriet KACT. I tillegg viser regresjonskoeffisienten en betaverdi på β

= .38. Selv om N er lav, så viser dette en sammenheng med den totale karakteren ved GKV utdanningen og resultater oppnådd i *Air Combat Training* (KACT). I tillegg så korrelerer hovedkarakter i den grunnleggende (GKVH) utdanningen signifikant med KAP1, den første av to eksamener i *Air Policing* faget.

I 1996 ble det gjennomført en meta-analyse om å undersøke validitets evidens brukt i seleksjon av piloter (Martinussen, 1996). Formålet var å oppdage moderatorer for sammenhengen mellom prediktorer og pilotprestasjoner. Meta-analysen bestod av 50 innsamlede studier fra 11 forskjellige nasjoner. Av flere resultat i analysen, kom det blant annet frem at den beste prediktoren for pilot prestasjon var tidligere trenings-erfaring. De fleste av de innsamlede studiene ble utført under grunnleggende trening, og derav noen med erfaring fra flyging før flyskolen kunne ha vært til hjelp for kandidatene. På samme tid var det usikkert hvordan dette forholdet ville være konsekvent over tid, og fortsatt være gyldig når utdanningsperioden er over (Martinussen, 1996). Med andre ord så kan tidligere erfaring være en fordel i starten av en utdanningsperiode, men etter en tid ut i utdanningen, vil sannsynligvis de andre elevene ta igjen fordelene som de elevene med mer erfaring har i starten. Selv om dette var en undersøkelse i følge med pilot-utdanning, kan noe av dette være overførbart til denne undersøkelsen. Siden GKVH korrelerer signifikant med første eksamen i *Air Policing* (KAP1) og ikke med andre eksamen (KAP2), kan kanskje dette argumentet brukes i dette tilfellet. AP eksamen 1(KAP1) blir utført tidligere i utdanningsperioden, og siste AP eksamen (KAP2) blir utført senere i utdanningen. Dermed kan elever som har hatt gode prestasjoner i GKV utdanningen stille med en fordel, sammenlignet med de som har prestert mindre bra.

Denne argumentasjonen blir vanskeligere å bruke for KACT kriteriet da det bare gjennomføres en eksamen i nærmere slutten av utdanningsperioden. På samme tid viser hovedkarakteren i GKV, som nevnt innledningsvis en sammenheng med *Air Combat Training* (KACT). For å kunne se etter forklaringer for dette, er det behov for å se på forskjellen mellom *Air Policing* (AP) og *Air Combat Training* (KACT). Som tidligere nevnt er AP ett fag hvilket krever stor grad av nøyaktighet og beregning for å kunne prestere bra på eksamen. For ACT er det også viktig med beregning i form av kommunikasjon men på en annen måte. Siden ACT innebærer å gi oppdateringer til pilot om hvor fiendtlige fly holder seg til enhver tid, krever dette en helt annen forståelse av hvordan luftkampen foregår. Informasjon må gis

til pilot på riktig måte, når piloten behøver det. Det er dermed i mindre grad viktig å gi piloten riktig ordre om hvilken retning flyet skal ta. Andre kriterier ligger til grunn for vurderingen av prestasjonen FC gjør, til forskjell fra AP. Dette gjelder også under eksamen. Dette kan kanskje forklare hvorfor GKVH korrelerer med KACT og GKVH ikke korrelerer med KAP2. Det spiller mer tilfældigheter inn under eksaminasjon for første og andre eksamen i *Air Policing* (KAP1/2), enn for KACT kriteriet.

Kan det være slik at den grunnleggende utdanningen kan fungere som en slags *work sample* test for *Air Combat Training*? Flere studier innenfor ATC kommer frem med at arbeidsprøve tester har god prediktiv validitet for ATC yrket (Eissfeldt, 1995; Martinussen et al., 2000). På samme tid blir det vanskelig å konkludere med at GKV utdanningen kan predikere for Kontrollutdanningen. Det mest naturlige kunne kanskje være Varsling Praktisk (GVP) kriteriet, siden dette er det kriteriet hvor elevene blir målt på prestasjon i simulator. Siden dette ikke har noen signifikante observasjoner er det ikke belegg for å hevde dette. På samme tid er det ikke til å unngå at det er noen indikasjoner som tyder på sammenhengen mellom den grunnleggende utdanningen og resultat oppnådd for *Air Combat Training*.

