

Seleksjon av offshore helikopterpiloter:

Predikerer psykologiske tester utfallet på
simulatorevaluering?

(Selection of offshore rotary wing pilots: Do psychological tests predict
simulator performance?)



Spesialistoppgave arbeids- og organisasjonspsykologi

Psykolog Tore Wingestad

Forsvarets Institutt for Ledelse

Abstract

The purpose of this study was to examine the predictive validity of psychological tests used for the selection of helicopter pilots. Nine psychological tests, cognitive and psychomotor, were used. In addition a psychological interview was conducted. The total sample was 100, and 67 % passed the simulator evaluation (criterion). Both the unweighted mean test score ($r = .51$) and the psychologist's ratings based on the interview ($r = .54$), predicted pass/fail in the simulator. Findings suggest that predictions of simulator performance are satisfactory, using psychological tests and evaluation. Thus, economical and safety issues are retained with this selection procedure.

Forord

Ønsket mitt da jeg skulle skrive oppgave for å fullføre spesialistutdanningen i arbeids- og organisasjonspsykologi, var å utføre en valideringsundersøkelse. Da jeg gjennom de siste årene har utført seleksjonsoppdrag for ulike offshore helikopterselskaper, var det naturlig å foreta en systematisk gjennomgang av hvordan testene som har blitt brukt, har predikert flygeradferd i simulator.

Jeg vil takke helikopterselskapene for data jeg har fått tilgang på og for et utmerket samarbeid i alle faser. Videre vil jeg takke min veileder dr. psychol Monica Martinussen (Universitetet i Tromsø/FIL) for meget nyttige tilbakemeldinger, det samme fra sjefpsykolog og forskningssjef i Forsvaret, prof. dr. philos Jon Christian Laberg og psykologspesialist Tore Torjussen. Det må også rettes en takk til psykolog Jon Lars Syversen for gjennomlesning og oversettelse. Ikke minst rettes takk til vernepliktig Erik Andreassen ved Forsvarets Institutt for Ledelse, for meget god hjelp med å få oppgaven ferdig.

Tore Wingestad

Akershus Festning 1. februar 2005

Innledning

Seleksjon er en betegnelse på en utvelgelsesprosess hvor man ved hjelp av en eller flere metoder/tester tilstreber å identifisere de personer i en søkermasse som med størst sannsynlighet vil lykkes i en bestemt utdanning eller jobb. Beslutningene om hvilke som skal velges ut eller ikke baseres på en vurdering av informasjon man mener er relevant for yrket eller skolen det gjelder. Det er derfor viktig å dokumentere at det er et samsvar mellom metode-/testresultater seleksjonen og de framtidige skole- eller jobbprestasjoner (kriterier). Et slikt samsvar mellom oppnådde resultater og kriteriet, kalles prediktiv validitet, og uttrykkes i korrelasjoner som går fra .0 – 1.0, hvor 1.0 er uttrykk for fullstendig positivt samsvar. Formålet med denne artikkelen er å undersøke den prediktive validiteten av metoder brukt i seleksjon av offshore helikopterpiloter, mot et indirekte kriterium, nemlig prestasjoner oppnådd i helikoptersimulator.

I Norge, og flere andre land, er det Forsvaret som utdanner de fleste flygere. Psykologiske metoder ble tidlig tatt i bruk for å velge ut elever til militære flyskoler. I Norge startet man like etter andre verdenskrig med å anvende psykologiske tester i utvelgelsen (Riis, 1955). Siden den gang er testbatteriet utvidet, videreutviklet (Torjussen & Hansen, 1999) og validert flere ganger (Martinussen & Torjussen, 1998).

Kort flygerseleksjonshistorikk

Litt spøkefullt er det sagt at den første flygerseleksjonen ble foretatt av brødrene Orville og Wilbur Wright i 1903, da de kastet mynt og kron om hvem som skulle ta den første flyturen med deres flykonstruksjon (Hunter & Burke, 1995). I en artikkel i *The Lancet* fra 1918: ”*Essential Characteristics of Successful Aviators*”, basert på observasjoner av piloter fra 1. verdenskrig, konkluderte man at flygere ikke var ”supermenn”, men gode sportsmenn med initiativ og humor (Turnbull, 1992). Testene man bruker i dag er mer sofistikerte, og et resultat av en mer systematisk utvikling og validering av metoder, gjennom nesten 100 år. (Hunter & Burke, 1995).

Å bli flyger har alltid vært populært. Derfor har det alltid vært mange søkere, og følgelig har en kunnet plukke ut de med gode skolekarakterer og attester. Likevel viste det seg at dette var et utilstrekkelig seleksjonsgrunnlag, fordi det viste seg at de fleste av de som på denne måten

ble valgt ut som elever falt igjennom som flygere. Derfor gikk psykologer og psykiatere ved flyskoler i USA og Canada, i samarbeid med erfarne flygere, i gang med å utvikle metoder med høyere prediktiv validitet (Storsve, 1983).

Utviklingen av flyet og flygeryrket har etter andre verdenskrig ført til at de fleste lands flyvåpen anvender psykologiske utvalgsmetoder ved siden av strenge medisinske krav og uttaksflyging. Årsaken til dette er først og fremst at a) det er kostbart å utdanne flygere og b) at konsekvensene av flygerfeil er så alvorlige. Man var opptatt av å finne fram til hvilke evner, anlegg og personlighetstrekk som hadde betydning for flygeryrket, og hvordan man kunne måle dette. Det er stor grad av internasjonal enighet om hvordan disse skal vektlegges og måles (Hunter & Burke, 1995). Evnetestene som tradisjonelt har vært brukt kan deles i to hovedkategorier:

- Generelt evnenivå: Består som regel av papir- og blyanttester som måler resonering, hukommelse, teknisk innsikt og romoppfatningsevne.
- Psykomotoriske ferdigheter: Måler koordineringsevne, reaksjonstid, informasjonsprosessering og simultankapasitet.

Meta-analyser av valideringsstudier har dokumentert god prediktiv validitet for evne- og psykomotoriske tester for flygersелеksjon (Hunter & Burke, 1995; Martinussen, 1997). I tillegg til evnetestene har en rekke personlighetstester vært undersøkt, men det har vært langt vanskeligere å dokumentere prediktiv validitet for disse testene sammenlignet med evnetestene (Martinussen, 1997).

Hvorfor er seleksjon viktig?

Hvorfor skal man bruke tid og penger på å selektere nye medarbeidere? Det er mye økonomi i systematisk personalutvelgelse, og det er en stor menneskelig gevinst ved å velge folk til stillinger de egner seg for og trives med. Ikke minst er sikkerhetsmessige aspekter viktige, noe som er sentralt spesielt blant flygere og flygeledere.

Forsvaret har i større grad enn sivile virksomheter hatt mulighet for å systematisere sine erfaringer med seleksjon. Dette skyldes ikke minst det store antallet kandidater som er testet og selektert til flygerutdanning gjennom mer enn 50 år. Selv om organisasjonen er spesiell og stiller spesielle krav til visse kategorier medarbeidere, så er erfaringene overførbare til sivile

forhold. I praksis kan man si at visse kategorier sivilt flygende personell har minst like store utfordringer som Forsvarets, kanskje spesielt innen offshore.

Luftforsvarets seleksjon til flygerutdanning består av en rekke tester og formelle krav (Martinussen & Torjussen, 1998; Torjussen & Hansen, 1999). I tillegg anvendes en personlighetstest og psykologintervju. Formålet med intervjuet er blant annet å vurdere søkerens motivasjon og personlige egenskaper, der evnen til å kommunisere er sentral.

Når det gjelder sivil flygerutdanning i Norge er det ingen spesielle krav til grunnleggende egenskaper som stilles, og det er derfor ulik praksis når det gjelder opptak. For mange skoler kreves det kun legeattest. Hvis man har klart sine eksamener og oppflyging for ulike sertifikater, er man godkjent som flyger. Dog kan ulike fly- eller helikopterselskaper stille krav selskapene selv mener er nødvendige. Disse kravene kan føre til at allerede utdannede og erfarne piloter ikke holder mål. Det er betenkelig at en del flygere er operative uten at sentrale evner og ferdigheter er evaluert før vedkommende starter sin pilotkarriere.

Seleksjon bør måle egenskaper som har lite utviklingspotensiale, det vil si at senere opplæring i begrenset grad kan kompensere for lavt nivå på disse. Eksempelvis vil en person som har under gjennomsnittelig evnenivå, eller dårlige spatiale egenskaper, fortsatt ha det selv om vedkommende gjennomgår trening.

Det mest "rettferdige" ville selvfølgelig være å la alle som søkte Forsvarets flyskole eller et sivilt selskap, få lov til å prøve seg under en kontrollert opplæringsperiode. Men det sier seg selv at dette ville bli praktisk ugjennomførbart, og særdeles kostbart. Nettopp på grunn av at grunnutdanningen foregår i så kontrollerte former kan det jo også være at egenskaper som er ugunstige for flygeryrket ikke blir avdekket der, men kan få fatale følger senere. Testene blir derfor et viktig virkemiddel for å redusere kostnader og ulykker.

Som testbruker eller kunde behøver man ikke vite særlig mye om utviklingen av en test, det holder å kjenne til noen grunnbegreper. Det er testkonstruktørens og testutviklerens oppgave å sørge for at normer, reliabilitet og validitet er tilfredsstillende (Madsen, 1991). Samtidig har også testbruker og oppdragsgiver et ansvar for riktig gjennomføring, anvendelse og tolkning

av testene. En test som er reliabel og valid for en gruppe behøver ikke å være det for en annen.

Reliabilitet og validitet

Reliabilitet er et uttrykk for testens nøyaktighet og pålitelighet. Er den et presist mål for det den er ment å måle, eller er det relativt tilfeldig hva folk skårer på testen? Hvis man måler samme person to ganger med den samme testen bør vedkommende helst oppnå det samme begge gangene (test-retest reliabilitet). En annen måte å anslå testens reliabilitet på, er ved hjelp av ulike mål på indre konsistens, for eksempel dele testen i to like deler og beregne korrelasjonen mellom delene ("split-half" reliabilitet) (Hellevik, 1999).

Validitet er knyttet til gyldigheten og anvendeligheten av slutninger man trekker på bakgrunn av testskårene. Måler den virkelig det den antas å måle? Måler intelligens testen virkelig intelligens, eller måler den noe helt annet, for eksempel utdanningsnivå? Dette betegnes som begrepsvaliditet. En annen slutning er om testen kan predikere framtidige jobbprestasjoner og vi snakker da om prediktiv validitet eller kriterievaliditet. Dette undersøkes vanligvis ved å sammenligne testprestasjoner med et mål på senere jobbprestasjoner (kriterium).