Korrelasjonsanalysen og regresjonsanalysen viser en sammenheng mellom de teoretiske fagene og *Air Combat training* (KACT). Dette viser korrelasjonsanalysen gjennom både Operative Prosedyrer og bestemmelser (GOPB) og Hovedkarakter GKV (GKVH). I tillegg er det korrelasjon mellom det teoretiske faget innenfor Kontrollutdanningen *Military Flying Regulations* (KMFR) og KACT. Det blir vanskelig å kalle den grunnleggende utdanningen innenfor Kontroll og Varsling en arbeidsprøvetest. Til det er sannsynligvis forskjellen mellom utdanningen for stor. Det kan heller være slik at gode resultat innenfor den grunnleggende utdanningen, viser god motivasjon og arbeidsinnsats. Kanskje motivasjon og arbeidsinnsats gir større utslag for oppnådd prestasjon innenfor *Air Combat Training*. De som velger videreutdanning underveis i den grunnleggende utdanningen, må nødvendigvis ha en større motivasjon. I så fall, kan det være slik at motivasjon og arbeidsinnsats, har noe å si for prestasjoner oppnådd i *Air Combat Training*.

Undersøkelsen til Pecena et al. (2013) viste prediktiv validitet for simulator (-.16). Selv om dette var i forbindelse med en undersøkelse hvor de undersøkte den prediktive validiteten for Intervju for ATC, viser dette at motivasjon har noe å si. De som søker seg direkte inn på ATC

utdanning, vet at det er akkurat dette de skal utdanne seg til. Dette gjelder ikke nødvendigvis i samme grad for de som velger utdanning innenfor K&V. De starter først på den grunnleggende utdanningen, før de eventuelt søker seg på videreutdanning. Dette kan bety at Intervjuprognosen stemmer bedre for den grunnleggende utdanningen, enn hva den gjør for Kontrollutdanningen. Det er først i løpet av den grunnleggende utdanningen kandidater bestemmer seg for om de skal ta videreutdanning innen Kontroll eller ikke. De som søker seg videre til denne utdanningen, har kanskje underveis fått mer motivasjon underveis. De mest motiverte, og med interesse for å bli FC, velger dermed denne utdanningen. Dette kan dermed være med på å forklare hvorfor de teoretiske fagene er med på å predikere noe av utdanningsresultatet til *Air Combat Training*.

Til slutt så viser både kriteriet GSFT og GOPB at de korrelerer signifikant med KMFR. Dette viser at det kan være en sammenheng mellom teoretiske prestasjoner i GKV og Kontrollutdanningen. Det ble ikke utført noen regresjonsanalyse med GSFT og GOPB som uavhengig variabel (prediktor). Det kan likevel ikke sies å være overraskende at dette stemmer overens, da både prediktor og kriterier er teoretiske fag.

6.4 Svakheter og styrker ved studien

Den utvilsomt største svakheten med denne studien var det lave utvalget. Spesielt for problemstilling 2 og 3. I tillegg har utvalget ved LKVS vært gjennom minimum en fase til med seleksjon i forkant. De aller fleste har vært gjennom seleksjon for inntak til LBSK før de gjennomfører seleksjon ved LSES. Denne avgrensningen er dermed en *Range Restriction* (RR) som undersøkelsen ble preget av. Siden utvalget er selektert og det var det selekterte utvalget det kunne hentes ut kriterieresultat av, ble størrelsen på korrelasjonen påvirket (Martinussen & Hunter, 2008, s. 119-120). Korrelasjonene i undersøkelsen er derfor lavere enn hva de ville vært uten RR. Dette i kombinasjon med lav N , har gjort det veldig vanskelig å få frem hvor sterk den prediktive validiteten for prediktorene ville vært.

En annen svakhet var utvalget mellom 2002 og 2006. Ved å sammenlikne N i de bivariate korrelasjonene mellom kriteriet *Air Policing 2* (KAP2) og AP, var det nesten ingen økning i N for Spasiale Evner (SE) og Informasjonsprosessering (IP). Med andre ord så har ikke LSES kunnet finne informasjon for SE og IP testene for LKSK kullene mellom 2002 og 2006. For AE, GT og Intervjuprognosen var det en økning i N . Til tross for økning i N , så gjorde det lite

utslag for korrelasjonene, da ingen av dem ble signifikante. Sannsynligvis vil dette også si, at dersom data hadde blitt samlet inn fra år 2000, ville det neppe hjelpet i nevneverdig grad.