Kriterier

Personellutvelgelse handler om å ansette de som er best egnet til jobben. Hvordan finner vi så ut om prognosen stemmer? Det gjøres ved å sammenligne testprognosen med enten objektive mål på produksjon eller mer subjektive vurderinger foretatt av en overordnet. Kriteriene bør ifølge Madsen (1991) oppfylle noen minimumskrav:

- De skal være relevante, det vil si mål på suksess eller produktivitet.
- De skal gå på jobbets vesentligste aspekter, ikke legge overdreven vekt på detaljer.
- De skal være praktisk brukbare, kriteriene skal kunne framskaffes ved rimelige omkostninger og prosedyrer. I tillegg bør de kunne måles på en reliabel måte og ha god begrepsvaliditet.

Kriterier på flygerprestasjoner har tradisjonelt vært bestått/ikke bestått flygerutdanning, og kun i svært sjeldne tilfeller har prestasjoner etter endt flygerutdanning vært brukt (Martinussen, 1997).

Seleksjon av offshore helikopterpiloter

Alle søkerne som blir vurdert har flygererfaring, fordi oljeselskapene (ikke helikopterselskapene) har satt som krav at nyansatte piloter må ha et minimum på 1000 timer flyerfaring. I tillegg må de igjennom et intervju, en legeundersøkelse og få godkjent simulatorevaluering før de kan begynne på opplæringsperioden. De siste årene har man i tillegg innført bruk av psykologiske tester og psykologintervju.

Grunnen til at man stiller høye krav for disse pilotene er kanskje at det er enda vanskeligere å manøvrere et helikopter i for eksempel offshore-oppdrag enn "vanlig flygning". Man flyr ofte med dårlig sikt (instrumentflyging) og sterke lyspunkter fra rigg. Samtidig kan det være kraftig turbulens og "hvite omgivelser". I motsetning til et fly som bare lander rett fram og hvor begge flygerne er med hele tiden, vil et helikopter både ha rene høyre- og venstrelandinger som ikke i like stor grad kan følges opp av sidemannen.

Problemstillinger

Hensikten med denne undersøkelsen er å undersøke om det er noen sammenheng mellom de psykologiske metodene (tester og intervju) som benyttes ved flygeruttak, og flygerens prestasjoner i simulator. Simulatortesting er en obligatorisk del av utvelgelsesprosedyren som helikopterselskapene bruker før ansettelse. Kan man redusere antall kandidater som skal prøves i simulator ved å teste dem først? Kan testene si noe om pass/fail i simulator?

Metode

Utvalget

Søkere til to større norske offshore helikopterselskaper deltok i undersøkelsen. Totalt består utvalget av 100 søkere (det er så få kvinner at jeg ser begge kjønn under ett). De ble testet i perioden 2000-2004. Søkernes alder varierte fra 22 til 46 år ($M = 32.7$, $SD = 5.7$) antall flytimer varierer fra 600 til 9000 timer ($M = 2690$, $SD = 1970$).

Beskrivelse av testene

Testbatteriet som er brukt her består av åtte forskjellige prøver. Testskårene (råskårene) på den enkelte test blir regnet om til en 9-delt skala (prøveklasser, hvor klasse ni er best og en

dårligst) basert på normer fra Luftforsvaret. I undersøkelsen er det tatt utgangspunkt i råskårene ved beregning av korrelasjoner.

Ravens progressive matriser (Raven,1994): Dette er en klassisk evnetest. Oppgavene går ut på at man skal finne system i et mønster og så velge den figuren som mangler. Det er i alt 36 oppgaver med stigende vanskelighetsgrad. Dette er den best dokumenterte testen for å kunne forutsi suksess ved høyere utdanning (Hjerkin, 1994). Testen er et godt mål på generell intelligens og anvendes av Forsvaret til seleksjon av både flygere og søkere til krigsskolen.

Tallrekker: Her skal søkeren finne ut hvilken regel en rekke med tall er satt opp etter, og så skrive ned de to neste tallene i rekken. Testen krever logisk resonnering.

Figurformer: Figurer er satt opp og nummerert på venstre side av et svarark. Kandidatene skal så finne igjen de enkelte bitene i høyre del av svararket. Den enkelte figur blir presentert rotert, men ikke speilvendt, blant andre figurene. Testen måler mental rotasjonsevne.

Instrumenttyding: Testen er ment å måle evnen til å visualisere hvordan et fly oppfører seg i luften, samt det å kunne lese av instrumentene og kombinere informasjonen fra dem. Kandidatene ser to instrumenter – kompass og kunstig horisont – og skal ut fra dette kunne avgjøre hvilken av fem avbildede flymaskiner instrumentene refererer seg til. Denne prøven måler både romoppfatning og resonnering (Hansen, 1987).

Retningsreversaler: Dette er en prøve som skal si noe om evnen til å holde greie på retninger (opp, ned, høyre og venstre) når man har liten tid til å tenke seg om. På svararket er det plassert 16 ruter slik at de danner fire rekker og fire kolonner. Ordrene om hvilke ruter som skal merkes blir gitt fra cd-spiller. Ordrene leses langsomt til å begynne med og så hurtigere etter hvert. Generelt måler testen evne til hurtig informasjonsbearbeiding .

Tallsøker: Testen består av 100 tallrubrikker. Kandidaten blir bedt om å finne, i nummerert rekkefølge, ulike tall så raskt som mulig. Denne testen er ment å måle perseptuell hurtighet, konsentrasjon og korttidshukommelse.