En annen mulighet for å kunne fått større N , hadde vært å henvende seg til andre land. Kanskje Danmark eller Nederland kunne vært en mulighet. Disse landene opererer med samme systemer til utførelse innenfor Kontroll og Varsling. Imidlertid ville det blitt en helt annen utfordring både med tanke på valg av prediktorer og kriterier. Både seleksjonssystemet og utdanningssystemet er annerledes. En meta-analyse kunne kanskje vært mulig. Utfordringen er at det er svært få studier i andre land, hvilket er så spesifikt rettet som denne studien.

Det har også vært en del andre utfordringer rundt denne lokale valideringsstudien som oppgaven bærer preg av. Den sannsynligvis største utfordringen var valget av kriterier for Kontrollutdanningen. Siden problemstillingen tok utgangspunkt i Kontrollutdanningen utført ved LKVS, kan det ikke sies at valget av kriterier var feil. Til tross for dette, kan det virke som reliabiliteten og validiteten for disse kriteriene kan diskuteres. Det første kan være eksamensgjennomføring. Selv om det er opp til tre sensorer som vurderer kandidat, kan det fortsatt være rom for at det går utover reliabiliteten. En god start på eksamensgjennomføring kan påvirke sensorene til å tro at vedkommende gjennomfører de andre oppgavene gjennom eksamen på like god måte (Martinussen & Hunter, 2008, s. 116). Denne såkalte smitteeffekten kan sannsynligvis ikke undervurderes. Andre aspekt kan også spille inn i prestasjon til kandidat under eksaminasjon, spesielt med tanke på selvpålagt press, og kanskje ytre press med så mange sensorer rundt seg. Mestring av stress og håndtering av eksaminasjon av kandidat er neppe irrelevant, og vanskelig å måle. Ett annet poeng er at kriteriene tar utgangspunkt i simulator som kan gi en del begrensninger. Spesielt i forhold til hvordan praktisk utførelse er i en reell situasjon på ARS. Selv om teknologien til stadighet blir bedre, vil det alltid være forskjell på trening og reel jobbutførelse. Kriterier fra ARS ville dermed sannsynligvis vært bedre. På samme tid ville da problemstillingen vært en helt annen. Valget av kriterier brukt i forbindelse med utdannelsen, kunne neppe vært noe annet.

Undersøkelsen har for det meste benyttet seg av teori knyttet til ATC. Selv om det er klare sammenhenger mellom ATC og FC er det fortsatt forskjeller mellom yrkene. Det kan diskuteres om tidligere forskning anvendt i oppgaven er direkte overførbart til denne

undersøkelsen. Dette var kanskje en enda større utfordring for problemstilling 1, da denne utdanningen ikke er direkte overførbart til FC utdanningen. Sammenligningen av tidligere forskning kan ha ført til en noe uklar diskusjon.

Selv om det er klare svakheter er det også noen styrker ved studien. For kriterier brukt ved GKV utdanningen var N høy nok til å avdekke prediktiv validitet for noen av prediktorene. Til tross for lav N for kriteriene til Kontrollutdanningen, har alle kandidater vært gjennom akkurat de samme seleksjonstestene. Det ble heller ikke brukt noe todelt kriterium i denne undersøkelsen, dette har vært med på å øke relevansen. Prediktorene i denne undersøkelsen er tatt fra testbatteriet til LSES. Dette testbatteriet er veldokumentert og gjennomtestet i flere år (Martinussen & Torjussen, 2004; Martinussen et al., 2004). Dette skal derfor ikke være noe tvil om at prediktorene måler det som de er tiltenkt for. Det samme gjelder for AE (Sundet et al., 2004). Selv om det kan være noen svakheter med kriteriene, må det tilføyes at fagkontoret til K&V i Luft Operativt Inspektorat, rutinemessig utfører inspeksjoner ved LKVS, for å se gjennom blant annet utdanningen. Dette bidraget til en standardisering burde bidra med høyere reliabilitet til kriteriene.