Tidsvurdering og romoppfatning: Dette er to prøver som løses samtidig, en tidsvurderingsdel og en romoppfatningsdel. Mens tidsintervallene avspilles skal søkeren arbeide med den andre oppgaven, som er en romoppfatningsoppgave. Personen får et ark med bilder av klosser som er i berøring med hverandre, og man skal telle antall kontaktpunkter mellom en spesiell kloss og andre som ligger rundt.

Tapping: Denne oppgaven administreres under psykologintervjuet. Personen holder en blyant i hver hånd, og skal flytte blyantene i takt med et lydsignal. Under prøven blir personen samtidig bedt om å løse regneoppgaver eller stave forskjellige ord for- eller baklengs. Denne testen er ikke normert, men det blir gitt en vurdering på rytme og motorikk, oppgaveløsning/simultankapasitet og stresstoleranse. Den foretatte vurderingen omgjøres til prøveklasser.

Flygerprognosen blir regnet ut på bakgrunn av alle disse testresultatene og resultatet på tappingen. Skåren er et uvektet gjennomsnitt av alle testene.

Prosedyre

Datainnsamlingen har foregått i to omganger. Først gjennomførte søkerne testene og ble intervjuet. Deretter ble kriterieoppnåelsen innhentet i form av prestasjoner i simulator. Alle, uansett testresultater, ble vurdert i simulator. Vurderingen i simulator skjedde rett etter at testingen fant sted. Instruktørene som foretok denne vurderingen kjente ikke til testresultatene eller psykologens vurderinger.

Intervju (personvurderingen)

Evnen til å kommunisere godt er sentral i vurderingen av søkerne. Intervjuet gir søkeren muligheten til å formulere seg muntlig, og intervjueren kan stille utdypende spørsmål. I motsetning til spørreskjemaer kan man utforske et tema inntil man har forstått hva vedkommende mener eller prøver å si. Intervjuet gir bedre enn skriftelige prøver, mulighet for å vurdere to forhold som er viktig for jobb- og skoleprestasjoner: motivasjon for arbeidet, og måten å forholde seg til andre mennesker på (Madsen, 1991).

Under intervjuet blir altså søkeren observert i forhold til kommunikative ferdigheter. Virker vedkommende nølende og tilbakeholden overfor noen spørsmål? Hva er grunnen til det? Kontakten under intervjuet gir en mulighet til å vurdere om og i hvilken grad kandidaten er

likefram eller unnvikende. Man kan se iver og interesse eller en mer lukket og mistroisk stil. Søkeren vurderes etter kriterier som motivasjon og planer, samarbeidsevne, selvinnsikt, kommunikasjonsevne, empati og eventuelle psykiske forstyrrelser.

Simulator

I simulatoren er det både flygerferdigheter, samarbeid og kommunikasjon i cockpit som evalueres. Det er noe ulike prosedyrer i de forskjellige selskapene, men informasjon om dette har ikke blitt inkludert i denne undersøkelsen fordi forskjellene har hatt liten betydning for resultatene.

Det man vurderer i simulatoren er om søkerne har god oversikt, om de er systematiske, om de samarbeider bra, om de tåler kritikk og stiller spørsmål. Det vurderes videre om de har god spatialorienteringsevne, utholdenhet, godt humør og om de har lederegenskaper. Den som blir testet, flyr sammen med en "co-pilot", som vedkommende kan bruke under flygningen. I simulator og under opplæring er det ofte instrumentflygningen (navigering uten bruk av visuelle holdepunkter) som er det største problemet. Dette kan skyldes både dårlig grunnutdanning og/eller marginale kandidater som vil kreve lang opplæringstid senere. En samlet vurdering av samhandling i cockpit og flygerferdigheter danner grunnlaget for bestått/ikke bestått.

Resultater

Av utvalget på 100 var det til sammen 67 personer som bestod simulatorevalueringen. Resultatene fra undersøkelsen er samlet i tre tabeller. I tillegg presenteres to figurer som illustrerer person- og flygerprøveklassene (vurderinger) i forhold til andelen bestått/ikke bestått i simulator.

Tabell 1 viser gjennomsnittsverdi og standardavvik for testene. Det bør påpekes at ikke alle resultatene er normalfordelte, det gjelder spesielt testen instrumenttydning, hvor 47% av søkerne havnet i prøveklasse 8 og 9. Dette kan forventes, siden søkerne har flygererfaring, og dermed er kjent med bruk av lignende instrumenter.

Tabell 1Deskriptiv statistikk for testene i testbatteriet ($N = 100$)

Tester	<i>M</i>	<i>SD</i>
Tallrekker	4.7 (9.7)	2.0 (4.2)
Raven	4.6 (22.9)	2.2 (6.0)
Figurformer	5.7 (23.3)	1.9 (7.9)
Instrumenttydning	6.8 (45.6)	2.0 (12.5)
Retningsreversalier	5.3 (35.4)	2.2 (15.4)
Tallsøker	4.6 (37.9)	2.0 (10.3)
Romoppfatning	6.4 (13.8)	2.2 (9.2)
Tidsvurdering	4.6 (58.5)	2.1 (33.6)
Tapping (Rytme/Motorikk)	5.6	1.9
Tapping (Sim. kap.)	5.3	2.1
Tapping (Stress)	5.3	2.1
Sum tapping	5.4	2.0
Personvurdering (Intervju)	5.7	2.1
Flygerprognose	5.3	1.4

Note: Råskårer (i parentes) er konvertert til stanineskala (1-9).