7 Konklusjon

Formålet med denne undersøkelsen var å se på om seleksjonstestene gjennomført ved Luftforsvarets seleksjonssenter (LSES), har en prediktiv validitet for oppnådde prestasjoner ved Luftforsvarets Kontroll og Varslingsskole (LKVS). Dette ble gjort ved å utforme problemstillinger som omhandlet utdanningsprestasjoner ved den grunnleggende utdanningen, og Kontrollutdanningen. I tillegg ble det utformet en problemstilling, for å se på om den grunnleggende utdanningen kan ha en prediktiv validitet opp mot utdanningsresultat på Kontroll. En gjennomgående utfordring for undersøkelsen var det lave utvalget. Spesielt for å undersøke seleksjonstestenes prediktive validitet opp mot utdanningsresultat innenfor Kontroll.

Både korrelasjonsanalyser og multippel hierarkisk regresjonsanalyse ble tatt i bruk for å få frem sammenhengen og den prediktive validiteten mellom seleksjonstestene og utdanningsprestasjoner ved LKVS. Seleksjonstestene og AE fra sesjon ble brukt som prediktorer. Til kriterier ble utdanningsresultat fra både teoretiske og praktiske fag valgt ut.

Denne undersøkelsen viser at det er en sammenheng mellom generell intelligens og prestasjoner oppnådd under GKV utdanningen. Av de evnespesifikke testene, pekte spesielt Engelsk seg ut med signifikante korrelasjoner. Dette tyder på at Engelsk kan være en viktig test for de som blir tatt inn til GKV utdanningen. Prediktoren Intervjuprognose viste signifikante korreleringer med to av fire kriterier. Intervjuprognosen er dermed en prediktor som kan være viktig å bruke i forbindelse med seleksjon inn til GKV. De spesifikke evnetestene innenfor SE og IP viste ingen signifikante korrelasjoner med utdanningsresultatene. Selv om seleksjonstestene ikke er primært for de som kun skal gjennomføre GKV utdanningen, viser denne undersøkelsen at GT testene i seleksjonsbatteriet er viktig for denne seleksjonen, i tillegg til Intervjuprognosen. For LSES er dermed den prediktive validiteten for GKV utdanningen god for GT testene og Intervjuprognosen. Dersom det i fremtiden skal være seleksjon kun til GKV og Varslingsutdanning, men ikke til Kontrollutdanning, er det kanskje ikke behov for å gjennomføre SE og IP testene.

Bakgrunnen for hvorfor LSES gjennomfører seleksjon til LKVS, er Kontrollutdanningen. Utdanningen kan relateres til ATC yrket. Undersøkelser viser at tester med vekt på SE og IP kan være formålstjenlig ved seleksjon til denne typen utdanninger. For Kontrollutdanningen

viste undersøkelsen få signifikante korreleringer. Til tross for å ha samlet inn data tilbake til 2002, er utvalget for lite innenfor Kontrollutdanningen til å nevne store funn. Av de få signifikante korrelasjonene, viste AE signifikant korrelasjon med første eksamen i *Air Policing* (KAP1), men ikke for andre eksamen. I tillegg var det ingen signifikant korrelasjon mellom første og andre eksamen i dette faget. Dette kan tyde på at det kanskje er andre faktorer som spiller inn, som kriteriet ikke fanger opp. Gruppedomenet GT korrelerte signifikant med det teoretiske kriteriet *Military Flying Regulations* (KMFR). Den prediktive validiteten mellom GT og KMFR blir også bekreftet gjennom regresjonsanalysen. Til tross for lav N , kan det virke som GT testene har noe å si for de teoretiske prestasjonene. De andre seleksjonstestene viste mer eller mindre ingen prediktiv validitet. Dette viser at det kan være hensiktsmessig med en jobbanalyse, for å få kartlagt hvilke evner en FC trenger. Undersøkelser viser også at arbeidsprøvetester kan være en god prediktor for en slik type seleksjon. Til tross for seleksjonstestenes manglende prediktive validitet opp mot utdanningsresultat for Kontroll, er det ingen grunn til ikke å fortsette med seleksjon for jagerflykontrollører. Til dette er det for mye forskning på at det faktisk er en sammenheng mellom seleksjonstester og senere prestasjoner.

I noen sammenhenger kan det være slik at tidligere prestasjoner kan predikere senere prestasjoner. Det var derfor av interesse av å se om tidligere prestasjoner oppnådd gjennom Varsling og GKV utdanningen kunne predikere prestasjoner ved Kontrollutdanningen. Hovedkarakteren i den grunnleggende utdannelsen (GKVH) viser prediktiv validitet for det praktiske faget *Air Combat Training*. GKVH viser noen grad av sammenheng med første eksamen i *Air Policing*, men ikke andre eksamen. Det er også en sammenheng mellom teoretiske fag fra den grunnleggende utdanningen (GKV) og Kontrollutdanningen. Varsling praktisk (GVP og TVP) kan ikke predikere for oppnådde prestasjoner innenfor Kontrollutdanning. Ut i fra denne undersøkelsen kan det sies at tidligere prestasjoner ved den grunnleggende utdanningen (GKV), kan predikere noe innenfor teoretiske fag, og for det ene praktiske faget, *Air Combat Training* (KACT).

7.1 Veien videre

Studien har vist det er behov for videre forskning innenfor dette spesifikke fagfeltet. Spørsmålet er, hva kan gjøres for å få seleksjonsprosessen til K&V enda bedre? Underveis i oppgaven har det dukket opp forskningslitteratur, som viser at det er flere områder innenfor

seleksjon til K&V hvor det er potensiale for å forbedre seleksjonssystemet. Kanskje spesielt innenfor *work sample tests* og jobbanalyse.

Eissfeldt (2002, s. 86) sier det så enkelt «*if it has to be judged whether an applicant can become a successful ATCO, just let him/her do air traffic control*». Det sies videre om arbeidsprøvetester at dersom søkeren ikke har noe tidligere kunnskap om oppgaven, burde oppgaven være så enkel så mulig, så det er mulig å lære seg. På samme tid komplisert nok til at det blir informativt med tanke på vurderingen av den individuelle kandidaten. Dersom kandidaten viser god utførelse vil også vedkommende med god sannsynlighet kunne utføre utdannelsen. Både for flygere og flygeledere har det blitt utviklet gode *Work sample tests*, eller arbeidsprøvetest (Martinussen, 2005). I en valideringsundersøkelse av *First European Air Traffic Controller Selection Tests* (FEAST), kom det frem at *Work sample tests* som vurderte *multi-tasking* i et kompleks scenario var viktig. Resultatene fra testen ble mer relevant, til lenger ut i utdanningsprogrammet enn kom (Damitz, Chetcuti, & Henriques, 2010). Evnene det var behov for i en slik type test var mindre relevant i de tidlige fasene. Dette fordi at lenger ut i utdanningen ville det komme sammenlignbare situasjoner som kom frem i *work sample* testen. Dette var basert på at det i de tidlige fasene var det lave og ikke signifikante korrelasjoner. I de senere fasene viste det en signifikant korrelasjon på $r = .55$, mellom arbeidsprøve tester og utdanningsresultat. Dette var da basert på $N = 163$. Det kan virke fornuftig å ha dette i et testbatteri til seleksjon. Også Schmidt og Hunter (1998) sin meta-analyse viser til at den beste prediktoren er *work samples* og dermed den beste prediktoren i lag med generell intelligens og intervju.

Utover arbeidsprøvetester, virker det å være et behov for en jobbanalyse. Både for Varsling og Kontroll. I forskningslitteraturen blir det nevnt at første steget for å gjennomføre en god seleksjon er jobbanalyse (Goeters et al., 2004; Martinussen & Hunter, 2008; Suresh et al., 2012). Den første delen av jobbanalysen for K&V funksjoner er allerede dekket. I forbindelse med seleksjon er det den andre delen av jobbanalysen som er viktig. Ved å gjennomføre en personspeifikasjon innenfor hver funksjon, vil det være mulig å kartlegge hvilke egenskaper en person bør ha for å kunne utføre jobben. Er ønsket å plukke ut de beste kandidatene, er dette et arbeid som bør gjøres.

7.2 Mulig videre forskning

Gjennom undersøkelsen er det flere områder hvor det anbefales å gjennomføre nye undersøkelser. Det første vil være å undersøke sammenhengen mellom seleksjonstester og prestasjoner oppnådd på ARS. Hvor godt presterer kandidatene etter endt utdanning? Dette kan være basert på både gjennomføringshastighet for de ulike utsjekkene og hvilke utsjekker hver enkelt kandidat oppnår. Også prestasjoner oppnådd i K&V bransjen i sammenheng med turnover kan være interessant. Den største utfordringen for slike undersøkelser er bruken av dikotome kriterier. K&V bransjen baserer seg først og fremst på disse kriteriene. Dette vil være et problem for reliabilitet og validitet for kriteriene. På samme tid har undersøkelser vist, at dette er mulig å få til. Det kunne også vært interessant å gjennomføre en større undersøkelse i samarbeid med andre land. Dette ville gitt muligheten til å få en større N og mer statistisk styrke i undersøkelsen.