Tabell 2Interkorrelasjoner mellom prediktorer ($N = 100$)

Tester	1	2	3	4	5	6	7	8	9	10	11	12	13
1 Tallrekker	-												
2 Raven	.64**	--											
3 Figurformer	.23*	.38**	-										
4 Instrumenttydning	.37**	.44**	-.13	-									
5 Retningsreversalier	.55**	.55**	.19	.54**	-								
6 Tallsøker	.34**	.47**	.29**	.32**	.47**	-							
7 Romoppfatning	.40**	.51**	.25*	.27**	.42**	.20*	-						
8 Tidsvurdering	-.25*	-.18	-.07	.02	-.11	-.09	-.06	-					
9 Sum tapping	.51**	.59**	.42**	.33**	.61**	.38**	.43**	-.11	-				
10 Personvurdering	.51**	.58**	.33**	.42**	.56**	.42**	.38**	-.13	.75**	-			
11 Flygerprognose	.73**	.82**	.44**	.55**	.79**	.62**	.63**	-.29**	.77**	.73**	-		
12 Flytimer (n=54)	-.09	-.16	-.32*	-.05	-.25	-.27*	-.22	-.09	-.28*	-.39**	-.29*	-	
13 Alder	-.20*	-.25*	-.14	-.17	-.32**	-.32**	-.05	.06	-.29**	-.31**	-.33**	.71**	-

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

Tabell 2 er en oversikt over interkorrelasjoner mellom prediktorene. Opprinnelig ble tapping-resultatene delt i tre kategorier (rytme/motorikk, sim. kap. og stress), men fordi disse hadde interkorrelasjoner på over 0.90 er de slått sammen til en variabel (sum tapping). De ulike evnetestene er som forventet positivt korrelert, med unntak av romoppfatning og tidsvurdering. Dette skyldes nok at disse to oppgavene utføres samtidig. Det vil si at personene delvis vurderer lengden på tidsintervall samtidig som en spatial oppgave løses. For de fleste er det nok slik at de velger å konsentrere seg om en av oppgavene, og derfor blir korrelasjonen mellom disse lav. Antall flytimer var noe overraskende negativt korrelert med de fleste prediktorene, med andre ord gjorde de som hadde flest flytimer på forhånd dårligst prestasjon på testene.

Tabell 3 er en oversikt over den prediktive validiteten til de ulike testene og intervjuet. Kriteriet som brukes er prestasjoner i simulator.

Tabell 3

Prediktiv validitet for testene og psykologvurdering mot simulatorprestasjoner (kriterium)

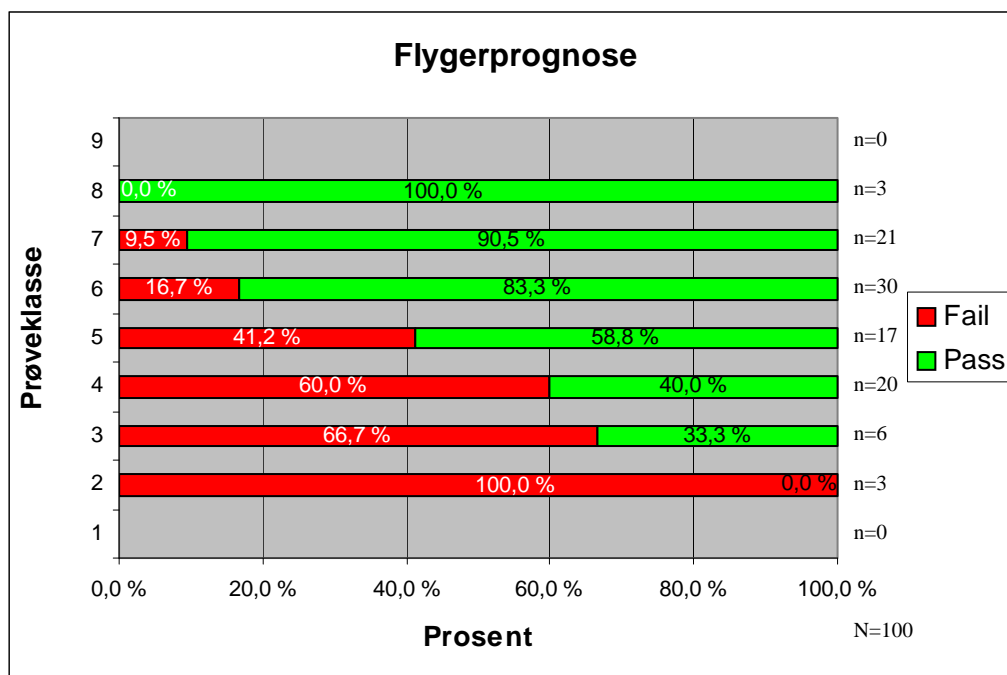
Tester	Simulatorprestasjoner (<i>r</i>)
Tallrekker	.41**
Raven	.44**
Figurformer	.20*
Instrumenttydning	.33**
Retningsreversalier	.38**
Tallsøker	.26**
Romoppfatning	.27**
Tidsvurdering	.02
Rytme/Motorikk	.56**
Sim. kap.	.55**
Stress	.57**
Sum tapping	.59**
Personvurdering	.54**
Flygerprognose	.51**
Flytimer (n=54)	-.23
Alder	-.18

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

Alle tester unntatt tidsvurdering var signifikant positivt korrelert med kriteriet. Raven og tallrekker korrelerer høyt med kriteriet (.44 og .41), noe også de andre testene gjør, med unntak av tidsvurdering. Tappingen (rytme/motorikk, sim.kap. og stress) skiller seg spesielt ut. Sum tapping har en korrelasjon på .59 med simatorevalueringen. Intervjuvurderingen korrelerer nesten like høyt (.54). Alder og erfaring har derimot ingen signifikant sammenheng med bestått/ikke bestått i simulator, selv om det kan se ut som det er en tendens til negativt samsvar (-.18 og -.23).