K&V bransjen rekrutterer også mange til AWACS miljøet i Geilenkirchen i Tyskland. Det kunne være interessant å se på sammenhengen mellom seleksjonstestene ved LSES og oppnådde prestasjoner i dette fagmiljøet. En siste mulighet for videre forskning kan også være om det er forskjell på personlighet for Varsling og Kontroll ved bruk av 5PF Mil. Selv om begge disse fagsøylene er i samme bransje, er det likevel forskjell på disse to fagretningene.

8 Litteraturliste

- Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied*, 1, 270-304.
- Boer, L. C. (1991). Spatial Ability and Orientation of Pilots. I R. Gal & A. D. Mangelsdorff (Red.), *Handbook of Military Psychology* (s. 103-114). Chichester: John Wiley & Sons Ltd.
- Broach, D., & Manning, C. A. (1997). Review of air traffic controller selection: An International Perspective DOT/FAA/AM-97/15. Washington, DC: Federal Aviation Administration Office of Aviation Medicine.
- Brown, K. G., Le, H., & Schmidt, F. L. (2006). Specific aptitude theory revisited: Is there incremental validity for training performance? *International Journal of Selection and Assessment*, 14(2), 87-100.
- Camp, G., Paas, F., Rikers, R., & van Merriënboer, J. (2001). Dynamic problem selection in air traffic control training: A comparison between performance, mental effort and mental efficiency. *Computers in Human Behavior*, 17(5), 575-595.
- Carretta, T. R., & King, R. E. (2008). USAF Enlisted air traffic controller selection: Examination of the predictive validity of the FAA air traffic selection and training battery versus training performance. *Aviation, Space, and Environmental Medicine*, 79, 36-43.
- Carretta, T. R., & Ree, M. J. (2000). Pilot Selection Methods *United States Air Force research laboratory*. San Antonio Center for Leadership Studies Our Lady of the Lake University
- Carroll, J. B. (2003). The higher-stratum structure of cognitive abilities: Current evidence supports g and about ten broad factors. *The scientific study of general intelligence: Tribute to Arthur R. Jensen*, 5-21.
- Cohen, J. (1988). Statistical power analysis for the behavioral sciences. *Hillsdale, NJ: Lawrence Earlbaum Associates*, 2.
- Command, N. A. A. (2016). NATO Air Policing. fra <https://www.airn.nato.int/page5931922/-nato-air-policing>
- Damitz, M., Chetcuti, F., & Henriques, P. (2010). Validation of the First European Air Traffic Controller Selection Tests (1.0 utg.). Belgium: EUROCONTROL.
- Damitz, M., & Eissfeldt, H. (2004). The Relevance of General Cognitive Ability (g) for training success of Ab-initio Air Traffic Controllers. I K.-M. Goeters (Red.), *Aviation Psychology: Practice and Research* (s. 135-151). Aldershot, England: Ashgate Publishing Limited.
- Eissfeldt, H. (1995). The Dynamic Air Traffic Control Test. I N. Johnston, R. Fuller & N. McDonald (Red.), *Aviation Psychology: Training and Selection* (s. 130-140). England: Ashgate Publishing Limited.
- Eissfeldt, H. (2002). ATC-Specific Tests. I H. Eissfeldt, M. C. Heil & D. Broach (Red.), *Staffing the ATM System: The selection of air traffic controllers* (s. 85-96). Aldershot: Ashgate Publishing Limited.
- Eissfeldt, H., & Heintz, A. (2002). Ability Requirements for DFS Controllers - Current and Future. I H. Eissfeldt, M. C. Heil & D. Broach (Red.), *Staffing the ATM system: The selection of air traffic controllers* (s. 13-24). Aldershot: Ashgate Publishing Limited.
- Engvik, H. (1994). "Big Five" på norsk. *Tidsskrift for Norsk Psykologforening* 30, 884-896.