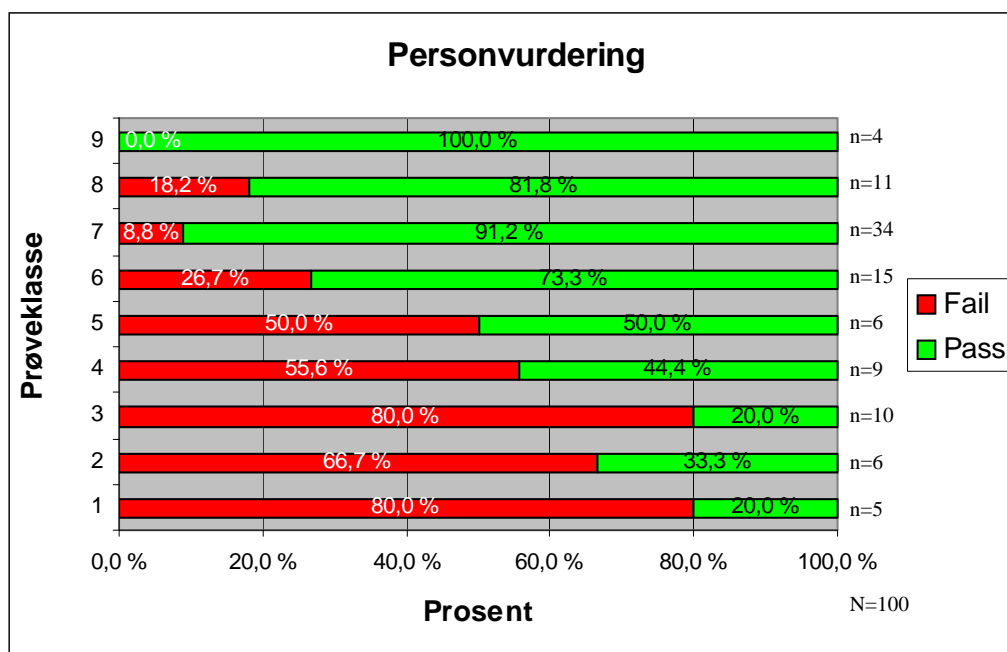
Sammenhengen mellom flygerprognose og andelen bestått er skissert i Figur 1, og sammenhengen mellom personvurdering og andelen bestått er skissert i Figur 2. Figurene angir prosentandelen som består for de ulike prøveklassene. Man får dermed et mer direkte mål på nytten av å anvende testen, enn det korrelasjonskoeffisientene angir.

Figur 1



Figur 1 viser at ingen med dårligere enn prøveklasse tre på flygerprognose har bestått simulator, og det blir suksessivt bedre til prøveklasse åtte, hvor alle har bestått.

Figur 2



Figur 2 viser at alle med personvurdering i prøveklasse ni har bestått, mens prosentandel bestått minker ettersom man beveger seg nedover i prøveklassene. I prøveklasse 1-5 på personvurderingen, får 2/3 av de som er i dette sjiktet ”ikke bestått” i simulator.

Diskusjon

I denne undersøkelsen har vi sett at psykologiske tester, med rimelig stor grad av treffsikkerhet predikerer om en søker til et offshore helikopterselskap på en tilfredsstillende måte vil klare å utføre ulike oppgaver i en helikoptersimulator. Selv om de fleste av søkerne får godkjent i simulator, er det likevel en relativt stor andel som ikke består (33 %). Dette gjelder uten unntak de av søkerne som er sivilt utdannet. Kandidater med militær bakgrunn og erfaring har klart et høyt nok nivå både på evnetester og simulator. Dette antakelig fordi de allerede er selektert på bakgrunn av ulike evne- og ferdighetstester. Denne gruppen er imidlertid ikke skilt ut som en egen gruppe fordi antallet søkere med forsvarsbakgrunn var relativt liten.

Hensikten med denne undersøkelsen var å se om det er en sammenheng mellom testresultater/personvurderingen og simulatorprestasjon. Hvis så var tilfellet, kunne man

kostnadseffektivisere seleksjonen ved at de som får dårlig prøveresultat ikke behøver å prøves i simulator. Resultatene i denne studien viste at det er en høy korrelasjon mellom prediktorer (tester og intervjuvurdering) og kriteriet (simulator).

Hva med de som ikke har gjort det tilfredsstillende på testene, men som har klart seg godt nok i simulator? Alle de som testes har flygererfaring, det vil si at de er mer på hjemmebane i en simulator enn i en testsetting. Testene måler mer spesifikke egenskaper enn simulatoren, og vil derfor si noe mer om hvor den enkeltes nivå ligger enn det man kan finne ut i en flygersetting. Derfor bør de som ikke er gode nok på testene, men klarer seg brukbart i simulator, havne i ”ikke-anbefalt” gruppen. Dette fordi det kan tenkes at denne gruppen ikke har god nok kapasitet når det virkelig gjelder.

Seleksjon som kun baserer seg på godkjente sertifikater og antall flytimer er ikke tilfredsstillende for denne type arbeid. Til det vil risikoen ved å plukke ut uegnede flygere være for høy, og omkostningene dermed for store. Det man har sett ved bruk av testene er at enkelte kan ha svært dårlig stresstoleranse, simultankapasitet og spatiale egenskaper. Denne så og si kollapsen i kognitiv og psykomotorisk fungering kan tenkes å øke risikoen for framtidige ulykker. Enkelte viste også en personlighetsmessig fungering som tilsa liten selvinnsikt, empati eller samarbeidsevne. Denne gruppen bør også være utelukket fra flygende tjeneste uavhengig av test- og simulatorresultater. Man kompenserer ikke for negative personlighets-trekk ved å gjøre det bra på de kognitive testene, siden dette er to ulike aspekter (intellektuell kapasitet og relasjonell kompetanse).

Noe av forklaringen på sammenhengen mellom dårlige testresultater og simulatorprestasjonen er mest sannsynlig at søkerne ikke er testet før flygerutdannelsen. Kravene som stilles for å kunne bli sertifisert flyger er medisinske, ikke psykologiske. Dette er betenkelig når man vet at det er en bred internasjonal konsensus på hvilke egenskaper som er av vesentlig betydning for en flyger for å kunne håndtere problematiske situasjoner (personlig egnethet, gode evner og psykomotorisk kapasitet) (Fallucco, 2002).

Ut fra Tabell 3 kan man lese at det er summen av de ulike testene som gir den beste prediktive validiteten. Den enkelttesten som skiller seg noe ut er tappingoppgaven. Denne kan dog ikke defineres som en standardisert test, siden den ikke er normert og validert. Den utgjør

likevel en viktig del av flygeseleksjonen i Forsvaret, og den anvendes også av det svenske og danske forsvaret. Grunnen til at den korrelerer så høyt som den gjør med simulatorprestasjoner, er kanskje at den frambringer den mest "cockpitlike" situasjonen. Man sitter ikke lenger anonym i et klasserom sammen med andre, men er under direkte observasjon, og kompleksitetsnivået er høyt. Oppgaver skal løses, samtidig som en skal forholde seg til både syn, hørsel og motorikk. Dette utgjør en type stressituasjon som krever god kapasitet i flere evner for å gjennomføre oppgaven på en tilfredsstillende måte.

Tidligere meta-analyser av valideringsstudier viste at gjennomsnittelig korrelasjon med kriteriet for kognitive tester var henholdsvis .22, psykomotoriske tester .20 og personlighet .13. (Martinussen, 1997; Hunter & Burke, 1995). Grunnen til at man har funnet høyere korrelasjoner i denne studien er at gruppen ikke har blitt pre-selektert. Eksempelvis i Forsvaret så vil kun de som har gode nok testresultater få lov til å prøve seg på flygerskolen. Dette fører til at variansen i egenskaper som måles er lavere. Denne reduserte spredningen (restriction of range) i testskårer fører til at den observerte korrelasjonen mellom tester og kriteriet blir lavere (Martinussen, 1997). I denne studien blir derimot kriteriedata innhentet på alle som er testet.

Videre viser det seg også at evnetestene (Raven og tallrekker) predikerte pass/fail i simulator bedre enn de psykomotoriske testene (med unntak av tappingen). Dette er forståelig ut i fra at det generelle evnenivået generelt sett er den beste prediktoren for å predikere skole- og jobbsuksess (Madsen, 1991).

I tidligere undersøkelser har flyerfaring vist seg å være en god prediktor på senere flygerprestasjoner (Martinussen, 1997). Derfor er det kanskje et noe overraskende funn at erfaring ikke er positivt korrelert med resultater i simulator. En mulig hypotese ville være at eldre, erfarne flygere som gruppe skulle gjøre det noe svakere på testene fordi de rett og slett er eldre og reagerer saktere på enkelte tester, men at dette ville bli kompensert for i simulator, hvor de kommer inn i et mer kjent miljø. Det viser seg imidlertid at det ikke er noen sammenheng mellom det å være erfaren og prestasjoner på tester og i simulator. Det kan kanskje komme av at noen av pilotene som har mange flytimer har flydd for selskaper hvor prosedyrene har vært dårlige, slik at de har tilegnet seg uvaner, eller at arbeidsmiljøet og selskapskulturen har vært dårlig. En annen mulig forklaring er at tidligere meta-analyser er

basert på ab-initio (lat. "fra begynnelsen av") seleksjon, det vil si personer uten flygererfaring. Erfaring kan kanskje fungere som prediktor for relativt uerfarne flygere, men kan spille mindre rolle når alle har over 1000 timer. Konsekvensen av det er at erfaring i seg selv ikke er et egnet seleksjonskriterium. Kanskje tvert imot, spesielt hvis erfaringen man har er fra selskaper som har dårlige rutiner og prosedyrer.

Det er stor variasjon i hvordan de ulike kandidatene framstår og presenterer seg selv, og måten de kommuniserer på. Den kategorien som ikke blir anbefalt som følge av intervjuet er de som ikke evner å samarbeide om å gi intervjuerne innblikk i egen personlige fungering. Noen kommuniserer dårlig, og må dras igjennom intervjuet. I tillegg kan det være personlig problematikk som depressive tendenser, eller liten grad av selvinnsikt om hvordan de er og virker inn på andre.

Hvis mye av vedkommendes liv er bundet opp i psykologiske eller relasjonelle problemer, vil det medføre at mentale ressurser og kapasitet er bundet opp i dette. Konsekvensen av dette er at det i kritiske situasjoner vil være mindre mental kapasitet til rådighet hvis fokus og konsentrasjon er bundet opp i en psykologisk konflikt. Manglende evne til kommunikasjon er også en kritisk faktor i cockpit, også dette vurderes under intervjuet.