-
- Engvik, H., & Føllesdal, H. (2005). The big five inventory på norsk. *Tidsskrift for Norsk Psykologforening*, 42, 128-129.
- EUROCONTROL. (2016). Who we are. fra <https://www.eurocontrol.int/articles/who-we-are>
- Fahey, R. P., Rowe, A. L., Dunlap, K. L., & deBoom, D. O. (2001). Synthetic Task Design: Cognitive task analysis of AWACS weapons director teams *United States Air Force research laboratory*. Mesa Arizona: Air Force research Laboratory human effectiveness directorate warfighter training research division
- Forsvaret. (2015). Kontroll og Varsling. fra <https://forsvaret.no/hogskolene/Luftkrigsskolen/utdanning/utdanningstilbud/kontroll-og-varsling>
- Forsvaret. (2016). Forsvarets opptak og seleksjon. fra <https://forsvaret.no/fos>
- Forsvarets Høgskole. (2014). *Forsvarets fellesoperative doktrine*. Oslo: Forsvarsstaben.
- Forsvarssjefen. (2008). *Forsvarssjefens beslutningsnotat nr 2/2008 - Seleksjon til befalsutdanning i Forsvaret*. Oslo: Forsvarssjefen.
- Gjestrud, G., Engetrøen, R., Stamsø, M. B., & Martinsen, Ø. L. (2009). *Perspektiver på ledelse* (3. utg. utg.). Oslo: Gyldendal Akademisk.
- Goeters, K.-M., Maschke, P., & Eissfeldt, H. (2004). Ability Requirements in Core Aviation Professions: Job Analysis of Airline Pilots and Air Traffic Controllers. I K.-M. Goeters (Red.), *Aviation Psychology: Practice and Research* (s. 99-119). Burlington, USA: Ashgate Publishing Limited.
- Huffcutt, A. I., Culbertson, S. S., & Weyhrauch, W. S. (2013). Employment Interview Reliability: New meta-analytic estimates by structure and format. *International Journal of Selection and Assessment*, 21(3), 264-276.
- Huffcutt, A. I., Roth, P. L., & McDaniel, M. A. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology*, 81(5), 459-473.
- Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological bulletin*, 96(1), 72-98.
- Hättig, H.-J. (1991). Selection of air traffic control cadets. I G. Reuven & A. D. Mangelsdorff (Red.), *Handbook of Military Psychology* (s. 115-129). Chichester: John Wiley & Sons.
- Inspektorat, S. L. (2016). *RFL 301-3 Reglement for trening, autorisering og vedlikeholdstrening av operativt personell innen kontrollfunksjoner - Kontroll og Varsling*. Rygge Sjef Luftoperativt Inspektorat.
- Jacobsen, D. I. (2015). *Hvordan gjennomføre undersøkelser?: innføring i samfunnsvitenskapelig metode* (3. utg. utg.). Oslo: Cappelen Damm akademisk.
- Jones, A. (1991). The contribution of psychologists to military officer selection. I G. Reuven & A. D. Mangelsdorff (Red.), *Handbook of Military Psychology* (s. 64-80). Chichester: John Wiley & Sons Ltd. .
- Kallus, K. W., Barbarino, M., & Van Damme, D. (1997). Model of the Cognitive aspects of air traffic control *European organisation for the safety of air navigation*. Belgium: EUROCONTROL.
- Luftforsvaret, G. f. (2016). *BFL 301 Bestemmelser om operativt personell i Luftforsvaret - Kontroll og Varsling*. Rygge.
- Martinussen, M. (1996). Psychological Measures As Predictors of Pilot Performance: A Meta-Analysis. *The International Journal of Aviation Psychology*, 6(1), 1-20.
- Martinussen, M. (2005). Seleksjon av flygere og flygeledere. *Tidsskrift for Norsk psykologforening*, 42, 291-299.

-
- Martinussen, M., & Hunter, D. R. (2008). *Luftfartspsykologi*. Bergen: Fagbokforlaget Vigmostad & Bjørke AS.
- Martinussen, M., Jenssen, M., & Joner, A. (2000). *Selection of air traffic controllers: Some preliminary findings from a meta-analysis of validation studies*. Paper presented at the Proceedings from the 24th EAAP (European Association for Aviation Psychology) conference
- Martinussen, M., & Torjussen, T. (1998). Pilot selection in the Norwegian Air Force: A validation and meta-analysis of the test battery. *The international Journal of Aviation Psychology*, 8, 33-45.
- Martinussen, M., & Torjussen, T. (2004). Initial validation of a computer-based assessment battery for pilot selection in the Norwegian Air Force. *Human Factors and Aerospace Safety*, 4, 233-243.
- Martinussen, M., Torjussen, T., Storsve, O., & Hjerkin, O. (2004). *Pilot selection in the Norwegian Air Force: From paper and pencil to computer-based assessment*. Paper presentert på 40th Applied Military Psychology Symposium.
- McClelland, D. C. (1993). Intelligence is not the best predictor of job performance. *Current Directions in Psychological Science*, 2(1), 5-6.
- McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. D. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of applied psychology*, 79(4), 599-616.
- Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology*, 79(6), 845-851.
- Olsen, L. J. (2002). *Seleksjon av kadetter til Luft Kommando og Kontroll*. Fordypningsoppgave Luftkrigsskolen, Trondheim.
- Passer, M. W., & Smith, R. E. (2011). *Psychology: The science of mind and behavior* (Fifth utg.). New York: The McGraw-Hill Companies.
- Pecena, Y., Keye, D., Conzelmann, K., Grasshoff, D., Maschke, P., Heintz, A., & Eißfeldt, H. (2013). Predictive validity of a selection procedure for air traffic controller trainees. *Aviation Psychology and Applied Human Factors*, 3(1), 19-27.
- Rathje, H., Golany, Z., & Eissfeldt, H. (2004). Pan-European selection test battery for air traffic controller applicants. I K.-M. Goeters (Red.), *Aviation Psychology: Practice and Research* (s. 171-201). Aldershot, England: Ashgate Publishing Limited
- Raven, J. C. (1941). Standardization of progressive matrices. *British Journal of Medical Psychology*, 19(1), 137-150.
- Ree, M. J., & Carretta, T. R. (1996). Central Role of g in military pilot selection. *The International Journal of Aviation Psychology*, 6(2), 111-123.
- Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science*, 1(3), 86-89.
- Ringdal, K. (2013). *Enhet og mangfold: samfunnsvitenskapelig forskning og kvantitativ metode* (3 utg.). Bergen: Fagbokforlaget.
- Schmidt, F. L. (2002). The role of general cognitive ability and job performance: Why there cannot be a debate. *Human Performance*, 15(2), 187-210.
- Schmidt, F. L., & Hunter, J. E. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological bulletin*, 124(2), 262-274.
- Schmidt, F. L., & Hunter, J. E. (2014). *Methods of meta-analysis: Correcting error and bias in research findings* (3 utg.). Thousand Oaks, California: Sage publications.

-
- Schmidt, F. L., Oh, I. S., & Shaffer, J. A. (2016). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 100 years of research findings. *Fox School of Business Research Paper*, 1-73.
- Schroeder, D. J., Broach, D., & Young, W. C. (1993). Contribution of personality to the prediction of success in initial air traffic control specialist training. *DOT/FAA/AM-93/4*. Washington, DC: Federal Aviation Administration Office of Aviation Medicine.
- Skoglund, T. H., Martinussen, M., & Lang-Ree, O. C. (2015). Papir versus PC. *Tidsskrift for Norsk psykologforening*, 51(6), 450-452.
- Spearman, C. (1924). The nature of "intelligence" and the principles of cognition. *Journal of Philosophy*, 21(11), 294-301.
- Stranden, A. L. (2015). Her velger psykologer ut lederspirer til Forsvaret. fra <http://forskning.no/arbeid-ledelse-og-organisasjon-organisasjonspsykologi/2015/07/her-velger-psykologer-ut-lederspirer>
- Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004). The end of the Flynn effect?: A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence*, 32(4), 349-362.
- Sundet, J. M., Tambs, K., Magnus, P., & Berg, K. (1988). On the question of secular trends in the heritability of intelligence test scores: A study of Norwegian twins. *Intelligence*, 12(1), 47-59.
- Suresh, A., Ramachandran, K., & Srivastava, A. (2012). Personality based job analysis of air traffic controller. *Indian Journal of Aerospace Medicine*, 56(2), 21-31.
- Utdanningsinspektorat, L. (2010). *BFL 600-1 Bestemmelser for utdanning i Luftforsvaret*
Rygge