Et intervju oppnår høyest validitet når det er strukturert, og med gode oppfølgingsspørsmål. Gode spørsmål hjelper kandidaten til å komme med relevant informasjon (Hermans og Mulder, 1998). Et annet moment som er viktig er intervjuerens ferdigheter, hvis ikke de er gode nok vil man kunne feile ved å legge for mye vekt på førsteinntrykk, stereotypier og "kloning" (verdsetting av egenskaper som er framtredd hos intervjuer) (Hunter & Burke, 1995).

Personvurderingene i denne undersøkelsen predikerte også pass/fail i simulator. Faktisk korrelerte personvurdering) med simulatorprestasjon (0.54) på samme nivå som tappingen (0.59). Dette henger antakelig sammen med at også de kommunikative og personlige egenskapene evalueres under simulatorøvelsen.

Oppsummering

Formålet med undersøkelsen var å se om testene predikerte jobbadferd. Til sammen 100 søkere til offshoreflyging ble undersøkt. Resultatene viste en sammenheng mellom testresultater og simulatorutførelse. Hvilke implikasjoner har så disse funnene? Selv om antall undersøkte er stort nok til at det gir statistisk holdbare resultater, er det jo et begrenset utvalg. Likevel er funnene nokså entydige, og man vil kunne trekke noen slutninger. Selv om man er utdannet flyger og har erfaring, betyr ikke det at man har de ønskede egenskapene. Disse ferdighetene er altså ikke noe som utvikles som følge av at man praktiserer som flyger.

Eventuelle oppfølgingsstudier bør følge kandidatene over en lengre periode. Hvordan klarer de som ansettes seg i opplæringsperioden og senere i operativ tjeneste? Samarbeider og kommuniserer de godt med kollegaer og andre? Får de kapteinutsjekk etter hvert?

Videre bør man også se nærmere på hva som går bra og hva som går dårlig i simulatoren. Var det kommunikasjon og samarbeid, stress, simultankapasitet eller spatiale egenskaper som ikke fungerte godt nok?

Konklusjonen er altså en anbefaling til de som ønsker å bli flygere om å først teste evnenivå, psykomotoriske egenskaper og personlig fungering, FØR det brukes tid og ressurser på en kostbar utdanning. Det vil både være en personlig fordel, en fordel for senere utdanningsinstitusjoner og arbeidsgivere, samt en sikkerhetsmessig gevinst. Ved å unnlate å bruke pre-seleksjon er det et skritt tilbake til mynt og kron utvelgelsen til brødrene Wright, hvor tilfeldighetene råder. Videre tyder resultatene på at et ofte brukt seleksjonskriterium som antall flytimer ikke bør tillegges vekt i seg selv. Det er kun en fordel hvis man likevel har de nødvendige egenskapene.

I praksis bør alle som skal fly, uansett om det er privat eller kommersielt, gå igjennom en grunnleggende psykologisk seleksjon på lik linje med legeundersøkelse. Dette bør bli et krav fra luftfartsmyndighetene, hvis ikke må de enkelte selskapene selv ta dette ansvaret.

Referanseliste

- Fallucco, S. J. (2002). *Aircraft Command Techniques*. Aldershot: Ashgate.
- Hansen, I. (1987). Beskrivelse av testene. *Militærpsykologiske Meddelelser (MPM)*, 15.
Oslo: Forsvarets Psykologiske og Pedagogiske Senter.
- Hellevik, O. (1999). *Forskningsmetode i sosiologi og statsvitenskap*. Oslo:
Universitetsforlaget.
- Hermans, P. H. & Mulder, H. W. (1998). Job analysis and the selection interview. Editor K.
M. Goeters, *Aviation Psychology: A Science and a Profession*, (pp.81-92). Aldershot:
Ashgate.
- Hjerkin, O. (1994). Det er forskjell på folk. *LUFTLED Norsk Luftmilitært Tidsskrift*, mars,
25-28.
- Hunter, D. R. & Burke E. (1995). *Handbook of Pilot Selection*. Aldershot: Ashgate.
- Madsen, J. P. (1991). *Systematisk personaludvælgelse: Teori og praksis*. Rungsted:
Forsikringshøjskolens Forlag.
- Martinussen, M. (1997). *Pilot Selection: A validation and meta-analysis of tests used for
predicting pilot performance*. Doktoravhandling. Universitetet i Tromsø.
- Martinussen, M. & Torjussen, T. (1998). Pilot Selection in the Norwegian Air Force: A
Validation and Meta-Analysis of the Test Battery. *International Journal of Aviation
Psychology*, 8, 33-45.
- Raven, J. (1994). *Occupational Users Guide: Raven's Advanced Progressive Matrices and
Mill Hill Vocabulary Scale*. Oxford: Oxford Psychologists Press.
- Riis, E. (1955). Psykologisk utvelgning av flygere. *Militærpsykologiske Meddelelser (MPM)*,
F-2. Oslo: Forsvarets Psykologiske og Pedagogiske Senter.
- Storsve, O. (1983). *Seleksjon av flygere til Luftforsvaret i Norge*. Magisteravhandling.
Universitetet i Oslo.
- Torjussen, T. & Hansen, I. (1999). Forsvaret best i test? Bruk av psykologiske tester i
Forsvaret, med spesiell vekt på flygerseleksjon. *Tidsskrift for Norsk Psykologforening*,
8, 772-779.
- Turnbull, G. J. (1992). A Review of Military Pilot Selection. *Aviation, Space, and
Environmental Medicine*, 63, 825-830.